

新規試験法提案書

眼刺激性試験代替法フルオレセイン漏出試験法

平成25年 1 月

国立医薬品食品衛生研究所

新規試験法提案書

平成 25 年 1 月 20 日

No. 2012-03

眼刺激性試験代替法フルオレセイン漏出試験法 (Fluorescein leakage test method) の提案

平成 24 年 10 月 1 日に東京、国立医薬品食品衛生研究所にて開催された新規試験法評価会議（通称：JaCVAM 評価会議）において以下の提案がなされた。

提案内容：フルオレセイン漏出試験法は、水溶性で眼腐食性・強度眼刺激性の化合物に限定したとき、トップダウン方式の最初の段階で用いることが許される眼刺激性試験代替法である。

この提案書は、European Center for the Validation of Alternative Methods (ECVAM) による第三者評価資料および OECD Test Guideline (TG) 460 の成立時により準備された資料をもとに、眼刺激性試験代替法評価委員会によりまとめられた文書を用いて JaCVAM 評価会議が評価および検討した結果、その有用性が確認されたことから作成された。

以上の理由により、行政当局の安全性評価方法として「フルオレセイン漏出試験法」の使用を提案するものである。

吉田武美



JaCVAM 評価会議 議長

西川秋佳



JaCVAM 運営委員会 委員長

JaCVAM 評価会議

吉田武美（日本毒性学会）：座長
浅野哲秀（日本環境変異原学会）
五十嵐良明（国立医薬品食品衛生研究所 生活衛生化学部）
大島健幸（日本化学工業協会）
小笠原弘道（独立行政法人 医薬品医療機器総合機構）
小野寺博志（独立行政法人 医薬品医療機器総合機構）
黒澤 努（日本動物実験代替法学会）
杉山真理子（日本化粧品工業連合会）
西川秋佳（国立医薬品食品衛生研究所 安全性生物試験研究センター）
長谷川隆一（独立行政法人 製品評価技術基盤機構）
牧 栄二（日本免疫毒性学会）
増田光輝（座長推薦）
横関博雄（日本皮膚アレルギー・接触皮膚炎学会）
吉田 緑（国立医薬品食品衛生研究所 安全性生物試験研究センター 病理部）
吉村 功（座長推薦）
渡部一人（日本製薬工業協会）

任期：平成 24 年 4 月 1 日～平成 26 年 3 月 31 日

JaCVAM 運営委員会

- 西川秋佳（国立医薬品食品衛生研究所 安全性生物試験研究センター）：委員長
大野泰雄（国立医薬品食品衛生研究所）
小川久美子（国立医薬品食品衛生研究所 安全性生物試験研究センター 病理部）
菅野 純（国立医薬品食品衛生研究所 安全性生物試験研究センター 毒性部）
斎藤和幸（独立行政法人 医薬品医療機器総合機構）
佐々木正広（厚生労働省 医薬食品局 化学物質安全対策室）
関野祐子（国立医薬品食品衛生研究所 安全性生物試験研究センター 薬理部）
高木篤也（国立医薬品食品衛生研究所 安全性生物試験研究センター 毒性部 動物
管理室）
長谷部和久（厚生労働省 医薬食品局 化学物質安全対策室）
広瀬明彦（国立医薬品食品衛生研究所 安全性生物試験研究センター 総合評価研究
室）
本間正充（国立医薬品食品衛生研究所 安全性生物試験研究センター 変異遺伝部）
光岡俊成（厚生労働省 医薬食品局 審査管理課）
小島 肇（国立医薬品食品衛生研究所 安全性生物試験研究センター 薬理部 新規
試験表評価室）：事務局

**JaCVAM statement
on the Fluorescein Leakage Test Method
for Identifying Ocular Corrosives and Severe Irritants**

At the meeting concerning the above method, held on 1 October 2012 at the National Institute of Health Sciences (NIHS), Tokyo, Japan, the members of the Japanese Center for the Validation of Alternative Methods (JaCVAM) Regulatory Acceptance Board unanimously endorsed the following statement:

The Fluorescein Leakage test method is acceptable as an initial step within a Top-Down Approach to identify ocular corrosives and severe irritants from all other classes for water-soluble chemicals.

Following the review of the results of the ESAC (ECVAM Scientific Advisory Committee) Statement on the Scientific Validity of Cytotoxicity/Cell Function Based *in vitro* Assays for Eye Irritation Testing and OECD (Organisation for Economic Co-operation and Development) Test Guideline No. 460, it is concluded that the Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants is clearly beneficial.

The JaCVAM Regulatory Acceptance Board has been regularly kept informed of the progress of the study, and this endorsement is based on an assessment of various documents, including, in particular, the evaluation report prepared by the JaCVAM ad hoc peer review panel for eye irritation testing.



Takemi Yoshida
Chairperson
JaCVAM Regulatory Acceptance Board



Akiyoshi Nishikawa
Chairperson
JaCVAM Steering Committee

20 January, 2013

The JaCVAM Regulatory Acceptance Board was established by the JaCVAM Steering Committee, and is composed of nominees from the industry and academia.

Mr. Takemi Yoshida (Japanese Society of Toxicology): Chairperson
Mr. Norihide Asano (Japanese Environmental Mutagen Society)
Mr. Yoshiaki Ikarashi (National Institute of Health Sciences: NIHS)
Mr. Takeyuki Oshima (Japan Chemical Industry Association)
Mr. Hiromichi Ogasawara (Pharmaceuticals and Medical Devices Agency)
Mr. Hiroshi Onodera (Pharmaceuticals and Medical Devices Agency)
Mr. Tsutomu Miki Kurosawa (Japanese Society for Animal Experimentation)
Ms. Mariko Sugiyama (Japan Cosmetic Industry Association)
Mr. Akiyoshi Nishikawa (Biological Safety Research Center: BSRC, NIHS)
Mr. Ryuichi Hasegawa (National Institute of Technology and Evaluation)
Mr. Eiji Maki (Japanese Society of Immunotoxicology)
Mr. Mitsuteru Masuda (nominee by Chairperson)
Mr. Hiroo Yokozeki (Japanese Society for Dermatoallergology and Contact Dermatitis)
Ms. Midori Yoshida (BSRC, NIHS)
Mr. Isao Yoshimura (nominee by Chairperson)
Mr. Kazuto Watanabe (Japan Pharmaceutical Manufacturers Association)

This statement was endorsed by the following members of the JaCVAM steering Committee after receiving the report from JaCVAM Regulatory Acceptance Board:

Mr. Akiyoshi Nishikawa (BSRC, NIHS): Chairperson
Mr. Yasuo Ohno (NIHS)
Ms. Kumiko Ogawa (Division of Pathology, BSRC, NIHS)
Mr. Jun Kanno (Division of Cellular and Molecular Toxicology, BSRC, NIHS)
Mr. Kazuyuki Saito (Pharmaceutical & Medical Devices Agency)
Mr. Masahiro Sasaki (Ministry of Health, Labour and Welfare)
Ms. Yuko Sekino (Division of Pharmacology, BSRC, NIHS)
Mr. Atsuya Takagi (Animal Management Section of the Division of Cellular and Molecular Toxicology, BSRC, NIHS)
Mr. Kazuhisa Hasebe (Ministry of Health, Labour and Welfare)
Mr. Akihiko Hirose (Division of Risk Assessment, BSRC, NIHS)
Mr. Masamitsu Honma (Division of Genetics and Mutagenesis, BSRC, NIHS)
Mr. Toshinari Mitsuoka (Ministry of Health, Labour and Welfare)
Mr. Hajime Kojima (Section for the Evaluation of Novel Methods, Division of Pharmacology, BSRC, NIHS):Secretary

眼刺激性試験代替法フルオレセイン漏出試験法

目次

評価会議報告書	1
第三者評価報告書	7
STATEMENT ON THE SCIENTIFIC VALIDITY OF CYTOTOXICITY/CELL FUNCTION BASED IN VITRO ASSAYS FOR EYE IRRITATION TESTING	17
OECD GUIDELINE FOR THE TESTING OF CHEMICALS No.460, Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants.....	23

眼刺激性評価のためのフルオレセイン漏出試験法の評価会議報告

JaCVAM 評価会議

平成 24 年（2012 年）10 月 1 日

JaCVAM 評価会議

吉田武美（日本毒性学会）：座長
浅野哲秀（日本環境変異原学会）
五十嵐良明（国立医薬品食品衛生研究所 生活衛生化学部）
大島健幸（日本化学工業協会）
小笠原弘道（独立行政法人 医薬品医療機器総合機構）
小野寺博志（独立行政法人 医薬品医療機器総合機構）
黒澤 努（日本動物実験代替法学会）
杉山真理子（日本化粧品工業連合会）
西川秋佳（国立医薬品食品衛生研究所 安全性生物試験研究センター）
長谷川隆一（独立行政法人 製品評価技術基盤機構）
牧 栄二（日本免疫毒性学会）
増田光輝（座長推薦）
横関博雄（日本皮膚アレルギー・接触皮膚炎学会）
吉田 緑（国立医薬品食品衛生研究所 安全性生物試験研究センター 病理部）
吉村 功（座長推薦）
渡部一人（日本製薬工業協会）

任期：平成 24 年 4 月 1 日～平成 26 年 3 月 31 日

以上

本報告の評価対象は、OECD の Test Guideline 460 (TG 460) で説明されている眼刺激性評価のための Fluorescein leakage test method である^{1,2)}。

以下ではこの試験法を当該試験法と記すが、具体名が必要な場合は、「フルオレセイン漏出試験法」あるいは「FL 試験法」と記すことにする。

当該試験法は、単層の細胞層を通過する sodium fluorescein (Na-FL) 量を測ることで、被験物質による強度の眼刺激性と眼腐食性を代替的に評価するものである³⁻⁵⁾。今回、眼刺激性試験評価委員会からの報告²⁾を受け、以下の 10 項目について評価したので報告する。

< 審議内容 >

1. 当該試験法は、どのような従来試験法を代替するものか。または、どのような毒性を評価あるいは予測するものか。

当該試験法は、ウサギを用いたドレイズ眼刺激性試験法で評価されてきた毒性の一部である、強度眼刺激性と眼腐食性を検出するための代替法である。

TG 460 では、当該試験法で陽性とされた被験物質は、強度眼刺激性あるいは眼腐食性を示すが、当該試験法での陰性結果は、被験物質が強度眼刺激性あるいは眼腐食性物質でないことを示すものではないとしている。

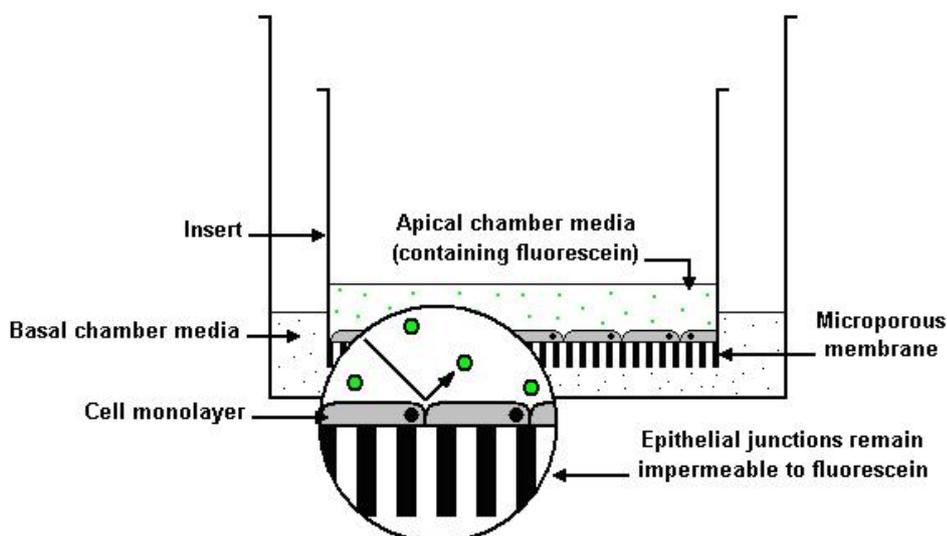
当該試験法は、トップダウン方式（末尾の用語集参照）による強度眼刺激性あるいは眼腐食性を検出する試験法である⁶⁾。

2. 当該試験法と従来試験法の間どのような科学的なつながりがあるか。

当該試験法は、インサート(別図参照)の内側に単層の細胞層を構築し、細胞層の上部に被験物質を曝露した後、細胞層の上部からフルオレセインを添加する。このフルオレセインの細胞層下部への漏出を測定することによって、被験物質による細胞への傷害や細胞間結合の脆弱化を評価するものである。

これに対してウサギを用いたドレイズ試験法は、被験物質の角膜、結膜、虹彩への傷害を、角膜 80 点、結膜 20 点、虹彩 10 点、最大が 110 点のスコアで評価している。

すなわち、ドレイズ試験法は、細胞間結合や細胞自身への傷害以外の損傷も評価しているという点で、当該試験法より多様な損傷を評価している。



3. 当該試験法とそのデータは、透明で独立な科学的評価を受けているか。

当該試験法の妥当性を示すデータは、J. Gartlon と R. Clothier によって background review document (BRD) としてまとめられており、そのデータの妥当性は、ECVAM によって透明で独立した評価を受けている。

4. 当該試験法は、従来試験法の代替法として、どのような物質又は製品を評価することを目的としているか。

当該試験法は、医薬品、医療機器、化粧品、医薬部外品、農薬など、ドレイズ眼刺激性試験法が適用される物質または製品の眼刺激性・眼腐食性を評価することを目的としている。

5. 当該試験法は、ハザード評価あるいはリスク評価のどちらに有用であるか。

当該試験法は、強度眼刺激性または眼腐食性を検出するものであるから、ハザード評価に有用であるが、リスク評価には利用できない。

6. 当該試験法は、目的とする物質又は製品の毒性を評価できるか。その場合、当該試験法の適用条件が明確になっているか。

当該試験法は、強度眼刺激性あるいは眼腐食性をトップダウン方式で検出する際に、最初の段階で実施される試験法である。

当該試験法で陽性の場合、毒性があると判断できるが、陰性の場合には、毒性がないと判断することはできない。陰性の場合に、その毒性がないことを確かめるには、他の試験法でさらに検討をすることが必要である。

適用可能な物質は水溶性物質に限られている。強酸、強塩基、組織固定液、強揮発性物質は適用外である。

7. 当該試験法はプロトコルの微細な変更に対して頑健であるか。

以下のことから、当該試験法は、プロトコルの微細な変更に対して頑健と考えられる。

- 1) 培養細胞として広く使用されている MDCK 細胞株を用いている。
- 2) 単層の細胞層が適切に構成されていることがフルオレセイン漏出の有無で容易に確認できる。
- 3) TG460 では特別な条件設定は見当らず、BRD にも特段の注意がない。

8. 当該試験法の技術習得は、適切な訓練と経験を経ている担当者にとって容易なものであるか。試験法の実施に特殊な設備が必要か。

単層の細胞層の構築以外は特殊技能が不要であり、特殊設備も不要である。

単層の細胞層の構築についても、TG 460 に詳細な説明があるので、適切な訓練と経験を経ている担当者であれば技術習得が容易である。

9. 当該試験法は、従来試験法と比べて時間的経費的に優れているか。

当該試験法は、ドレイズ眼刺激性試験法に比べて、時間的経費的に優れている。

10. 当該試験法は、動物福祉の観点及び科学的見地から、目的とする物質又は製品の毒性を評価する代替法として、行政上利用することは可能か。

当該試験法は、動物を用いずに強度眼刺激性・眼腐食性を評価できるので、動物福祉の観点から有用である。

目的とする物質又は製品の毒性を評価する代替法として、科学的見地から行政上利用することが可能である。

しかし、トップダウン方式でのみ利用可能であること、水溶性物質のみが適用可能であること、からその有用性は限定的である。

参考文献

- 1) 吉村功, 山本直樹: FL 試験法 (Fluorescein leakage test method ; フルオレセイン漏出試験法) の概要, JaCVAM 評価会議 (4, June, 2012)
- 2) Proposal for a New Test Guideline 460: Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants.
- 3) Gartlon, J., Clothier, R.: Fluorescein Leakage Assay Background Review Document as an Alternative Method for Eye Irritation Testing. Available under Validation Study Documents, Section Eye Irritation at [<http://ecvam.jrc.it>]
- 4) Gartlon, J., Clothier, R.: Fluorescein Leakage Assay Background Review Document as an Alternative Method for Eye Irritation Testing: Appendices and Annexes. Available under Validation Study Documents, Section Eye Irritation at [<http://ecvam.jrc.it>]
- 5) Fluorescein Leakage (FL) Test DB-ALM Protocol no.71. Available under Validation Study Documents, Section Eye Irritation at [<http://ecvam.jrc.it>]
- 6) ECVAM Retrospective Validation Study on Cytotoxicity/Cell-Function Based in vitro Assays for the prediction of Eye Irritation: ESAC Peer Review. Available under Validation Study Documents, Section Eye Irritation at [<http://ecvam.jrc.it>]

用語集

トップダウン方式

一つの被験物質に対して複数の試験法を逐次適用する試験方式の一つ。前段階の試験法で毒性があると判定された場合は、それで試験を終了し、毒性が無いと判断された場合は、次段階の試験法を適用して、無毒性が真であるかどうかを確認するという毒性判定法。

眼刺激性試験代替法フルオレセイン漏出試験法
(Fluorescein leakage test method ; FL 試験法) の
第三者評価報告書

2012年6月

眼刺激性試験代替法評価委員会

眼刺激性試験代替法評価委員会

吉村 功（東京理科大学）

山本 直樹（藤田保健衛生大学）

小坂 忠司（残留農薬研究所）

竹内 小苗（P&G）

細井 一弘（参天製薬株式会社）

加藤 雅一（株式会社 J-TEC）

簾内 桃子（国立医薬品食品衛生研究所）

増田 光輝（国立医薬品食品衛生研究所）

まえおき

OECD は公に認めた毒性試験法について、試験ガイドライン (Test Guideline: TG) を作成し公開している。TG 編纂の主体は Working Group of National Coordinators of the Test Guideline Programme (WNT) である。

FL 試験法にはいくつかのプロトコルがある。WNT が焦点を当てて検討したのは、階層的試験方針で眼腐食性・強度眼刺激性を確認するトップダウン方式の INVITTOX Protocol No. 71 (プロトコル 71) と、階層的試験方式で眼刺激性が無いことを確認するボトムアップ方式の INVITTOX Protocol No. 120 (プロトコル 120) である。2011 年 4 月 12~14 日にパリの OECD 本部で行われた WNT 第 23 回会合に提出・検討されたのは、トップダウン方式のプロトコル 71 の TG 案 (Draft OECD Guideline for the Testing of Chemicals : Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants) であった。この TG 案は、同年 11 月 8 日の会合で修正され、12 月 5 日締め切りで意見公募が行われ、2012 年 4 月に改訂が検討され、6 月に最終案が提出された。

本報告では、主としてこの TG 案に基づいたフルオレセイン漏出試験法 (FL 試験法) の概要を紹介し、評価を行う。

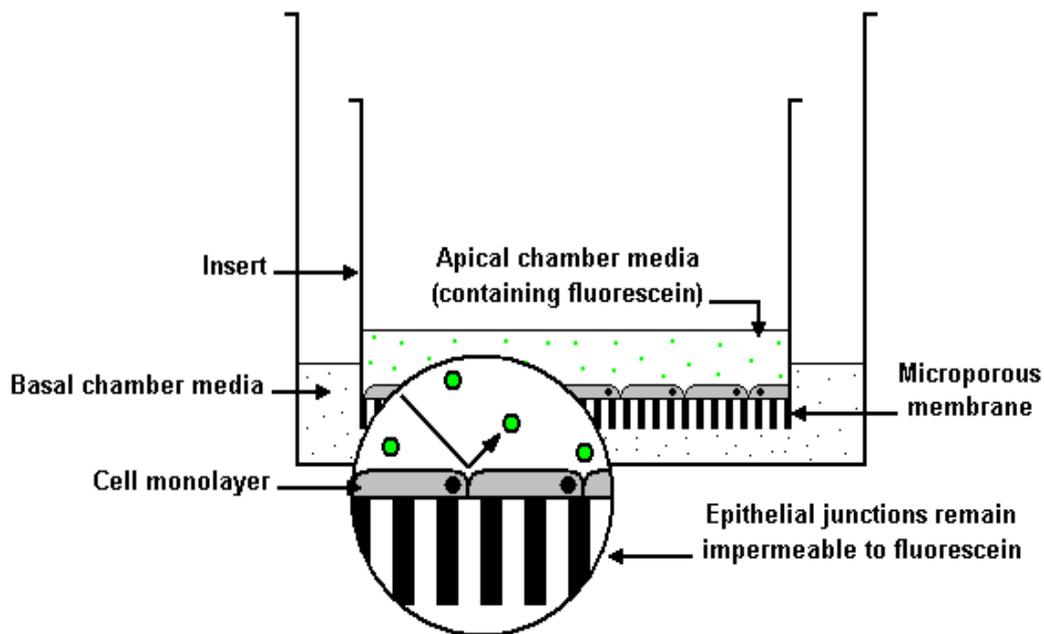
1. 試験法の位置づけ

FL 試験法は、眼腐食性・強度眼刺激性物質、すなわち UN-GHS, EU-CLP, U.S.-EPA でカテゴリー I の化合物・複合物 (以下、化学物質) を検出するためのインビトロ試験法である。強度眼刺激性物質とは、それに触れることで、21 日経っても回復しない眼組織傷害、あるいは、強度の視力損傷、を引き起こすものである。眼腐食性物質とは、眼に回復不能な組織傷害を与えるものである。上記 TG 案では、FL 試験法を、眼腐食性・強度眼刺激性を持つ水溶性物質をトップダウン方式で検出する際に、最初に用いる試験法としている。

2. 試験法の原理

いろいろな物質が眼の中に入るのを阻止するのは、角膜と結膜の重要な役割である。これは細胞間結合で制御されている。FL 試験法は、その傷害を測るために、透過性インサートの薄膜上に Madin-Darby Canine Kidney (MDCK) 細胞、すなわち MDCK-CB997 尿細管上皮細胞を単層培養して、コンフルエント状態になっているモデルを用いる。試験では、フルオレセインナトリウム (sodium fluorescein; Na-F) が単層培養した細胞の間をどれだけ通過するかを測って、短時間の被験物質曝露による毒性発現を評価する。この Na-F 漏出量は、化学的に引き起こされた細胞間結合の損傷に比例するから、これで被験物質の眼刺激性が評価できる。

インサート薄膜上の MDCK 細胞の状態は下図の通りである。単層培養でコンフルエント状態になった MDCK はインサート内の半透過性薄膜上に生成されていて、インサートは 24 穴プレートの各ウェルに置かれている。



試験では、被験物質をインサートの薄膜上に形成された単層培養でコンフルエント状態の細胞に滴下する。滴下 1 分後に被験物質を払拭し、明るい蛍光を発する無毒性染料の Na-F を 30 分間、単層細胞の上に適用する。被験物質による細胞間結合の損傷は、この時間の間に単層細胞とインサートの薄膜を通過するフルオレセイン量 (fluorescein leakage; FL) で同定される。

単層細胞とインサートの薄膜を通過してウェル内の一定量溶液 (Basal chamber media) に届いた Na-F 量は、ウェル内のフルオレセイン濃度を分光蛍光光度計で測定することで得られる。すなわち、FL は、ブランク対照と最大通過対照 (maximum leakage control) におけるフルオレセイン強度値 (fluorescein intensity; FI) を参照して計算できる。被験物質の用量ごとに、対照との比較で、漏出率 (%), すなわち細胞間結合の損傷を測定する。無処理の単層培養でコンフルエント状態のインサートと無細胞のインサートにおける値と対比して 20% の FL が得られる濃度 FL_{20} (mg/mL) を計算する。この FL_{20} 値を眼腐食・強度眼刺激の確認に用いる。

3. 試験手順と結果の判定

3.1 細胞単層の作成

MDCK-CB997 細胞の単層培養は、DMEM/Nutrient Mix F12 を入れた細胞培養フラスコ内で培養されたサブコンフルエント状態の細胞を使って作成する。細胞間結合を完全にするために、FL 試験法での溶媒/溶液では、カルシウム濃度を常に 1.8mM (200mg/L) から 1.0mM (111mg/L) の間に収めておくことが重要である。

均一かつ再現可能な細胞間結合を生成するには、継代数をある範囲にしておくべきである。試験結果の再現性を確保するために、できれば、その継代数の範囲は解凍後

3～30 継代にするべきである。この範囲であれば細胞が類似した機能性を保つからである。

FL 試験法の実施に先立って、トリプシンを適用してフラスコから細胞を外し、解凍し、一定量を 24 穴プレート内のインサートに播種する。細胞を播種するインサートとしては、厚み 80 ～150 μm 、孔サイズ (pore size) 0.45 μm の混合セルロースエステル薄膜を持つ直径 12 mm のものを用いるべきである。バリデーションでは、Millicell-HA 12mm インサートが用いられた。インサートと膜タイプの性質は、細胞成長と化学結合に影響するので、他のインサートを用いるときは、§ 7.2 に示す熟達度確認用化合物 (proficiency chemicals) を使って同等性を確かめるべきである。

ある種の化学物質は Millicell-HA インサートの薄膜と結合して試験結果を説明しがたいものにする。たとえば塩化ベンザルコニウムなどのカチオンは、荷電薄膜と化学結合する傾向がある。インサート薄膜との化学結合は、化学的な曝露時間を増やして化学物質の毒性を過大に見せかける。しかし同時に、インサートの薄膜へのカチオンの化学結合が FL を物理的に減らし、化学物質の毒性を過小に見せかけることもある。いずれも結果の説明を困難にする。

化学結合の発現は、無細胞のインサートの薄膜を最大濃度の被験物質に曝露させた後で、標準時間、標準濃度の Na-F 染色を行うことで監視できる。Na-F 染色が起こったら、被験物質を払拭した後でインサート薄膜が黄色になるからである。細胞に被験物質を適用した結果が適切に説明できるためには、被験物質の化学結合の性質を知っておくことが必須である。

インサート上に播種した細胞は、化学物質曝露の際に、単層のコンフルエント状態を作っていないなければならない。一つのインサートに 1.6×10^5 個の細胞があるように、細胞には、濃度 4×10^5 cells/mL の懸濁液 400 μL を加えなければならない。この条件であれば、播種後 96 時間でコンフルエント状態の細胞単層が構成できる。

MDCK 細胞培養は、CO₂ 濃度が $5 \pm 1\%$ 、温度が $37 \pm 1^\circ\text{C}$ で、湿度が保たれている培養器で行わなければならない。細胞がバクテリア、ウイルス、マイコプラズマ、真菌類に汚染されないように注意すべきである。

3.2 被験物質の適用と対照物質

被験物質の新鮮な保存溶液は実験ごとに用意し、30 分以内に使用しなければならない。被験物質は、血清蛋白結合を避けるために、1.0～1.8mM 濃度でカルシウムを含みフェノールレッドを含まない Hanks' Buffered Saline Solution (HBSS) 内に用意しなければならない。また、被験物質が HBSS に 250mg/mL まで溶解することを実験前に確かめておかななければならない。化学物質が、この濃度で、均一で 2 層には分かれていない安定した懸濁液・乳液状態を 30 分以上維持するのであれば、HBSS を溶剤として使うことができる。

化学物質がこの濃度で HBSS に可溶でないならば、FL 試験法とは異なる試験法を使うべきである。HBSS に不溶であるためにミネラルオイル (light mineral oil) を

溶剤とすると、FL 試験法が適切な結果をもたらすかどうか分からないからである。

被験物は、カルシウム (1.0~1.8mM) を含みフェノールレッドを含まない HBSS で 1, 25, 100, 250 mg/mL の 5 用量に調製された溶液と、原液又は飽和溶液である。

固体物質の場合は、750 mg/mL という高濃度も含める。この濃度のときは、ポジティブ・ディスプレイメント・ピペットで細胞に適用してもよい。25 mg/mL と 100mg/mL で毒性が発現したときは、1, 25, 50, 75, 100 mg/mL で 2 回再実験をしなければならない。1 mg/mL で毒性が発現したときは、0.10, 0.1, 0.25, 1, 10 mg/mL で 2 回再実験をしなければならない。

被験物は、培養液を除去した後で、除菌され、37°Cに温められている、カルシウムを含みフェノールレッドを含まない HBSS で 2 回洗浄したコンフルエント状態の細胞単層に適用する。実験では、用意された各濃度が、少なくとも 3 回反復して実験されなければならない。

室温で 1 分間曝露した後、注意して被験物を吸引除去し、滅菌して 37°Cに温められた、カルシウムを含みフェノールレッドを含まない HBSS を用いてコンフルエント状態の細胞単層を 2 回洗浄し、すぐ FL を測定する。

インサート上の単層培養された細胞の整合性と、実験での細胞の感度が過去の実績範囲内にあることを確かめるために、各実験実行 (run) には陰性対照 (NC) と陽性対照 (PC) を含めなければならない。

陽性対照としては、Brij 35 (CAS No. 9002-92-0) の 100 mg/mL の使用が勧められている。この濃度は、ほぼ 30% (20%~40%でよい) の FL をもたらすはずである。

陰性対照としては、カルシウムを含みフェノールレッドを含まない HBSS が勧められている。FL₂₀ を計算するには、最大漏出対照も各実験実行に含めることが必要である。最大漏出は無細胞インサート対照を用いて求める。

3.3 フルオレセイン通過性の検量

被験物質と対照物質を除去したら、直ちに十分な量の 0.1% (w/v) Na-F 液を Millicell-HA インサートに添加し、細胞を室温に 30 分間置いておく。フルオレセインを添加した培養の最後に、注意して各ウェルからインサートを取り除く。各フィルタを視認で検査し操作中に生じた損傷があったら記録する。

細胞単層とインサートを通過したフルオレセイン量はインサートを取り除いた後のウェルに残っている溶液で定量する。測定は、波長 485 nm と 530 nm の励起・放出波長の各々で、分光蛍光光度計を用いて行う。分光蛍光光度計の感度は最大 FL 値 (無細胞インサート) と最小 FL 値 (陰性対照) の差の最大値が含まれるようにしておかなければならない。使用する分光蛍光光度計の違いが影響しないように、最大漏出対照に対して、FI が 4000 を超えるように感度を定めておく。ただし、最大 FL 値は 9999 を超えないようにする。

3.4 結果の説明と予測モデル

被験物質の各濃度での相対的な FL 値，すなわち %FL 値は，各実験実行における陰性対照での FI 値と，最大漏出対照での FI 値を参照して，被験物質の FI 値から次のようにして計算する。

すなわち，反復実行に対する，最大漏出 FI の平均値を x とし，陰性対照での FI の平均値を y とすると，100%漏出の平均値 z は， $z = x - y$ となる。各用量で，単層を通過する FL の相対値，%FL は， $\%FL = [(m - y) / z] \times 100\%$ で計算する。ここで m は，各濃度での FI の 3 反復測定 of 平均値である。

20%漏出に対応する用量 FL_{20} は，用量反応曲線の直線補間で求める。すなわち，20%より小さい %FL 値 B とそれをもたらしている用量 M_B ，及び 20%より大きい %FL 値 C とそれをもたらしている用量 M_C を調べ，次式で計算する。

$$FL_{20} = [(20 - B) / (C - B)] \times (M_C - M_B) + M_B$$

眼腐食性・強度眼刺激性の予測においては， $FL_{20} \leq 100 \text{ mg/mL}$ のとき，被験物質をカテゴリー I と判定する。すなわち，臨界値 (cut-off value) は 100 である。

3.5 結果の承認

最大漏出値 (x) は 4000 以上，0%漏出値 (y) は 300 以下，100%漏出値 (z) は 3700 と 6000 の間でなければならない。陽性対照の %FL が 20%から 40%の間であれば試験結果は承認できる。

4. 試験法の正確性

文献 1 によれば，水溶性で眼腐食性・強度眼刺激性の化学物質の偽陽性率は 7% (GHS と CLP で 7/103) から 9% (EPA で 9/99)，偽陰性率は 54% (EPA で 15/28) から 56% (GHS と CLP で 27/48) である。括弧内の比率表示における分母は，プロトコル 71 に基づくバリデーション研究に用いられた被験物質数，分子は判定が誤りであった被験物質数である。これより，バリデーション研究で得られた予測の正確度 (concordance rate) は，77.5% (GHS と CLP で 117/151) から 82.7% (EPA で 105/127) となる。

眼腐食性・強度眼刺激性を確かめるといふ目的に限定すれば，注目すべき値は，正確度や偽陰性率ではなく偽陽性率である。正確度が大きくないことや，偽陰性率が大きいことは重大な弱点でない。階層的試験方式と重み付け評価方式を用いる場合，後に続くインビトロ代替法で，FL 試験法で見逃された偽陰性物質が確認できるからである。

注：上記の数値は，文献 2, 3 に基づいて得られたものの筈である。ところがこれらの文献にあるデータ (文献 2 の Table 6.2.4.2.3) は，次のものである。EPA での

偽陽性率に「9% (9/99)」と「8.3% (9/109)」という食い違いがある。本評価委員会はまだその理由を確かめることができないでいる。文献 2 には他にも集計ミスがあるので、文献 1 の数値が正しいと感じられる。

分類システム	正確度	偽陽性率	偽陰性率
EU			
	77.9% (113/145)	7.1% (7/99)	54.3% (25/46)
GHS			
	77.5% (117/151)	6.8% (7/103)	56.3% (27/48)
EPA			
	82.5% (113/137)	8.3% (9/109)	53.6% (15/28)

上の表の数値は、次に示す 3 分類表（文献 2 の Tables 6.2.4.2.1 a, b, c）を後付け的に 2 分類に縮約したものである。後付けの宿命として、過大評価になっている可能性を認識しておくべきであろう。

判定	EU 分類				判定	GHS 分類				判定	EPA 分類		
	NC	R36	R41			NC	Cat2	Cat1			III/IV	II	I
NC	41	19	21		NC	36	24	14		III/IV	36	16	10
R36	12	20	13		Cat2	4	32	13		II	23	15	5
R41	4	3	21		Cat1	3	4	21		I	3	6	13
計	57	42	46		計	43	60	48		計	62	37	28

どのような化学物質で偽陽性が出やすいかということについては、データが見あたらない。

5. 試験結果の再現性

文献 2, 3 では、施設内再現性、技術移転性、施設間再現性を、過去の文献で得られるデータで多種多様な観点から検討をしている。しかし検討しているデータは、プロトコルの違いがあったり、検討対象の被験物質が界面活性剤に偏っているというような特性があったり、技術移転性と施設間再現性が交絡していたりして、総体としての結論が明らかでない。ここでは、プロトコル 71 に適用できると思われる結果を参考として示すだけにする。

5.1 施設内再現性

施設内で、単純反復の生データ (raw data) が入手できて、実験作業や反復について施設内再現性が検討できる場合について調べた結果では、FL₂₀ (mg/mL) の変動係数 (coefficient variation) の中央値及び平均値は、56.5%~63.2%であった。

5.2 技術移転性

プロトコル 71 に関して技術移転性を検討できるデータは、文献 4 の 60 化合物 4 施設の実験データである。そこでは FL₂₀ についての実験間ピアソン相関係数（計量値の相関係数）が報告されている。報告されている値は 0.167～0.778 であるが、相関係数の値が低いのは特定のある施設と他の施設の間だけである。問題の施設では、プロトコルが忠実に守られていなかったようである。再現性のある結果を得るには、プロトコルの曖昧さを減らすと同時に、プロトコルを忠実に守るトレーニングが必須である。細胞を処理する技術はそれほど困難なものではないからである。

5.3 施設間再現性

前記の文献 4 では FL₂₀ についての施設間の結果の乖離をピアソン相関係数で評価している。得られている値は 0.214～0.841 である。これについても前項で述べたことと同じで、特定のある施設が低い相関係数の源になっている。これをもって施設間再現性が悪いというのか、プロトコルが悪いというのか、それとも技術移転性が悪いというのか、結論は明らかでない。

6. 試験法の適用範囲

この試験法の適用対象は水溶性化合物のみである。文献 1 では、水溶性で、希釈によって毒性が変わらない強度眼刺激性物質は、この試験法で正確に確かめることができるとしている。

強酸・強塩基・定着薬・強揮発性物質はこの試験法の適用範囲外である。これらの化学物質は FL 試験法で評価できない仕組み、例えば、広範な凝固、鹼化・加水分解、ある種の特殊な化学反応等、を伴うからである。

着色性や粘着性のある被験物質も予測性を侵すので適用が妥当でない。これらのタイプの物質は、短時間曝露後に細胞単層から取り除くのが困難である。

液体中に懸濁している固体は、急速に沈殿する傾向があるので適用濃度を正確に定めることが困難である。化学的あるいは物理的にこれらの性質を持つ物質を評価から除外すると、EU, EPA, GHS のどのクラス分けシステムで評価しても、FL 試験法の性能は非常によい。

7. その他の注意

7.1 適切な技術移転が達成できたことを確かめるには、表 1 の熟達度確認用化合物を使う。

表1 熟達度確認用化合物

化合物	CAS 番号	化合物分類	物理性状	トリス試験の結果	FL 試験の結果
#1	8001-54-5	オニウム 化合物	液体	カテゴリー1	腐食性/強刺激性
#2	58-33-3	アミン/アミジン...	固体	カテゴリー1	腐食性/強刺激性
#3	1310-73-2	アルカリ	液体	カテゴリー1	腐食性/強刺激性
#4	151-21-3	カルボキシル酸 (塩)	液体	カテゴリー1	腐食性/強刺激性
#5	619-66-9	カルボキシル酸, ...	固体	カテゴリー2(A)	非腐食性/強刺激性
#6	6484-52-2	無機塩	固体	カテゴリー2(A)	非腐食性/強刺激性
#7	609-14-3	ケトン, エステル	液体	カテゴリー2(B)	非腐食性/強刺激性
#8	56-81-5	アルコール	液体	カテゴリー無	非腐食性/強刺激性

#1: Benzalkonium chloride (5%), #2: Promethazine hydrochloride, #3: Sodium hydroxide (10%),
 #4: Sodium lauryl sulfate (15%), #5: 4-carboxy-benzaldehyde, #6: Ammonium nitrate,
 #7: Ethyl-2-methylaceto-acetate, #8: Glycerol

CAS 番号 : Chemical Abstract Service Registry Number

7.2 実験での注意点

MDCK 細胞の培養には特別な技術的制約がある。Na-F が単層培養でコンフルエント状態になった細胞間の通過を阻止する細胞間結合は、培養継代数が増えると弱まる傾向がある。不完全に形成された細胞間接合は無処置対照で FL を増加させる。細胞は時間経過と共に変異するので、実験では、許容できる培養継代数を定めて置くことが必要である。

8. FL 試験法の概要の本委員会としての結論

日本の施設で確認した実験データがないので外国文献上のデータを信用して評価すると、FL 試験法の偽陽性率は、水溶性で眼腐食性・強度眼刺激性の化合物に限定したとき、7% (GHS と CLP で 7/103) から 9% (EPA) である。FL 試験法はこの性能で十分と思われる目的に対して、トップダウン方式の最初の段階で用いることが許される試験法である。

トップダウン方式で用いるのであれば、偽陽性率は小さい方が望ましい。陽性であるという判定の臨界値を小さくすることで、偽陽性率を小さくすることは今後の検討課題である。

本委員会は、FL 試験法の日本語訳として、フルオレセイン漏出試験法 (FL 試験法と略記) を提案する。

9. 文献

1. OECD WNT, 2011. Revised draft OECD Guideline for the Testing of Chemicals: Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants.
2. Gartlon, J., Clothier, R., 2008. Fluorescein Leakage Assay Background Review Document as an Alternative Method for Eye Irritation Testing.
3. Gartlon, J., Clothier, R., Fluorescein Leakage Assay Background Review Document as an Alternative Method for Eye Irritation Testing: Appendices and Annexes.
4. Balls, M., Botham, P.A., Bruner, L.H., Spielmann, H., 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test. *Toxicology In Vitro*, 9(6): 871-929.

10. 用語

UN-GHS : Globally Harmonized System of Classification and Labeling of Chemicals by the United Nation (国連による化学物質の分類とラベル付けの世界統一システム) の省略形である。これは、有害作用情報を伝えることで、雇用者、労働者、輸送者、消費者、事故被害者等の人々と環境を護ろうとして、図表示・用語警告・危険声明・警戒宣言・安全情報シートなどの伝達手段を用いて、物理的・健康的・環境的危険の、標準化されたタイプとレベルに応じて、化学物質をクラス分けするシステムである。このシステムのカテゴリー I (GHS Category I) は、これに属する被験物質を眼の表面に適用したとき、21 日経っても完全には回復しない眼組織損傷、あるいは視力障害をもたらすことで特徴付けられる。

EU-CLP : European Commission Regulation on the Classification, Labelling and Packaging of Substance and Mixtures の省略形である。化学物質分類の UN-GHS システムを欧州に導入したものである。

EPA Category I : 米国環境保護庁 (Environmental Protection Agency) が定めているクラス分けのカテゴリー I のことで、その内容は、21 日以上にわたって、腐食 (眼組織を回復できないように破壊すること)、角膜関連損傷、あるいは刺激をもたらす化学物質である。



1
2 **STATEMENT ON THE SCIENTIFIC VALIDITY OF CYTOTOXICITY/CELL-**
3 **FUNCTION BASED IN VITRO ASSAYS FOR EYE IRRITATION TESTING**
4
5

6 At its 31st meeting, held on 7 and 8 July, 2009 at the European Centre for the Validation of
7 Alternative Methods (ECVAM), Ispra, Italy, the non-Commission members of the ECVAM
8 Scientific Advisory Committee (ESAC)¹ unanimously endorsed the following statement:
9

10 The replacement of traditional animal-based test methods by alternative ones should ideally
11 be obtained by one-to-one replacements: to keep the testing regime simple and economical
12 one single alternative method should, wherever feasible, be sufficient to generate data of
13 equal or better quality than the traditional test.
14

15 However, in the case of eye irritation it is currently generally accepted that, in the foreseeable
16 future, no single *in vitro* eye irritation test will be able to replace the *in vivo* Draize eye test to
17 predict across the full range of irritation for different chemical classes. However, strategic
18 combinations of several alternative test methods within a (tiered) testing strategy may be able
19 to replace the Draize eye test.
20

21 A possible conceptual framework for such a (tiered) testing strategy has been developed
22 within an ECVAM workshop (Ref. 1). The framework is based on alternative eye irritation
23 methods that vary in their capacity to detect either severe irritant substances (EU R41; GHS
24 'Category 1') or substances considered non-irritant (EU 'Non-Classified'; GHS 'No Category').
25 According to this framework the entire range of irritancy may be resolved by arranging tests
26 in a tiered strategy that may be operated from either end: to detect first severe irritants and
27 resolve absence of irritancy ("Top-Down Approach") or to proceed inversely, starting with the
28 identification of non-irritants first ("Bottom-Up Approach"). Mild irritancy will be resolved in
29 a last tier in both approaches.
30

31 To evaluate the scientific validity of possible building blocks of such a test strategy and to
32 assess their possible placement within a Bottom-Up and Top-Down Approach, ECVAM has
33 undertaken a retrospective validation study of four cell-based *in vitro* methods.
34

35 The test methods evaluated were:

- 36
- 37 a. Cytosensor Microphysiometer (INVITTOX Protocols 97 and 102 modified)²
 - 38 b. Fluorescein Leakage (INVITTOX Protocols 71, 82, 86 and 120);
 - 39 c. Neutral Red Release (INVITTOX Protocol 54 and PREDISAFETM);
 - 40 d. Red Blood Cell haemolysis (INVITTOX Protocols 37 and 99),
41

42 The four test methods, including ten protocol variations, were subjected to independent,
43 expert review with respect to their use to either

¹ Details can be found in the PRP report

² Invitox protocols can be downloaded from ECVAM's database service on Alternative Methods to Animal Experimentation, DBALM: <http://ecvam-dbalm.jrc.ec.europa.eu>



- 44 a) initiate a Bottom-Up Approach, for consideration for regulatory use to identify non-
45 irritants (EU: 'Non Classified'; GSH: 'No Category'; EPA: 'Category IV') from all other
46 classes as part of a tiered testing strategy, or
47 b) to initiate a Top-Down Approach, to identify ocular corrosives and severe irritants
48 (EU R41, GHS 'Category 1', and EPA 'Category I') from all other classes as part of a
49 tiered testing strategy.

50 In the absence of internationally agreed performance criteria for either approach, the PRP of
51 the ESAC applied the following criteria:

- 52 • any test used to initiate a Top-Down Approach must balance specificity and sensitivity
53 to correctly identify a substantial proportion of severe irritants, with a false positive
54 rate that would not lead to the over-classification of an unreasonable number of
55 materials of lower ocular irritancy potential – an over-classification rate (false
56 positives) of <10% was considered acceptable
57 • any test used to initiate a Bottom-Up Approach should ideally give no false negatives
58 with respect to human safety, and no false negative should be produced by high-
59 moderate or severe irritants.

60
61 Following independent ESAC peer review of this retrospective validation study and
62 considering the potential test strategies in which the tests may be used, the ESAC concluded
63 the following:
64

65 1. CYTOSENSOR MICROPHYSIOMETER TEST METHOD

66
67 The Cytosensor Microphysiometer test method can be used for two of the three EU and GHS
68 classification categories used for the endpoint of ocular irritation:
69

70 A. The **Cytosensor Microphysiometer test method (INVITTOX Protocol 102 modified)** is
71 considered to have been scientifically validated and to be ready for consideration for
72 regulatory use as an initial step within a **Top-Down Approach** to identify ocular corrosives
73 and severe irritants (EU R41, GHS Category 1, and EPA Category I) from all other classes for
74 the chemical applicability domain of water-soluble chemicals (substances and mixtures).
75

76 B. Furthermore, the **Cytosensor Microphysiometer test method (INVITTOX Protocol 102**
77 **modified)** is considered to have been scientifically validated and to be ready for consideration
78 for regulatory use as an initial step within a **Bottom-Up Approach** to identify non-irritants
79 (EU:NC; GHS: NC; EPA: cat IV) from all other classes only for water-soluble surfactants and
80 water-soluble surfactant-containing mixtures.
81

82 C. On the basis of a thorough evaluation of the data compiled in the course of the ECVAM
83 validation study, the ESAC concludes that the **Cytosensor Microphysiometer** test method
84 does NOT correctly identify moderate and mild ocular irritants (EU: R36; GHS: Cat 2A/B;
85 EPA: Cat II/III). Therefore, the test method can only be employed to make decisions on two
86 of the three categories of the eye irritation classification scheme (see A and B). Consequently,
87 ESAC does NOT recommend this test method as a full replacement method. It should be
88 noted in this context that the **Top-Down and Bottom-Up Approach** foresees the theoretical
89 possibility of a *default* mild/moderate categorization (e.g. EU R36 or GHS Cat 2) of all those
90 substances neither identified as ocular corrosives and severe irritants (see A) nor as "non-



91 classified" substances (see B) in the first two tiers of the strategy. However, the test method's
92 high false negative rate (9-55%) when initiating a top-down approach and high false positive
93 rate (50-69%) when initiating a bottom-up approach exclude the possibility to use the method
94 for default categorization. The test methods can thus not be considered a full-replacement
95 method on its own using the Top-Down and Bottom-Up approach.

96
97 Although these recommendations are based on the evaluation of data sets obtained using
98 specific hard- and software, it is anticipated that other Cytosensor Microphysiometer
99 equipment and software may become available with either equivalent or better performance
100 and will need to be efficiently validated. Depending on the similarity of new equipment with
101 respect to the validated one, this may be performed as a *Similar Method Validation* ('me-too')
102 or an *Update Validation*. ESAC therefore recommends the development of Performance
103 Standards for the Cytosensor Microphysiometer test method.

104
105 The current chemical applicability domain is limited: whilst in some cases this might be
106 increased by expanding the data set of studied compounds, the test method is not amenable to
107 testing non-water soluble solids, suspensions, or viscous materials.

108
109

110 2. FLUORESCHEIN LEAKAGE TEST METHOD

111
112 The **Fluorescein Leakage test method (INVITTOX Protocol 71)** is considered to have been
113 scientifically validated and to be ready for consideration for regulatory use as an initial step
114 within a **Top-Down Approach** to identify ocular corrosives and severe irritants (EU R41,
115 GSH Category 1, and EPA Category I) from all other classes for water-soluble chemicals
116 (substances and mixtures).

117
118 Additional testing and further refinement, in particular with respect to variability and
119 definition of the applicability domain, by expanding the dataset of tested chemicals and direct
120 comparison with *in vivo* data is recommended and should be kept under review.

121
122 With regard to the

- 123 • Neutral Red Release (INVITTOX Protocol 54 and PREDISAFETM);
- 124 • Fluorescein Leakage (INVITTOX Protocols 82, 86 and 120);
- 125 • Red Blood Cell haemolysis (INVITTOX Protocols 37 and 99),

126 ESAC considers that the available evidence is insufficient³ to support a recommendation that
127 they are ready for consideration for regulatory use.

128
129 Similarly, the available evidence for Fluorescein Leakage INVITTOX Protocol 71 does not
130 support a recommendation for its use to initiate a Bottom-Up Approach for regulatory use.

131
132
133
134
135

³ Details can be found in the PRP report



EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
In vitro methods Unit

European Centre for the Validation of Alternative Methods (ECVAM)

136 This statement takes account of the dossiers prepared for peer review; the views of
137 independent experts of the ESAC Peer Review Panel (PRP) who evaluated the dossiers
138 against defined validation criteria as well as supplementary submissions made by the
139 Validation Management Group.

140

141 In agreement with common practice upon completion of a validation study, ESAC
142 recommends the development of Performance Standards for the Cytosensor
143 Microphysiometer and the Fluorescein Leakage assays to allow the validation of *similar test*
144 *methods* or *modifications of the validated test methods* based on pre-defined evaluation and
145 acceptance criteria.

146

147 Joachim Kreysa

148 Head of Unit

149 In vitro methods Unit

150 European Centre for the Validation of Alternative Methods

151

152 Ispra, 10th July 2009



EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
In vitro methods Unit
European Centre for the Validation of Alternative Methods (ECVAM)

153 **REFERENCECS**

154

- 155 1. Scott, L. et al. (2009) A proposed eye irritation testing strategy to reduce and replace
156 in vivo studies using Bottom-Up and Top-Down approaches. *Toxicol In Vitro*. May
157 31. [*Epub ahead of print*]

158



EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
In vitro methods Unit
European Centre for the Validation of Alternative Methods (ECVAM)

159

160 The ESAC was established by the European Commission, and is composed of nominees from
161 the EU Member States, industry, academia and animal welfare organisations, together with
162 representatives of the relevant Commission services.

163

164 This statement was endorsed by the following members of the ESAC:

165

166 Ms Argelia Castaño (Spain)
167 Ms Maija Dambrova (Latvia)
168 Ms Alison Gray (ESTIV)
169 Ms Katalin Horvath (Hungary)
170 Ms Dagmar Jírová (Czech Republic)
171 Mr Roman Kolar (Eurogroup for Animals)
172 Ms Elisabeth Knudsen (Denmark - acting as moderator at the meeting)
173 Mr Manfred Liebsch (Germany)
174 Mr Gianni Dal Negro (EFPIA)
175 Mr. Walter Pfaller (Austria)
176 Mr Tõnu Püssa (Estonia)
177 Mr Dariusz Sladowski (Poland)
178 Mr Jon Richmond (UK)
179 Ms Vera Rogiers (ECOPA)
180 Mr Michael Ryan (Ireland)
181 Ms Annalaura Stamatii (Italy)
182 Mr Jan van der Valk (The Netherlands)
183 Mr Carl Westmoreland (COLIPA)
184 Mr Timo Ylikomi (Finland)

185

186 The following Commission Services and Observer Organisations were involved in the
187 consultation process, but not in the endorsement process itself:

188 **Commission services**

189 Mr Joachim Kreysa (DG JRC, Head of In vitro methods Unit/ECVAM, chairman)
190 Mr Claudius Griesinger (DG JRC, ESAC secretariat)
191 Ms Susanne Hoke (DG ENTR)
192 Ms Susanna Louhimies (DG ENV)
193 Mr Juan Riego Sintés (DG JRC)

194

195 **The following observers were present**

196 Mr Hajime Kojima (JaCVAM)
197 Mr William Stokes (NICEATM)
198 Ms Marilyn Wind (ICCVAM)

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants

INTRODUCTION

1. The Fluorescein Leakage (FL) test method is an *in vitro* test method that can be used under certain circumstances and with specific limitations to classify chemicals (substances and mixtures) as ocular corrosives and severe irritants, as defined by the United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (Category 1), the European Union (EU) Regulation on Classification, Labelling and Packaging of Substances and Mixtures (CLP) (Category 1), and the U.S. Environmental Protection Agency (EPA) (Category I) (1) (2) (3). For the purpose of this Test Guideline, severe irritants are defined as chemicals that cause tissue damage in the eye following test substance administration that is not reversible within 21 days or causes serious physical decay of vision, while ocular corrosives are chemicals that cause irreversible tissue damage to the eye. These chemicals are classified as UN GHS Category 1, EU CLP Category 1, or U.S. EPA Category I.

2. While the FL test method is not considered valid as a complete replacement for the *in vivo* rabbit eye test, the FL is recommended for use as part of a tiered testing strategy for regulatory classification and labelling. Thus, the FL is recommended as an initial step within a Top-Down approach to identify ocular corrosives/severe irritants, specifically for limited types of chemicals (*i.e.* water soluble substances and mixtures) (4)(5).

3. It is currently generally accepted that, in the foreseeable future, no single *in vitro* eye irritation test will be able to replace the *in vivo* eye test (TG 405 (6)) to predict across the full range of irritation for different chemical classes. However, strategic combinations of several alternative test methods within a (tiered) testing strategy may be able to replace the *in vivo* eye test (5). The Top-Down approach (5) is designed to be used when, based on existing information, a chemical is expected to have high irritancy potential.

4. Based on the prediction model detailed in paragraph 35, the FL test method can identify substances within a limited applicability domain as ocular corrosives/severe irritants (UN GHS Category 1; EU CLP Category 1; U.S. EPA Category I) without any further testing. The same is assumed for mixtures although mixtures were not used in the validation. Therefore, the FL test method may be used to determine the eye irritancy/corrosivity of chemicals, following the

© OECD, (2012)

You are free to use this material for personal, non-commercial purposes without seeking prior consent from the OECD, provided the source is duly mentioned. Any commercial use of this material is subject to written permission from the OECD.

sequential testing strategy of TG 405 (6). However, a chemical that is not predicted as ocular corrosive or severe irritant with the FL test method would need to be tested in one or more additional test methods (*in vitro* and/or *in vivo*) that are capable of accurately identifying i) chemicals that are *in vitro* false negative ocular corrosives/severe irritants in the FL (UN GHS Category 1; EU CLP Category 1; U.S. EPA Category I); ii) chemicals that are not classified for eye corrosion/irritation (UN GHS No Category; EU CLP No Category; U.S. EPA Category IV); and/or iii) chemicals that are moderate/mild eye irritants (UN GHS Categories 2A and 2B; EU CLP Category 2; U.S. EPA Categories II and III).

5. The purpose of this Test Guideline is to describe the procedures used to evaluate the potential ocular corrosivity or severe irritancy of a test substance as measured by its ability to induce damage to an impermeable confluent epithelial monolayer. The integrity of trans-epithelial permeability is a major function of an epithelium such as that found in the conjunctiva and the cornea. Trans-epithelial permeability is controlled by various tight junctions. Increasing the permeability of the corneal epithelium *in vivo* has been shown to correlate with the level of inflammation and surface damage observed as eye irritation develops.

6. In the FL test method, toxic effects after a short exposure time to the test substance are measured by an increase in permeability of sodium fluorescein through the epithelial monolayer of Madin-Darby Canine Kidney (MDCK) cells cultured on permeable inserts. The amount of fluorescein leakage that occurs is proportional to the chemical-induced damage to the tight junctions, desmosomal junctions and cell membranes, and can be used to estimate the ocular toxicity potential of a test substance. [Annex I](#) provides a diagram of MDCK cells grown on an insert membrane for the FL test method.

7. Definitions are provided in [Annex II](#).

INITIAL CONSIDERATIONS AND LIMITATIONS

8. This Test Guideline is based on the INVITTOX protocol No. 71 (7) that has been evaluated in an international validation study by the European Centre for the Validation of Alternative Methods (ECVAM) (8), in collaboration with the US Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the Japanese Center for the Validation of Alternative Methods (JaCVAM).

9. The FL test method is not recommended for the identification of chemicals which should be classified as mild/moderate irritants or of chemicals which should not be classified for ocular irritation (substances and mixtures) (i.e. GHS Cat. 2A/2B, no category; EU CLP Cat. 2, no category; US EPA Cat. II/III/IV), as demonstrated by the validation study (4) (8).

10. The test method is only applicable to water soluble chemicals (substances and mixtures). The ocular severe irritation potential of chemicals that are water soluble and/or where the toxic effect is not affected by dilution is generally predicted accurately using the FL test method (8). To categorise a chemical as water soluble, under experimental conditions, it should be soluble in sterile calcium-containing (at a concentration of 1.0-1.8 mM), phenol red-free, Hanks' Buffered Salt Solution (HBSS) at a concentration ≥ 250 mg/mL (one dose above the cut-off of 100 mg/mL). However, if the test substance is soluble below the concentration 100 mg/mL,

but already induces a FL induction of 20 % at that concentration (meaning $FL_{20} < 100$ mg/mL), it can still be classified as GHS Cat. 1 or EPA Cat. 1.

11. The identified limitations for this test method exclude strong acids and bases, cell fixatives and highly volatile chemicals from the applicability domain. These chemicals have mechanisms that are not measured by the FL test method, *e.g.* extensive coagulation, saponification or specific reactive chemistries. Other identified limitations for this method are based upon the results for the predictive capacity for coloured and viscous test substance (8). It is suggested that both types of chemicals are difficult to remove from the monolayer following the short exposure period and that predictivity of the test method could be improved if a higher number of washing steps was used. Solid chemicals suspended in liquid have the propensity to precipitate out and the final concentration to cells can be difficult to determine. When substances within these chemical and physical classes are excluded from the database, the accuracy of FL across the EU, EPA, and GHS classification systems is substantially improved (8).

12. Based on the purpose of this test method (*i.e.* to identify ocular corrosives/severe irritants only), false negative rates (see Paragraph 13) are not critical since such substances would be subsequently tested with other adequately validated *in vitro* tests or in rabbits, depending on regulatory requirements, using a sequential testing strategy in a weight of evidence approach (6) (see also paragraphs 3 and 4).

13. Other identified limitations of the FL test method are based on false negative and false positive rates. When used as an initial step within a Top-Down approach to identify water soluble ocular corrosive/severe irritant substances and mixtures (UN GHS Category 1; EU CLP Category 1; U.S. EPA Category I), the false positive rate for the FL test method ranged from 7% (7/103; UN GHS and EU CLP) to 9% (9/99; U.S. EPA) and the false negative rate ranged from 54% (15/28; U.S. EPA) to 56% (27/48; UN GHS and EU CLP) when compared to *in vivo* results. Chemical groups showing false positive and/or false negative results in the FL test method are not defined here.

14. Certain technical limitations are specific to the MDCK cell culture. The tight junctions that block the passage of the sodium-fluorescein dye through the monolayer are increasingly compromised with increasing cell passage number. Incomplete formation of the tight junctions results in increased FL in the non-treated control. Therefore, a defined permissible maximal leakage in the non-treated controls is important (see paragraph 38: 0% leakage). As with all *in vitro* assays there is the potential for the cells to become transformed over time, thus it is vital that passage number ranges for the assays are stated.

15. The current applicability domain might be increased in some cases, but only after analyzing an expanded data set of studied test substances, preferably acquired through testing (4). This Test Guideline will be updated accordingly as new information and data are considered.

16. For any laboratory initially establishing this assay, the proficiency chemicals provided in [Annex III](#) should be used. Laboratories can use these chemicals to demonstrate their technical competence in performing the FL test method prior to submitting FL assay data for regulatory hazard classification purposes.

PRINCIPLE OF THE TEST

17. The FL test method is a cytotoxicity and cell-function based *in vitro* assay that is performed on a confluent monolayer of MDCK CB997 tubular epithelial cells that are grown on semi-permeable inserts and model the non-proliferating state of the *in vivo* corneal epithelium. The MDCK cell line is well established and forms tight junctions and desmosomal junctions similar to those found on the apical side of conjunctival and corneal epithelia. Tight and desmosomal junctions *in vivo* prevent solutes and foreign materials penetrating the corneal epithelium. Loss of trans-epithelial impermeability, due to damaged tight junctions and desmosomal junctions, is one of the early events in chemical-induced ocular irritation.

18. The test substance is applied to the confluent layer of cells grown on the apical side of the insert. A short 1 min exposure is routinely used to reflect the normal clearance rate in human exposures. An advantage of the short exposure period is that water-based substances and mixtures can be tested neat, if they can be easily removed after the exposure period. This allows more direct comparisons of the results with the chemical effects in humans. The test substance is then removed and the non-toxic, highly fluorescent sodium-fluorescein dye is added to the apical side of the monolayer for 30 minutes. The damage caused by the test substance to the tight junctions is determined by the amount of fluorescein which leaks through the cell layer within a defined period of time.

19. The amount of sodium-fluorescein dye that passes through the monolayer and the insert membrane into a set volume of solution present in the well (to which the sodium-fluorescein dye leaks in) is determined by measuring spectrofluorometrically the fluorescein concentration in the well. The amount of fluorescein leakage (FL) is calculated with reference to fluorescence intensity (FI) readings from two controls: a blank control, and a maximum leakage control. The percentage of leakage and therefore amount of damage to the tight junctions is expressed, relative to these controls, for each of the set concentrations of the test substance. Then the FL₂₀ (*i.e.* concentration that causes 20% FL relative to the value recorded for the untreated confluent monolayer and inserts without cells), is calculated. The FL₂₀ (mg/mL) value is used in the prediction model for identification of ocular corrosives and severe irritants (see paragraph 35).

20. Recovery is an important part of a test substance's toxicity profile that is also assessed by the *in vivo* ocular irritation test. Preliminary analyses indicated that recovery data (up to 72 h following the chemical exposure) could potentially increase the predictive capacity of INVITTOX Protocol 71 but further evaluation is needed and would benefit from additional data, preferably acquired by further testing (7). This Test Guideline will be updated accordingly as new information and data are considered.

PROCEDURE

Preparation of the cellular monolayer

21. The monolayer of MDCK CB997 cells is prepared using sub-confluent cells growing in cell culture flasks in DMEM/Nutrient Mix F12 (1x concentrate with L-glutamine, 15 mM HEPES, calcium (at a concentration of 1.0-1.8 mM) and 10% heat-inactivated FCS/FBS). Importantly, all media/solutions used throughout the FL assay should contain calcium at a concentration between 1.8 mM (200 mg/L) and 1.0 mM (111 mg/L) to ensure tight junction formation and integrity. Cell passage number range should be controlled to ensure even and

reproducible tight junctions formation. Preferably, the cells should be within the passage range 3-30 from thawing because cells within this passage range have similar functionality, which aids assay results to be reproducible.

22. Prior to performing the FL test method, the cells are detached from the flask by trypsinisation, centrifuged and an appropriate amount of cells is seeded into the inserts placed in 24-well plates (see [Annex I](#)). Twelve mm diameter inserts with membrane of mixed cellulose esters, a thickness of 80-150 µm and a pore size of 0.45 µm, should be used to seed the cells. In the validation study, Millicell-HA 12 mm inserts were used. The properties of the insert and membrane type are important as these may affect cell growth and chemical binding. Certain types of chemicals may bind to the Millicell-HA insert membrane, which could affect the interpretation of results. Proficiency chemicals (see [Annex III](#)) should be used to demonstrate equivalency if other membranes are used.

23. Chemical binding to the insert membrane is more common for cationic chemicals, such as benzalkonium chloride, which are attracted to the positively charged membrane (8). Chemical binding to the insert membrane may increase the chemical exposure period, leading to an over-estimation of the toxic potential of the chemical, but can also physically reduce the leakage of fluorescein through the insert by binding of the dye to the cationic chemical bound to the insert membrane, leading to an under-estimation of the toxic potential of the chemical. This can be readily monitored by exposing the membrane alone to the top concentration of the chemical tested and then adding sodium-fluorescein dye at the normal concentration for the standard time (no cell control). If binding of the sodium-fluorescein dye occurs, the insert membrane appears yellow after the test material has been washed-off. Thus, it is essential to know the binding properties of the test substance in order to be able to interpret the effect of the chemical on the cells.

24. Cell seeding on inserts should produce a confluent monolayer at the time of chemical exposure. 1.6×10^5 cells should be added per insert (400 µL of a cell suspension with a density of 4×10^5 cells / mL). Under these conditions, a confluent monolayer is usually obtained after 96 hours in culture. Inserts should be examined visually prior to seeding, so as to ensure that any damages recorded at the visual control described at paragraph 30 is due to handling.

25. The MDCK cell cultures should be kept in incubators in a humidified atmosphere, at $5\% \pm 1\%$ CO₂ and 37 ± 1 °C. The cells should be free of contamination by bacteria, viruses, mycoplasma and fungi.

Application of the Test and Control Chemicals

26. A fresh stock solution of test substance should be prepared for each experimental run and used within 30 minutes of preparation. Test substances should be prepared in calcium-containing (at a concentration of 1.0-1.8 mM), phenol red-free, HBSS to avoid serum protein binding. Solubility of the chemical at 250 mg/mL in HBSS should be assessed prior to testing. If at this concentration the chemical forms a stable suspension or emulsion (*i.e.* maintains uniformity and does not settle or separate into more than one phase) over 30 minutes, HBSS can still be used as solvent. However, if the chemical is found to be insoluble in HBSS at this concentration, the use of other test methods instead of FL should be considered. The use of light mineral oil as a solvent, in cases where the chemical is found to be insoluble in HBSS, should be

considered with caution as there is not enough data available to conclude on the performance of the FL assay under such conditions.

27. All chemicals to be tested are prepared in sterile calcium-containing (at a concentration of 1.0-1.8 mM), phenol red-free, HBSS from the stock solution, at five fixed concentrations diluted on a weight per volume basis: 1, 25, 100, 250 mg/mL and a neat or a saturated solution. When testing a solid chemical, a very high concentration of 750 mg/mL should be included. This concentration of chemical may have to be applied on the cells using a positive displacement pipette. If the toxicity is found to be between 25 and 100 mg/mL, the following additional concentrations should be tested twice: 1, 25, 50, 75, 100 mg/mL. The FL₂₀ value should be derived from these concentrations provided the acceptance criteria were met.

28. The test substances are applied to the confluent cell monolayers after removal of the cell culture medium and washing twice with sterile, warm (37°C), calcium-containing (at a concentration of 1.0-1.8 mM), phenol red-free, HBSS. Previously, the filters have been visually checked for any pre-existing damages that could be falsely attributed to potential incompatibilities with test chemicals. At least three replicates should be used for each concentration of the test substance and for the controls in each run. After 1 min of exposure at room temperature, the test substance should be carefully removed by aspiration, the monolayer should be washed twice with sterile, warm (37°C), calcium-containing (at a concentration of 1.0-1.8 mM), phenol red-free, HBSS, and the fluorescein leakage should be immediately measured.

29. Concurrent negative (NC) and positive controls (PC) should be used in each run to demonstrate that monolayer integrity (NC) and sensitivity of the cells (PC) are within a defined historical acceptance range. The suggested PC chemical is Brij 35 (CAS No. 9002-92-0) at 100 mg/mL. This concentration should give approximately 30% fluorescein leakage (acceptable range 20-40% fluorescein leakage, *i.e.* damage to cell layer). The suggested NC chemical is calcium-containing (at a concentration of 1.0-1.8 mM), phenol red-free, HBSS (untreated, blank control). A maximum leakage control should also be included in each run to allow for the calculation of FL₂₀ values. Maximum leakage is determined using a control insert without cells.

Determination of fluorescein permeability

30. Immediately after removal of the test and control substances, 400 µL of 0.1 mg/mL sodium-fluorescein solution (0.01% (w/v) in calcium-containing [at a concentration of 1.0-1.8 mM], phenol red-free, HBSS) is added to the Millicell-HA inserts. The cultures are kept for 30 minutes at room temperature. At the end of the incubation with fluorescein, the inserts are carefully removed from each well. Visual check is performed on each filter and any damage which may have occurred during handling is recorded.

31. The amount of fluorescein that leaked through the monolayer and the insert is quantified in the solution which remained in the wells after removal of the inserts. Measurements are done in a spectrofluorometer at excitation and emission wavelengths of 485 nm and 530 nm, respectively. The sensitivity of the spectrofluorometer should be set so that there is the highest numerical difference between the maximum FL (insert with no cells) and the minimum FL (insert with confluent monolayer treated with NC). Because of the differences in the used spectrofluorometer, it is suggested that a sensitivity is used which will give fluorescence intensity > 4000 at the maximum fluorescein leakage control. The maximum FL value should not be

greater than 9999. The maximum fluorescence leakage intensity should fall within the linear range of the spectrofluorometer used.

Interpretation of results and Prediction model

32. The amount of FL is proportional to the chemical-induced damage to the tight junctions. The percentage of FL for each tested concentration of chemical is calculated from the FL values obtained for the test substance with reference to FL values from the NC (reading from the confluent monolayer of cells treated with the NC) and a maximum leakage control (reading for the amount of FL through an insert without cells).

The mean maximum leakage fluorescence intensity = x

The mean 0% leakage fluorescence intensity (NC) = y

The mean 100% leakage is obtained by subtracting the mean 0% leakage from the mean maximum leakage,

i.e. $x - y = z$

33. The percentage leakage for each fixed dose is obtained by subtracting the 0% leakage to the mean fluorescence intensity of the three replicate readings (m), and dividing this value by the 100% leakage, *i.e.* $\%FL = [(m-y) / z] \times 100\%$, where:

m = the mean fluorescence intensity of the three replicate measurements for the concentration involved

% FL = the percent of the fluorescein which leaks through the cell layer

34. The following equation for the calculation of the chemical concentration causing 20% FL should be applied:

$$FL_D = [(A-B) / (C-B)] \times (M_C - M_B) + M_B$$

Where:

D = % of inhibition

A = % damage (20% fluorescein leakage)

B = % fluorescein leakage < A

C = % fluorescein leakage > A

M_C = Concentration (mg/mL) of C

M_B = Concentration (mg/mL) of B

35. The cut-off value of FL_{20} for predicting chemicals as ocular corrosives/severe irritants is given below:

FL₂₀ (mg/mL)	UN GHS C&L	EU CLP C&L	U.S. EPA C&L
≤ 100	Category 1	Category 1	Category I

C&L: classification and labelling

36. The FL test method is recommended only for the identification of water soluble ocular corrosives and severe irritants (UN GHS Category 1, EU CLP Category 1, U.S. EPA Category I) (see paragraphs 1 and 10).

37. In order to identify water soluble chemicals (substances and mixtures) (4) (7) (8) as "inducing serious eye damage" (UN GHS/EU CLP Category 1) or as an "ocular corrosive or severe irritant" (U.S. EPA Category I), the test substance should induce an FL₂₀ value of ≤ 100 mg/mL.

Acceptance of results

38. The mean maximum fluorescein leakage value (x) should be higher than 4000 (see paragraph 31), the mean 0% leakage (y) should be equal or lower than 300, and the mean 100% leakage (z) should fall between 3700 and 6000.

39. A test is considered acceptable if the positive control produced 20% to 40% damage to the cell layer (measure as % fluorescein leakage).

DATA AND REPORTING

Data

40. For each run, data from individual replicate wells (*e.g.* fluorescence intensity values and calculated percentage FL data for each test substance, including classification) should be reported in tabular form. In addition, means \pm SD of individual replicate measurements in each run should be reported.

Test Report

41. The test report should include the following information:

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;
- Chemical CAS number, if known;
- Purity and composition of the substance or mixture (in percentage(s) by weight), to the extent this information is available;
- Physical-chemical properties relevant to the conduct of the study (*e.g.* physical state, volatility, pH, stability, water solubility, chemical class);
- Treatment of the test/control substance prior to testing, if applicable (*e.g.* warming, grinding);
- Storage conditions;

Justification of the Test Method and Protocol Used

- Should include considerations regarding applicability domain and limitations of the test method;

Test Conditions

- Description of cell system used, including certificate of authenticity and the mycoplasma status of the cell line;
- Details of test procedure used;
- Test substance concentration(s) used;
- Duration of exposure to the test substance;
- Duration of incubation with fluorescein;
- Description of any modifications of the test procedure;
- Description of evaluation criteria used;
- Reference to historical data of the model (*e.g.* negative and positive controls, benchmark chemicals, if applicable);
- Information on the technical proficiency demonstrated by the laboratory;

Results

- Tabulation of data from individual test substances and controls for each run and each replicate measurement (including individual results, means and SDs);
- The derived classification(s) with reference to the prediction model and/or decision criteria used;
- Description of other effects observed;

Discussion of the Results

- Should include considerations regarding a non-conclusive outcome (paragraph 35: FL₂₀ > 100 mg/mL) and further testing;

Conclusions

LITERATURE

1. UN (2009), United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS), Third revised edition, New York & Geneva: United Nations Publications. ISBN: 978-92-1-117006-1. Available at: [\[http://www.unece.org/trans/danger/publi/ghs/ghs_rev03/03files_e.html\]](http://www.unece.org/trans/danger/publi/ghs/ghs_rev03/03files_e.html)
2. EC (2008), Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006, Official Journal of the European Union L353, 1-1355.
3. U.S. EPA (1996), Label Review Manual: 2nd Edition, EPA737-B-96-001, Washington DC: U.S. Environmental Protection Agency.
4. EC-ECVAM (2009), Statement on the scientific validity of cytotoxicity/cell-function based *in vitro* assays for eye irritation testing. Available under *Publications* at: [\[http://ecvam.jrc.it/index.htm\]](http://ecvam.jrc.it/index.htm)
5. Scott, L. *et al.* (2010), A proposed eye irritation testing strategy to reduce and replace *in vivo* studies using Bottom-Up and Top-Down approaches, *Toxicol. In Vitro* 24, 1-9.
6. OECD (2002), *Test No. 405: Acute Eye Irritation/Corrosion*, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing. doi: [10.1787/9789264070646-en](https://doi.org/10.1787/9789264070646-en)
7. EC-ECVAM (1999), INVITOX Protocol 71: Fluorescein Leakage Test, Ispra, Italy: European Centre for the Validation of Alternative Methods (ECVAM). Available at: [\[http://ecvam-dbalm.jrc.ec.europa.eu\]](http://ecvam-dbalm.jrc.ec.europa.eu)
8. EC-ECVAM (2008), Fluorescein Leakage Assay Background Review Document as an Alternative Method for Eye Irritation Testing. Available under *Validation Study Documents*, Section *Eye Irritation* at: [\[http://ecvam.jrc.it/index.htm\]](http://ecvam.jrc.it/index.htm)
9. OECD (2005), *Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment*, OECD Series on Testing and Assessment No. 34. OECD, Paris. Available at: [\[http://www.oecd.org/env/testguidelines\]](http://www.oecd.org/env/testguidelines)

ANNEX I

DIAGRAM OF MDCK CELLS GROWN ON AN INSERT MEMBRANE FOR THE FL TEST METHOD

A confluent layer of MDCK cells is grown on the semi-permeable membrane of an insert. The inserts are placed into the wells of 24 well plates.

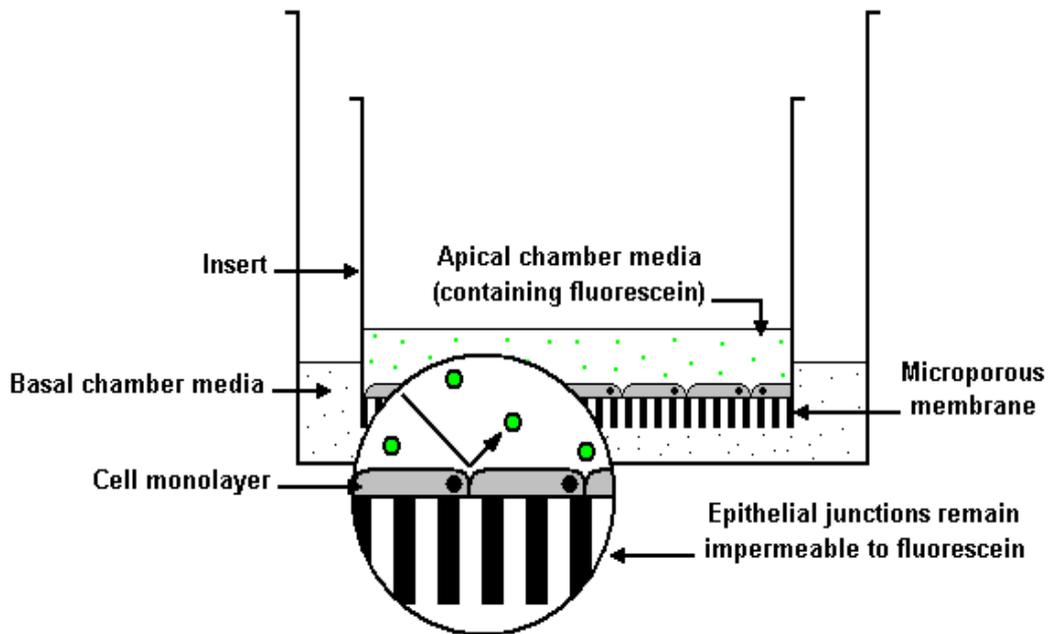


Figure taken from: Wilkinson, P.J. (2006), Development of an *in vitro* model to investigate repeat ocular exposure, Ph.D. Thesis, University of Nottingham, UK.

ANNEX II

DEFINITIONS

Accuracy: The closeness of agreement between test method results and accepted reference values. It is a measure of test method performance and one aspect of “relevance.” The term is often used interchangeably with “concordance”, to mean the proportion of correct outcomes of a test method.

EPA Category I: Chemicals that produce corrosive (irreversible destruction of ocular tissue) or corneal involvement or irritation persisting for more than 21 days (4).

EU CLP (European Commission Regulation on the Classification, Labelling and Packaging of Substances and Mixtures): Implements in the European Union (EU) the UN GHS system for the classification of chemicals (substances and mixtures) (3).

False negative rate: The proportion of all positive chemicals falsely identified by a test method as negative. It is one indicator of test method performance.

False positive rate: The proportion of all negative chemicals that are falsely identified by a test method as positive. It is one indicator of test method performance.

FL₂₀: Can be estimated by the determination of the concentration at which the tested chemical causes 20% of the fluorescein leakage through the cell layer.

Fluorescein leakage: the amount of fluorescein which passes through the cell layer, measured spectrofluorometrically.

GHS (Globally Harmonized System of Classification and Labeling of Chemicals by the United Nation (UN)): A system proposing the classification of chemicals (substances and mixtures) according to standardized types and levels of physical, health and environmental hazards, and addressing corresponding communication elements, such as pictograms, signal words, hazard statements, precautionary statements and safety data sheets, so that to convey information on their adverse effects with a view to protect people (including employers, workers, transporters, consumers and emergency responders) and the environment (2).

GHS Category 1: Production of tissue damage in the eye, or serious physical decay of vision, following application of a test substance to the anterior surface of the eye, which is not fully reversible within 21 days of application.

Hazard: Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub) population is exposed to that agent.

Mixture: Used in the context of the UN GHS (2) as a mixture or solution composed of two or more substances in which they do not react.

Negative control: An untreated replicate containing all components of a test system. This sample is processed with test substance-treated samples and other control samples to determine whether the solvent interacts with the test system.

Not-classified: Chemicals that are not classified as UN GHS Categories 1, 2A, or 2B; EU CLP Categories 1 or 2; or U.S. EPA Categories I, II, or III ocular irritants (2) (3) (4).

Ocular corrosive: (a) A chemical that causes irreversible tissue damage to the eye. (b) Chemicals that are classified as UN GHS Category 1; EU CLP Category 1; or U.S. EPA Category I ocular irritants (2) (3) (4).

Ocular irritant: (a) A chemical that produces a reversible change in the eye following application to the anterior surface of the eye; (b) Chemicals that are classified as UN GHS Categories 2A, or 2B; EU CLP Category 2; or U.S. EPA Categories II or III ocular irritants (2)(3)(4).

Ocular severe irritant: (a) A chemical that causes tissue damage in the eye following application to the anterior surface of the eye that is not reversible within 21 days of application or causes serious physical decay of vision. (b) Chemicals that are classified as UN GHS Category 1; EU CLP Category 1; or U.S. EPA Category I ocular irritants (2) (3) (4).

Positive control: A replicate containing all components of a test system and treated with a chemical known to induce a positive response. To ensure that variability in the positive control response across time can be assessed, the magnitude of the positive response should not be extreme.

Proficiency Chemicals: A sub-set of the list of Reference Chemicals that can be used by a naïve laboratory to demonstrate proficiency with the validated reference test method.

Relevance: Description of relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method (9).

Reliability: Measures of the extent that a test method can be performed reproducibly within and between laboratories over time, when performed using the same protocol. It is assessed by calculating intra- and inter-laboratory reproducibility and intra-laboratory repeatability.

Replacement test: A test which is designed to substitute for a test that is in routine use and accepted for hazard identification and/or risk assessment, and which has been determined to provide equivalent or improved protection of human or animal health or the environment, as applicable, compared to the accepted test, for all possible testing situations and chemicals.

Sensitivity: The proportion of all positive/active chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method (9).

Serious eye damage: Is the production of tissue damage in the eye, or serious physical decay of vision, following application of a test substance to the anterior surface of the eye, which is not fully reversible within 21 days of application.

Solvent/vehicle control: An untreated sample containing all components of a test system, including the solvent or vehicle that is processed with the test substance-treated and other control samples to establish the baseline response for the samples treated with the test substance dissolved in the same solvent or vehicle. When tested with a concurrent negative control, this sample also demonstrates whether the solvent or vehicle interacts with the test system.

Specificity: The proportion of all negative/inactive chemicals that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method.

Substance: Used in the context of the UN GHS as chemical elements and their compounds in the natural state or obtained by any production process, including any additive necessary to preserve the stability of the product and any impurities deriving from the process used, but excluding any solvent which may be separated without affecting the stability of the substance or changing its composition.

Tiered testing strategy: A stepwise testing strategy where all existing information on a test substance is reviewed, in a specified order, using a weight-of-evidence process at each tier to determine if sufficient information is available for a hazard classification decision, prior to progression to the next tier. If the irritancy potential of a test substance can be assigned based on the existing information, no additional testing is required. If the irritancy potential of a test substance cannot be assigned based on the existing information, a step-wise sequential animal testing procedure is performed until an unequivocal classification can be made.

Validated test method: A test method for which validation studies have been completed to determine the relevance (including accuracy) and reliability for a specific purpose. It is important to note that a validated test method may not have sufficient performance in terms of accuracy and reliability to be found acceptable for the proposed purpose (9).

Weight-of-evidence: The process of considering the strengths and weaknesses of various pieces of information in reaching and supporting a conclusion concerning the hazard potential of a chemical.

ANNEX III

PROFICIENCY CHEMICALS FOR THE FL TEST METHOD

Prior to routine use of a test method that adheres to this Test Guideline, laboratories should demonstrate technical proficiency by correctly identifying the ocular corrosivity classification of the 8 chemicals recommended in Table 1. These chemicals were selected to represent the range of responses for local eye irritation/corrosion, which is based on results in the *in vivo* rabbit eye test (TG 405) (*i.e.*, Categories 1, 2A, 2B, or No Category according to the UN GHS and EU CLP (1)(2)(6)). However, considering the validated usefulness of the FL assay (*i.e.*, to identify ocular corrosives/severe irritants only), there are only two test outcomes for classification purposes (corrosive/severe irritant or non-corrosive/non-severe irritant) to demonstrate proficiency. Other selection criteria were that chemicals are commercially available, there are high quality *in vivo* reference data available, and there are high quality data from the FL test method. For this reason, the proficiency chemicals were selected from the "Fluorescein Leakage Assay Background Review Document as an Alternative Method for Eye Irritation Testing" (8), which was used for the retrospective validation of the FL test method.

Table 1: Recommended chemicals for demonstrating technical proficiency with FL

Chemical	CAS NR	Chemical Class ¹	Physical Form	In Vivo Classification ²	In Vitro Classification ³
Benzalkonium chloride (5%)	8001-54-5	Onium compound	Liquid	Category 1	Corrosive/ Severe Irritant
Promethazine hydrochloride	58-33-3	Amine/Amidine, Heterocyclic, Organic sulphur compound	Solid	Category 1	Corrosive/ Severe Irritant
Sodium hydroxide (10%)	1310-73-2	Alkali	Liquid	Category 1	Corrosive/ Severe Irritant
Sodium lauryl sulfate (15%)	151-21-3	Carboxylic acid (salt)	Liquid	Category 1	Corrosive/ Severe Irritant
4-carboxy-benzaldehyde	619-66-9	Carboxylic acid, Aldehyde	Solid	Category 2(A)	Non-corrosive/ Non-severe irritant
Ammonium nitrate	6484-52-2	Inorganic salt	Solid	Category 2(A)	Noncorrosive/ Non-severe irritant
Ethyl-2-methylacetoacetate	609-14-3	Ketone, Ester	Liquid	Category 2(B)	Noncorrosive/ Non-severe irritant
Glycerol	56-81-5	Alcohol	Liquid	No Category	Noncorrosive/ Non-severe irritant

Abbreviations: CAS NR = Chemical Abstracts Service Registry Number

¹Chemical classes were assigned to each test substance using a standard classification scheme, based on the National Library of Medicine Medical Subject Headings (MeSH) classification system (available at <http://www.nlm.nih.gov/mesh>)

²Based on results from the *in vivo* rabbit eye test (OECD TG 405) and using the UN GHS and EU CLP (1)(2)(6).

³Based on results obtained with FL (INVITTOX Protocol No. 71).

Fluorescein Leakage Assay Background Review Document As An Alternative Method for Eye Irritation Testing

**Dr Joanne Gartlon
Dr Richard Clothier**

**CCR.IHCP.C431273.X0
January 2008**



**The University of
Nottingham**

Acknowledgements

The generous contributions from the following companies and individuals who provided *in vitro* and/or *in vivo* data considered in this Background Review Document are gratefully acknowledged:

Avon Products Inc., Suffern, USA
Uma Santhanam, Steve Gettings

Cosmetic, Toiletry, and Fragrance Association, Washington, USA
Carol Eisenmann

European Centre for the Validation of Alternative Methods, Ispra, Italy
Chantra Eskes, Valerie Zuang, Pilar Prieto

European Chemicals Bureau, Ispra, Italy
Andrew Worth

Fund for the Replacement of Animals in Medical Experiments (FRAME)
Michael Balls

Institute for In Vitro Sciences, Inc., Gaithersburg, USA
Roger Curren, John Harbell

J&J Consumer Products Worldwide R&D&E, Skillman, USA
Lauren Bernhofer, Lily Agarwal

L'Oreal Research, Aulnay Sous Bois, France
Christine van den Berghe, Odile de Silva

Fluorescein Leakage Assay Background Review Document

List of contents

Based on the Modular Approach to Validation and on the ECVAM
Validation Workshop

Number	Page
1. Data collection	1
1.1. Description of the methods used to collect data, including literature searches or other sources, and number of studies collected	1
1.2. Brief description of data collected on overall study management	6
2. Test Definition (Module 1)	14
2.1. Rationale for the proposed test method	14
2.1.1. Intended uses/ purpose	14
2.1.2. Regulatory rationale and applicability	15
2.1.3. Scientific basis for the test (mechanistic)	15
2.1.4. Similarities and differences of modes of action in the test method and the reference species	19
2.2. Test method protocols	27
2.2.1. Description of protocol components and rationale for differences, if available	51
2.2.2. Proposed critical components of the protocol that impact on reproducibility and/or predictive capacity of the assay include PM	60
2.2.3. List of studies with similar protocols (no protocol differences, or no impact of protocol differences on predictive capacity)	61
2.2.4. Known applicability and limitations of the assay (including ranges of irritancy, types of substances, technical limitations)	63
2.2.5. Others	67

3.	Within-laboratory reproducibility (Module 2)	70
3.1.	Table presenting the results and relevant information for studies	71
3.2.	Compilation of results	74
3.2.1.	Statistical approach(es) used: description & rationale for the approach used	74
3.2.2.	Results and discussion	76
3.2.3.	Additional analyses of operator and time variability	96
3.3	Additional studies where raw data are not available: attempt to combine the data using weight-of-evidence approaches	104
4.	Transferability (Module 3)	107
4.1.	Brief description of study results on transferability and availability of Standardised Operating Procedures (SOPs)	107
4.2.	Facilities and major fixed equipment needed	109
4.3.	Required level of training, expertise, and demonstrated proficiency needed	109
5.	Between-laboratory reproducibility (Module 4)	111
5.1.	Table presenting the relevant results and information for studies where raw data are available	112
5.2.	Discussions from the literature	113
5.3.	Compilation of results	115
5.3.1.	Statistical approach(es) used: description & rationale for the approach used to determine between-laboratory reproducibility	115
5.3.2.	Relevant results and information for each study	116
5.4.1.	Additional studies where raw data are not available: attempt to combine the data using weight-of-evidence approaches	122
5.4.2.	Discussions from the literature	123
5.4.3.	Compilation of results	125
5.4.3.1.	Statistical approach(es) used: description & rationale for the approach used to determine between-laboratory reproducibility	125

5.4.3.2	Compilation of results	126
5.4.3.3	Attempt to combine the data using weight-of-evidence approaches	138
6.	Predictive Capacity (Module 5)	140
6.1.	Studies with available raw data	141
6.1.1.	Table presenting the relevant information for each study where raw <i>in vitro</i> and <i>in vivo</i> data were available	141
6.1.2.	<i>In vivo</i> reference data used to assess the performance of the FL assay for studies where raw <i>in vitro</i> and <i>in vivo</i> data were available	142
6.1.3.	Brief description of the studies with available raw <i>in vitro</i> and <i>in vivo</i> data	143
a.	Statistical approaches detailed in the CTFA publications (Gettings <i>et al.</i> , 1996)	143
b.	Principal results reported in the literature	143
6.1.4.	Compilation of data on predictive capacity of the test method from studies with raw data available	145
6.1.4.1.	Description & rationale for the PM applied and statistical approaches used	145
a.	Classification systems	145
b.	Prediction models (PMs)	145
6.1.4.2.	Description of the performance compared to the reference and eventually the human situation for each study	146
6.1.4.3.	Discussions	151
a.	Descriptions of limitations of the test method (i.e. applicability domain based on the results of the compiled data)	151
b.	Possible rationale for the differences observed	152
6.2.1.	Table presenting the relevant information for each study where raw data were not available	153
6.2.2.	<i>In vivo</i> reference data used to assess the performance of the FL assay	157
6.2.3.	Brief description of the studies without raw data available	161
a.	Statistical Approaches	161
b.	Main results	164
6.2.4.	Compilation of data on predictive capacity of the test method from studies	

without raw data	170
6.2.4.1. Description & rationale for the prediction model(s) applied and statistical approach(es) used	170
6.2.4.2. Description of performance compared to reference and eventually, to the human situation for each study	176
6.2.4.3. Discussions	201
a. Description of the limitations of the test method (e.g., applicability domain based on results from compilation of data)	201
b. Possible rationale for differences observed	206
6.3.1. Attempt to combine the data using weight-of-evidence approaches	208
6.3.2. Modifications to the assay if designed today	213
7. Applicability Domain	215
8. Supporting materials	222
8.1. Relevant publications, other scientific reports and reviews	222
8.2. Relevant unpublished data	223
9. References	224

Appendices

Appendix I	Standard letter inviting companies to contribute to the FL assay BRD, and accompanying FL assay questionnaire	2
Appendix II	Complete list of protocol steps	6

Annexes

Annex I	Standard FL Protocols: a. The Fluorescein Leakage Test, INVITTOX Protocol No. 71; b. Fluorescein Leakage Test - SOP of Company # 4, INVITTOX Protocol No. 120; c. Trans-epithelial Permeability (TEP) Assay, INVITTOX Protocol No. 86; d. Fixed Dose Procedure For The Fluorescein Leakage Test, INVITTOX Protocol No. 82; e. Company # 3 Robotic Trans-Epithelial Permeability (TEP) Assay; f. ECVAM Prevalidation Study Phase II Protocol; g. ECVAM Prevalidation Study Phase III Protocol	63
Annex IIa	Raw data used in the statistical analyses for intra-laboratory variability	144
Annex IIb	Raw data used in the statistical analyses for intra-laboratory variability – CTFA Phase III TEP Assay data	145
Annex III	Between-laboratory FL assay data for a. COLIPA test materials: i) chemicals ii) formulations b. EC/HO chemicals; c. ECVAM Prevalidation Phase II; d. ECVAM Prevalidation Phase III	147
Annex IVa	Information available on the chemicals tested to assess FL assay predictivity of <i>in vivo</i> eye irritation	148
Annex IVb	Information available on the products/formulations tested to assess FL assay predictivity of <i>in vivo</i> eye irritation	188
Annex Vai	CTFA Study Phase III data set containing predicted and actual EU, GHS and EPA classifications for each formulation tested using the	

	TEP assay/ INVITTOX Protocol No. 86. Raw <i>in vitro</i> and <i>in vivo</i> data were available	203
Annex Vaii	Data sets containing predicted and actual EU, GHS and EPA classifications for each test material. Data sets contained summarised <i>in vitro</i> data.	208
Annex Vb	<i>In vivo</i> reference data as reported in the literature for each substance tested using the FL assay.	209
Annex Vc	<i>In vitro and in vivo</i> reference data as submitted by Company # 3 Consumer Products Worldwide R&D&E for 17 formulations.	210
Annex VI	Table presenting the studies where FL assay results were used to assess the assay's predictive capacity for <i>in vivo</i> ocular irritation	211
Annex VII	Raw <i>in vivo</i> data as entered into ECVAM v6 template	222
	Formulation Annexes	223
Annex A	CTFA Study Phase III formulation compositions (from draft HET-CAM BRD: Appendix C2 (ICCVAM/NICEATM, 2004)	224
Annex B	COLIPA Study test chemicals and formulations compositions (from COLIPA)	233
Annex C	Formulation compositions from Company # 3	246

List of Tables

1.1. Results of the different search terms in the various databases	1
1.2 Description of studies featuring FL assay protocol information and/or data, in chronological order.	6
2.1.3. Summary of the physicochemical properties that the FL assay is (in)capable of modelling.	18
2.1.4.1 Summary of the events involved in chemical-induced eye irritation <i>in vivo</i> which are (not) modelled by the FL assay.	23
2.1.4.2. Examples of chemical classes and their effects through the different mechanisms categorised (Scott <i>et al.</i> , (<i>under preparation</i>)).	24
2.2. Test Method Protocols	27
2.2.1.1. Insert properties (modified from Ward <i>et al.</i> , 1997a)	55
2.2.1.2. PM for data generated using INVITTOX Protocol No. 71 as presented in the EC/HO Study publication (Balls <i>et al.</i> , 1995).	58
2.2.1.3. The PM for INVITTOX Protocol No. 82 (taken from the publication of Clothier <i>et al.</i> , (1994)).	59
2.2.1.4. PM from INVITTOX Protocol No. 120	59
2.2.1.5 PM from INVITTOX Protocol No. 86 (TEP assay)	59
3.1. Table presenting the results and relevant information for studies	71
3.2.2. Summary table of the overall reproducibility per study. There were 25 data sets (compared to 14 in table 3.1) as some studies contained multiple data sets (e.g. ECVAM Prevalidation study (Southee, 1998)).	77
3.2.2.1. The main protocol differences between the FL assay protocols for which data were available to assess intra-laboratory variability.	84
3.2.2.2. Relevant <i>in vitro</i> and <i>in vivo</i> information for the types of test materials and potencies.	90
3.2.2.3. PM from the COLIPA study for a version of INVITTOX Protocol No. 120.	94
3.2.2.4. Variation in classifications assigned according to the COLIPA PM for the three laboratories that participated in the ECVAM	

Prevalidation study Phase III (n ≥2).	95
3.2.3.1. Results and relevant information for each study	97
3.2.3.2. Operator variability for all the data sets featured in table 3.2.3.1	100
3.2.3.3. Impact of failed experimental runs on overall reproducibility.	101
3.2.3.4. The number of materials for which two or more outliers were identified in relation to the total number of materials within the data set	102
3.3. Presents the results and relevant information for each study containing non-raw data.	104
4.2. The fixed equipment required to carry out the FL assay	109
5.1. Table presenting the relevant results and information for studies where raw data are available	112
5.3.2.1.a. Summarised ECVAM Prevalidation Phase II FL ₂₀ (mg/ml) data (All raw data for Phase II, for each laboratory provided in Annex III on CD)	116
5.3.2.1.b. Summarised ECVAM Prevalidation Phase III FL ₂₀ (mg/ml) data (All raw data for Phase III, for each laboratory provided in Annex III on CD)	116
5.3.2.2. PM applied to ECVAM Prevalidation study Phase III data	120
5.3.2.3. Mean FL ₂₀ (mg/ml) T4 data and classification for each test material in each laboratory.	120
5.3.2.4. Summary table of proportion of identical predicted classifications from each laboratory	121
5.4.1. Relevant results and information for studies where raw data were not available	122
5.4.3.2.1. Between-laboratory FL ₂₀ (mg/ml) data for COLIPA study test materials.	126
5.4.3.2.2. Between-laboratory FL ₂₀ (mg/ml) data for EC/HO study test chemicals	128
5.4.3.2.3. Inter-laboratory correlation of the FL ₂₀ (mg/ml) data from the four laboratories (from Balls <i>et al.</i> , 1995).	129

5.4.3.2.4. Modified PM from EC/HO study used to assign irritant and non-irritant classifications.	130
5.4.3.2.5. Predicted classifications for EC/HO study data generated using the PM featured in table 5.4.3.2.4.	131
5.4.3.2.6. Summary table of proportion of identical predicted classifications from each laboratory.	134
5.4.3.2.7. Predicted classifications for COLIPA Study data generated using the PM featured in table 5.3.2.2.	135
5.4.3.2.8. Summary table of proportion of identical predicted classifications from each laboratory.	137
5.4.3.2.9. FL ₂₀ (mg/ml) data from the EC/HO study and the COLIPA study for the nine common test chemicals (from Zanvit <i>et al.</i> , 1999). A version of INVITTOX Protocol No. 120 was used.	138
6.1.1. Table presenting the relevant information for each study where raw <i>in vitro</i> and <i>in vivo</i> data were available	141
6.1.2. <i>In vivo</i> reference data used to assess the performance of the FL assay for studies where raw <i>in vitro</i> and <i>in vivo</i> data were available	142
6.1.3. The findings reported in the literature for those studies which used PMs to assess the predictive capacity of the TEP assay.	144
6.1.4.1.1. PM from INVITTOX Protocol No. 86	145
6.1.4.1.2. Modified TEP assay PM for predicting EU, GHS and EPA classifications.	146
6.1.4.2.1.a. Contingency tables for CTFA Phase III data generated using INVITTOX Protocol No. 86 and the PM featured in table 6.1.4.1.2. -EU classification system	147
6.1.4.2.1.b. Contingency tables for CTFA Phase III data generated using INVITTOX Protocol No. 86 and the PM featured in table 6.1.4.1.2. -GHS classification system	147
6.1.4.2.1.c. Contingency tables for CTFA Phase III data generated using INVITTOX Protocol No. 86 and the PM featured in table 6.1.4.1.2. -EPA classification system	148

6.1.4.2.2.a. Evaluation of the performance of INVITTOX Protocol No. 86 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest	150
6.1.4.2.2.b. Evaluation of the performance of INVITTOX Protocol No. 86 for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest.	150
6.1.4.2.2.c. Evaluation of the performance of INVITTOX Protocol No. 86 for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest	150
6.2.1. Table presenting the relevant information for each study where raw data were not available.	153
6.2.2. <i>In vivo</i> reference data used to assess the performance of the FL assay	157
6.2.3. The findings reported in the literature for those studies which used PMs to assess the predictive capacity of the various FL assay protocols.	164
6.2.4.1.1. PM submitted to ECVAM for FL ₂₀ (mg/ml) values with recovery data, based upon the PM featured in the EC/HO study (Balls <i>et al.</i> , 1995).	171
6.2.4.1.2. ICCVAM/NICEATM criteria for assigning GHS classifications to Draize data (from draft HET-CAM BRD (2004)).	172
6.2.4.1.3. ICCVAM/NICEATM criteria for assigning EU classifications to Draize data (from draft HET-CAM BRD (2004)).	173
6.2.4.1.4. EPA classification system for ocular irritation (from draft HET-CAM BRD (2004)).	173
6.2.4.1.5. PM for FL ₂₀ (mg/ml) values.	174
6.2.4.1.6. PM for FL ₂₀ (mg/ml) values produced 4h following a 15 minute chemical exposure period (modified from Brantom <i>et al.</i> , 1997).	174
6.2.4.1.7. Modified PM for INVITTOX Protocol No. 82 data (from Clothier <i>et al.</i> , (1994)).	175
6.2.4.2.1.a. Contingency table for EC/HO study data generated using	

INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. -EU classification system	177
6.2.4.2.1.b. Contingency table for EC/HO study data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. -GHS classification system	178
6.2.4.2.1.c. Contingency table for EC/HO study data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. -EPA classification system	179
6.2.4.2.2.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest.	180
6.2.4.2.2.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest.	180
6.2.4.2.2.c. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest.	180
6.2.4.2.3.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Severe irritants versus the rest.	182
6.2.4.2.3.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Severe irritants versus the rest.	182
6.2.4.2.3.c. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EPA classification system –Severe irritants versus the rest.	182
6.2.4.2.4.a. Contingency table for data generated using INVITTOX Protocol No. 71 with recovery time-point and the PM featured in table 6.2.4.1.1. -EU classification system.	183
6.2.4.2.4.b. Contingency table for data generated using INVITTOX Protocol No. 71 with recovery time-point and the PM featured in	

table 6.2.4.1.1. -GHS classification system.	184
6.2.4.2.4.c. Contingency table for data generated using INVITTOX Protocol No. 71 with recovery time-point and the PM featured in table 6.2.4.1.1. -EPA classification system.	184
6.2.4.2.5.a Evaluation of the performance of INVITTOX Protocol No. 71 with recovery data for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest	186
6.2.4.2.5.b. Evaluation of the performance of INVITTOX Protocol No. 71 with recovery data for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest.	186
6.2.4.2.5.c. Evaluation of the performance of INVITTOX Protocol No. 71 with recovery data for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest.	186
6.2.4.2.6. Evaluation of the performance of INVITTOX Protocol No. 71 with recovery data for predicting ocular irritation according to the GHS classification system –Severe irritants versus the rest.	187
6.2.4.2.7.a. Contingency table for data generated using INVITTOX Protocol No. 120 and the PM featured in table 6.2.4.1.6. –EU classification system.	188
6.2.4.2.7.b. Contingency table for data generated using INVITTOX Protocol No. 120 and the PM featured in table 6.2.4.1.6. –GHS classification system.	189
6.2.4.2.7.c. Contingency table for data generated using INVITTOX Protocol No. 120 and the PM featured in table 6.2.4.1.6. –EPA classification system.	189
6.2.4.2.8.a Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the EU classification system –Non-irritant versus the rest.	190
6.2.4.2.8.b. Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the GHS classification system –Non-irritant versus the rest.	190

6.2.4.2.8.c. Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the EPA classification system –Non-irritant versus the rest.	190
6.2.4.2.9.a. Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the GHS classification system –Severe irritants versus the rest.	191
6.2.4.2.9.b. Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the EPA classification system –Severe irritants versus the rest.	191
6.2.4.2.10. Contingency table for data generated using INVITTOX Protocol No. 82 and the PM featured in table 6.2.4.1.7. –EU classification system.	192
6.2.4.2.11. Evaluation of the performance of INVITTOX Protocol No. 82 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest.	193
6.2.4.2.12.a. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –EU classification system.	194
6.2.4.2.12.b. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –GHS classification system.	194
6.2.4.2.12.c. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –EPA classification system.	195
6.2.4.2.13.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest.	196
6.2.4.2.13.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest.	196
6.2.4.2.13.c. Evaluation of the performance of INVITTOX Protocol No. 71	

for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest.	196
6.2.4.2.14.a. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –EU classification system.	197
6.2.4.2.14.b. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –GHS classification system.	197
6.2.4.2.14.c. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –EPA classification system.	198
6.2.4.2.15.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest.	199
6.2.4.2.15.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest.	199
6.2.4.2.15.c. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest.	199
6.2.4.2.16.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Severe irritants versus the rest.	200
6.2.4.2.16.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Severe irritants versus the rest.	200
6.2.4.2.16.c. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EPA classification system –Severe irritants versus the rest.	200

List of Figures

2.1.3.1. Diagram of MDCK cells grown on an insert membrane for the FL assay (from Wilkinson, 2006).	16
2.1.3.2. Equation for the calculation of the chemical/formulation concentration causing 20 percent FL (FL ₂₀) (from INVITTOX Protocol No. 71).	17
2.1.4. Weighted scoring system for ocular lesions (from Draize <i>et al.</i> , 1944).	21
2.2.2.1. Critical elements of FL assay protocols	60
3.2.1. ◇ SD and ◆ CV values plotted against the mean FL assay results;	
a. FL ₂₀ (mg/ml)	74
b. FL (%)	75
c. FL (mM) (n≥2)	75
3.2.2.1. The CV values for INVITTOX Protocol No. 71 ◆, and INVITTOX Protocol No. 120 ◇ plotted against FL assay results for all available raw data.	87
3.2.2.2. The CV values for TEP assay data plotted against the mean TEP assay values (n=2) for Company # 3 formulations	88
3.2.2.3. The effect of time on the ECVAM Prevalidation Phase II SLS positive control FL% values for the four participating laboratories at a.) T0; b.) T4; □, FAL; ■, Company # 7; ▲, Company # 3; Δ, ECVAM	89
3.2.2.4. The CV values plotted against mean FL ₂₀ (mg/ml) T4 values for the ten surfactants tested in three laboratories; □, FAL; ▲, Company # 3; Δ, ECVAM	91
3.2.2.5. The CVs plotted against the mean T4 FL ₂₀ (mg/ml) values for surfactants and surfactant-based formulations tested by the FAL as part of the COLIPA study.	92
3.2.2.6. The CVs plotted against the mean EC ₅₀ (%) values for a. formulations tested by Company # 3 Consumer Products Worldwide	

R&D&E; b. CTFA Phase III data.	93
5.3.2.1. Mean FL% \pm SD values for Phase II positive control data (0.16mg/ml SLS) from each laboratory at T0 (pattern) and T4 (blank).	118
5.3.2.2. Mean FL% \pm SD values for Phase III positive control data (0.16mg/ml SLS) from each laboratory at T0 (pattern) and T4 (blank).	119
6.2.4. Correlation of FR ₈₅ (%) data with EU classifications.	175

Abbreviations

AB	alamar blue
ANOVA	analysis of variance
ATCC	American Type Culture Collection
BMedSci	Bachelor of medical science placement student
BRD	background review document
CAPB	cocamidopropylbetaine
CI	confidence interval
CMC	critical micelle concentration
COLIPA	European Cosmetic, Toiletry and Perfumery Association
CTAB	cetyl trimethyl ammonium bromide
CTFA	Cosmetics, Toiletries and Fragrance Association
CV	coefficient of variance
DDSW	double distilled sterile water
DMEM	Dulbecco's modified eagle medium
DMSO	dimethyl sulphoxide
EC	European Commission
EC50	effective concentration causing 50% fluorescein leakage
ECACC	European Collection of Cell Cultures
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
EC/HO	European Community/ Home Office
ECVAM	European Centre for the Validation of Alternative Methods
EGF	epidermal growth factor
EGTA	ethylenebis(oxyethylenenitrilo)tetraacetic acid
EMEM	Eagle's modified essential medium
EPA	Environmental Protection Agency
EU	European Union
FAL	FRAME Alternatives Laboratory
FBS	foetal bovine serum
FCS	foetal calf serum
FD	fluorescein isothiocyanate dextran
FHSA	Federal Hazardous Substances Act
FL	fluorescein leakage
FLT	fluorescein leakage test
FL%	percentage of fluorescein leakage for a set chemical concentration
FR	fluorescein retention
FRAME	Fund for the Replacement of Animals in Medical Experiments
GLP	good laboratory practice
h	hour
HBSS	hanks balanced salt solution
HCE-T	human corneal epithelium -transfected
hi	heat inactivated
HRIPT	human repeated insult patch testing
IC	inhibitory concentration

IIVS	Institute for <i>In Vitro</i> Sciences
IRAG	US Interagency Regulatory Alternatives Group
J-HCET	Japanese human corneal epithelium transfected
KB	kenacid blue
kD	kilodalton
LVET	low volume eye test
MAS	maximum average score
MDCK	Madin-Darby Canine Kidney
MEM	mininum essential medium
min	minute
MMAS	modified maximum average score
MTT	[3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide]
Na	sodium chloride
NR	3-amino-7-di-methylamino-2-methylphenazinehydrochloride
NRR	neutral red release
OD	optical density
OECD	Organisation for Economic and Community Development
OIC	ocular irritation categories
PBS	phosphate buffered solution
PI	prediction interval
PM	prediction model
PTFE	PolyTetraFluoroEthylene
RT	room temperature
RTECS	Registry of Toxic Effects of Chemical Substances
SCNM	study criteria not met
SD	standard deviation
SDS	sodium dodecyl sulphate
secs	seconds
SEM	standard error of the mean
SIS	science information service
SV40	simian virus 40
T	time-point
TEP	trans-epithelial permeability
TG	test guideline
w/o	without
w/v	weight/volume
w/w	weight/weight

1. Data collection

1.1. Description of the methods used to collect data, including literature searches or other sources, and number of studies collected

Initially, three internet-based scientific literature databases PubMed (www.ncbi.nlm.nih.gov/entrez/query.fcgi), Toxnet (www.toxnet.nlm.nih.gov), and Science Direct (www.sciencedirect.com), were searched in order to attain an initial overview of the work carried out using the fluorescein leakage (FL) assay. The following search terms were used; “fluorescein leakage assay,” “fluorescein leakage test,” and “trans-epithelium permeability assay.” The trans-epithelium permeability (TEP) assay and the fluorescein leakage assay are essentially the same and protocol information and results from the TEP assay are included in this background review document (BRD), i.e. general remarks pertaining to the FL assay are also applicable for the TEP assay unless otherwise stated. The number of publications from the different search terms in the various databases are shown (table 1.1.).

Table 1.1. Results of the different search terms in the various databases

Search term	Database		
	PubMed	Toxnet	Science Direct
fluorescein leakage assay	3 (129)	5 (6)	4 (4)
fluorescein leakage test	3 (132)	6 (8)	4 (4)
trans-epithelium permeability assay	0 (3)	0 (5)	0 (0)

Bold figures refer to the actual number of relevant papers, results in brackets refers to number of hits per search term. NB. It is to be noted that the Science Direct database required speech-marks around the search term in order to produce the relevant results.

In general, these databases led to credible scientific publications regarding the following types of FL assay information; assay development, protocol variations, and use of the assay for predicting *in vivo* eye irritation. The search terms failed to find a number of papers which were known feature the FL assay, e.g. the EC/HO International validation study on alternatives to the Draize eye irritation test (Balls *et al.*, 1995), and the USA Cosmetics, Toiletries and Fragrance Association (CTFA) Evaluation of alternatives program: an evaluation of *in vitro* alternatives to the Draize primary eye irritation test, Phase III (Gettings *et al.*, 1996).

The relevant papers found using the search terms featured in table 1.1. are listed for each database in chronological order:

Fluorescein leakage assay:

-Pubmed

Botham, P., Osborne, R., Atkinson, K., Carr, G., Cottin, M., van Buskirk, R.G., 1997. IRAG working group 3. Cell function-based assays. Interagency Regulatory Alternatives Group. Food and Chemical Toxicology, 35(1):67-77.

Jones, P. A., Budynsky, E., Cooper, K. J., Decker, D., Griffiths, H. A., Fentem, J. H., 2001. Comparative evaluation of five *in vitro* tests for assessing the eye irritation potential of hair-care products. ATLA, 29(6):669-92.

Clothier, R., Starzec, G., Pradel, L., Baxter, V., Jones, M., Cox, H., Noble, L., 2002. The prediction of human skin responses by using the combined *in vitro* fluorescein leakage/Alamar Blue (resazurin) assay. ATLA, 30(5):493-504.

-Toxnet

Gautheron, P., Duprat, P., Hollander, C.F., 1994. Investigations of the MDCK permeability assay as an *in vitro* test of ocular irritancy. In Vitro Toxicology, 7:33-43.

Botham, P., Osborne, R., Atkinson, K., Carr, G., Cottin, M., van Buskirk, R.G., 1997. IRAG working group 3. Cell function-based assays. Interagency Regulatory Alternatives Group. Food and Chemical Toxicology, 35(1):67-77.

Ward, R.K., Mungall, S., Carter, J., Clothier, R.H., 1997. Evaluation of tissue culture insert membrane compatibility in the fluorescein leakage assay. Toxicology In Vitro, 11:761-768.

Clothier, R.H., Starzec, G., Stipho, S., Kwong, Y.C., 1999. Assessment of initial damage and recovery following exposure of MDCK cells to an irritant. Toxicology In Vitro, 13:713-717.

Zanvit, A., Meunier, P. A., Clothier, R., Ward, R., Buiatti-Tcheng, M., 1999. Ocular irritancy assessment of cosmetics formulations and ingredients: fluorescein leakage test. Toxicology In Vitro, 13:385-391.

-Science Direct

Shaw, A. J., Balls, M., Clothier, R.H., Bateman, N.D., 1991. Predicting ocular irritancy and recovery for injury using MDCK cells. Toxicology In Vitro, 5:569-571

Botham, P., Osborne, R., Atkinson, K., Carr, G., Cottin, M., van Buskirk, R.G., 1997. IRAG working group 3. Cell function-based assays. Interagency Regulatory Alternatives Group. Food and Chemical Toxicology, 35(1):67-77.

Ward, R.K., Mungall, S., Carter, J., Clothier, R.H., 1997. Evaluation of tissue culture insert membrane compatibility in the fluorescein leakage assay. Toxicology In Vitro, 11:761-768.

Clothier, R.H., Starzec, G., Stipho, S., Kwong, Y.C., 1999. Assessment of initial damage and recovery following exposure of MDCK cells to an irritant. *Toxicology In Vitro*, 13:713-717.

Fluorescein leakage test:

-Pubmed

Botham, P., Osborne, R., Atkinson, K., Carr, G., Cottin, M., van Buskirk, R.G., 1997. IRAG working group 3. Cell function-based assays. Interagency Regulatory Alternatives Group. *Food and Chemical Toxicology*, 35(1):67-77.

Jones, P.A., Budynsky, E., Cooper, K. J., Decker, D., Griffiths, H.A., Fentem, J.H., 2001. Comparative evaluation of five *in vitro* tests for assessing the eye irritation potential of hair-care products. *ATLA*, 29(6):669-92.

Clothier, R., Starzec, G., Pradel, L., Baxter, V., Jones, M., Cox, H., Noble, L., 2002. The prediction of human skin responses by using the combined *in vitro* fluorescein leakage/Alamar Blue (resazurin) assay. *ATLA*, 30(5):493-504.

-Toxnet

Clothier, R.H., Morgan, S.J., Atkinson, K.A., Garle, M.J., Balls, M., 1994. Development of a fixed-dose approach for the fluorescein leakage test. *Toxicology In Vitro*, 8: 883-884.

Gautheron, P., Duprat, P., Hollander, C.F., 1994. Investigations of the MDCK permeability assay as an *in vitro* test of ocular irritancy. *In Vitro Toxicology*, 7:33-43.

Botham, P., Osborne, R., Atkinson, K., Carr, G., Cottin, M., van Buskirk, R.G., 1997. IRAG working group 3. Cell function-based assays. Interagency Regulatory Alternatives Group. *Food and Chemical Toxicology*, 35(1):67-77.

Ward, R.K., Mungall, S., Carter, J., Clothier, R.H., 1997. Evaluation of tissue culture insert membrane compatibility in the fluorescein leakage assay. *Toxicology In Vitro*, 11:761-768.

Cottin, M., Zanvit, A., 1999. Fluorescein leakage test: A useful tool in ocular safety assessment. *Toxicology In Vitro*, 11:399-405.

Zanvit, A., Meunier, P. A., Clothier, R., Ward, R., Buiatti-Tcheng, M., 1999. Ocular irritancy assessment of cosmetics formulations and ingredients: fluorescein leakage test. *Toxicology In Vitro*, 13: 385-391.

-Science Direct

Clothier, R.H., Morgan, S.J., Atkinson, K.A., Garle, M.J., Balls, M., 1994. Development of a fixed-dose approach for the fluorescein leakage test. *Toxicology In Vitro*, 8: 883-884.

Brantom, P.G., Bruner, L.H., Chamberlain, M., deSilva, O., Dupuis, J., Earl, L.K., Lovell, D.P., Pape, W.J.W., Uttley, M., Bagley, D.M., Baker, F.W., Brachter, M., Courtellemont, P., Declercq, L., Freeman, S., Steiling, W., Walker, A.P., Carr, G.J., Dami, N., Thomas, G., Harbell, J., Jones, P.A., Pfannenbecker, U., Southee, J.A., Tcheng, M., Argembeaux, H., Castelli, D., Clothier, R., Esdaile, D.J., Itigaki, H., Jung, K., Kasai, Y., Kojima, H., Kristen, U., Larnicol, M., Lewis, R.W., Marenus, K., Moreno, O., Peterson, A., Rasmussen, E. S., Robles, C., Stern, M., 1997. A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. *Toxicology In Vitro*, 11(N1-2): 141-179.

Cottin, M., Zanvit, A., 1999. Fluorescein leakage test: A useful tool in ocular safety assessment. *Toxicology In Vitro*, 11:399-405.

Zanvit, A., Meunier, P. A., Clothier, R., Ward, R., Buiatti-Tcheng, M., 1999. Ocular irritancy assessment of cosmetics formulations and ingredients: fluorescein leakage test. *Toxicology In Vitro*, 13: 385-391.

Trans-epithelium permeability assay:

-no relevant publications found in any of the databases

An internet search using the 'Google' search engine was performed in order to access a greater range of material concerning the FL assay, such as conference abstracts and journal comments. From these pieces a list of companies and organisations which use, or have used the FL assay were compiled. The Google search engine was also helpful in acquiring information pertaining to the field of *in vitro* testing for ocular irritation. Protocol information and/or data acquired using the Google search engine were only used in this BRD if they led to credible sources, e.g. journal websites or official company/institute websites known in the field of *in vitro* science. For example, one TEP assay INVITTOX Protocol and three FL assay INVITTOX Protocols were taken from the ECVAM Science Information System (SIS) website (<http://ecvam-sis.jrc.it>) which required registration to access the on-line protocols.

The internet searches were very useful for quickly gaining an overview of the publications featuring the FL assay. The journal articles were particularly useful for acquiring protocol information. In comparison, the FL assay data featured in journals had a tendency to be summarised to varying degrees, e.g. from mean \pm SD of repeated experiments, to statistical correlations regarding performance. Where contact details were available and current, the authors of all FL assay related publications were approached and invited to submit 'complete' protocol information and raw data. A standard letter and questionnaire was sent to the various companies and research institutes (Appendix I).

The majority of raw data featured in this BRD were primarily taken from the Fund for the Replacement of Alternative Medical Experiments (FRAME) in-house database. This raw data were produced as a result of the FRAME Alternatives Laboratory's (FAL) participation in FL assay validation and evaluation studies. Additional raw data were given by those companies which submitted data and protocol information to be used in this BRD. As internet searches only revealed a small number of companies and institutes that have reported their use of the FL assay, many other research

divisions of cosmetic and pharmaceutical companies were contacted. Of 17 companies contacted, two responded as not having FL assay protocol information or data to submit, 14 did not respond at all, and two agreed to participate in the study. Protocol information and data were provided by Company # 3 and Company # 4. Additional TEP assay data from Company # 3 which was part of the CTFA study Phase III was also submitted through collaboration with the CTFA, 1101 17th Street, NW Suite 300, Washington D.C., 20036-4702, USA. From all the available sources, 33 data sets generated using various FL assay protocols were collected and feature in this BRD.

The publications that were obtained and reviewed for this BRD are listed in table 1.2. It is acknowledged that many of these publications were not discovered as a result of the search terms shown in table 1.1.. Many of the publications were acquired by reviewing the reference sections of those papers obtained as a result of the initial searches as described in table 1.1.. Additional papers were also obtained due to the authors, of this BRD, awareness of relevant publications regarding the FL assay.

1.2. Brief description of data collected on overall study management

Table 1.2. Description of studies featuring FL assay protocol information and/or data, in chronological order

Study name	Study management	Chemical selection, chemical coding?	Process of data collection and statistical analysis	Avail. of standard prediction model	Quality of data (<i>in vivo</i> and <i>in vitro</i> GLP compliant)	Avail. of standard protocol*	Data format in publication
Trans-epithelial Permeability of Fluorescein <i>In Vitro</i> as an Assay to Determine Eye Irritants (Tchao, 1988). Poster	In-house	In-house. No chemical coding	In-house	N	<i>In vitro</i> : non-GLP	N	Tween concentrations producing similar FL; amount of FL not stated. FL summarised in graphs T0
Loss of Trans-epithelial Impermeability of a Confluent Layer of Madin-Darby Canine Kidney (MDCK) Cells as a Determinant of Ocular Irritancy Potential (Shaw <i>et al.</i> , 1990)	In-house	In-house. No chemical coding	In-house	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP	N (<i>INVITTOX Protocol No. 71</i>)	Mean FL20, FL50 (mg/ml) \pm SEM T0
Predicting Ocular Irritancy and Recovery from Injury using Madin-Darby Canine Kidney Cells (Shaw <i>et al.</i> , 1991)	In-house	In-house. No chemical coding	In-house	N	<i>In vitro</i> : non-GLP	N (<i>INVITTOX Protocol No. 71</i>)	Mean FL20 (mg/ml) \pm SEM T0, T72
Human Corneal Epithelial Primary Cultures and Cell Lines with Extended Life Span: <i>In Vitro</i> Model for Ocular Studies (Kahn <i>et al.</i> , 1993)	In-house	N/A –no chemicals tested	In-house	N	<i>In vitro</i> : non-GLP, <i>In vivo</i> : GLP	N	Mean FR% \pm SD or SEM T0 summarised in graphs

Study name	Study management	Chemical selection, chemical coding?	Process of data collection and statistical analysis	Avail. of standard prediction model	Quality of data (<i>in vivo</i> and <i>in vitro</i> GLP compliant)	Avail. of standard protocol*	Data format in publication
Investigations of the MDCK Permeability Assay as an <i>In Vitro</i> Test of Ocular Irritancy (Gautheron <i>et al.</i> , 1994)	In-house	In-house. No chemical coding	In-house	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP (historical data from Gautheron <i>et al.</i> , (1994); Kennah <i>et al.</i> , (1989) Grant , (1986)	N	Mean FL20, FL50 (mg/ml) w/o SD or SEM T0
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994)	In-house	Surfactants in-house, formulations independent; formulations coded	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP not stated	N (INVITTOX Protocol No. 71)	Mean FL (mg/ml), (%) ± SEM T0, T24, T48 and raw data
Development of a Fixed Dose Approach for The Fluorescein Leakage Test (Clothier <i>et al.</i> , 1994)	In-house	In-house. Chemical coding	In-house	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP (historical data from Botham <i>et al.</i> , (1989))	N (INVITTOX Protocol No. 82)	Mean FL% ± SEM T0, T72
The Evaluation of Pesticide Ingredients and Formulations <i>In Vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995)	In-house	In-house. No chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP (OECD Guidelines for Testing of Chemicals (1987) No. 45 "Acute Eye Irritation/ Corrosion)	INVITTOX Protocol No. 71	Mean FL20 (mg/ml) ±SEM T0, T72

Study name	Study management	Chemical selection, chemical coding?	Process of data collection and statistical analysis	Avail. of standard prediction model	Quality of data (<i>in vivo</i> and <i>in vitro</i> GLP compliant)	Avail. of standard protocol*	Data format in publication
The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test (Balls <i>et al.</i> , 1995)	Independent	Independent. Chemical coding	Independent -analyses performed by BIBRA.	Y	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP	Y -not strictly followed (INVITTOX Protocol No. 71)	FL20 (mg/ml) correlations from 4 labs with <i>in vivo</i> MMAS data T0
CTFA Evaluation of Alternatives Program: An Evaluation of <i>In Vitro</i> Alternatives to the Draize Primary Eye Irritation Test: Phase III (Gettings <i>et al.</i> , 1996)	Independent	Independent. Chemical coding	Independent	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP, concurrent <i>in vivo</i> and <i>in vitro</i> testing	N (INVITTOX Protocol No. 86)	Mean EC50 (%) w/o SD or SEM T0
Effects of Surfactant Re-treatment <i>In Vitro</i> : A Method to Evaluate Changes in Cell Junctions and Cell Viability (Clothier and Sansom, 1996)	In-house	In-house. No chemical coding	In-house	N	<i>In vitro</i> : non-GLP	N (-adaptation of INVITTOX Protocol No. 71 and No. 80)	Mean FL% ±SD T0, T1, T24, T72 summarised in graphs
Evaluation of Tissue Culture Insert Membrane Compatibility in the Fluorescein Leakage Assay (Ward <i>et al.</i> , 1997a)	In-house	In-house. Chemical coding for EC/HO test chemicals only	In-house	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP (historical data from Balls <i>et al.</i> , (1995); ECETOC (1992))	N (INVITTOX Protocol No. 71)	Mean FL20 (mg/ml) ±SD T0, T72
Evaluation of Chemically Induced Toxicity Using an <i>In Vitro</i> Human Corneal Epithelium (Ward <i>et al.</i> , 1997b)	In-house	In-house. No chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP, <i>In vivo</i> : GLP not stated	N	Mean TEP relative fluorescein retention (%) ± SD on graphs T0, T24, T48

Study name	Study management	Chemical selection, chemical coding?	Process of data collection and statistical analysis	Avail. of standard prediction model	Quality of data (<i>in vivo</i> and <i>in vitro</i> GLP compliant)	Avail. of standard protocol*	Data format in publication
Fluorescein Leakage Test: a Useful Tool in Ocular Safety Assessment (Cottin and Zarvit, 1997)	In-house	In-house. No chemical coding	In-house	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : according to Officiel de la Republique Francais, 24/10/1984.	N	Mean FL10 FL20 (mg/ml), w/o SD or SEM T0, T24
LAB A IRAG Working Group 3: Cell Function-based Assays (Botham <i>et al.</i> , 1997)	Independent	Unknown. Chemical coding unknown	In-house	N	<i>In vitro</i> : GLP stated in publication as unknown <i>In vivo</i> : GLP (historical data)	N	Pearson's correlation coefficients for <i>in vitro-in vivo</i> data T0
LAB B IRAG Working Group 3: Cell Function-based Assays (Botham <i>et al.</i> , 1997)	Independent (one data set from CTFA Phase III (Gettings <i>et al.</i> , 1996)	Independent for CTFA data set, unknown for 2nd data set. Chemical coding	In-house	N	<i>In vitro</i> : GLP not stated; <i>In vivo</i> : GLP for CTFA data; 2nd data set GLP stated in publication as unknown.	N	Pearson's correlation coefficients for <i>in vitro-in vivo</i> data T0
A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom <i>et al.</i> , 1997)	Independent-including selection of participating laboratories.	Independent. Chemical coding	Independent	Y - protocol common to all labs	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP	N (INVITTOX Protocol No.120)	Mean FL20 (mg/ml) w/o SD or SEM T4 from both labs summarised in <i>in vitro- in vivo</i> correlation graphs

Study name	Study management	Chemical selection, chemical coding?	Process of data collection and statistical analysis	Avail. of standard prediction model	Quality of data (<i>in vivo</i> and <i>in vitro</i> GLP compliant)	Avail. of standard protocol*	Data format in publication
Evaluation of a Human Corneal Epithelial Cell Line as an <i>In Vitro</i> Model for Predicting Ocular Irritation (Kruszewski <i>et al.</i> , 1997)	In-house	In-house. No chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP, <i>In vivo</i> : GLP not stated	N	Mean FR85(%) T0
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase II) (Southee, 1998)	Independent	Independent. No chemical coding	Independent -analyses performed by BIBRA.	Y- (based on Brantom <i>et al.</i> , (1997))	<i>In vitro</i> : non-GLP	N	Raw FL20 (mg/ml) and mean FL20 (mg/ml) T0, T4 summarised in graphs.
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	Independent	Independent. Chemical coding	Independent -analyses performed by BIBRA.	Y- (based on Brantom <i>et al.</i> , (1997))	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP unknown (historical data from Company # 1 and Gautheron <i>et al.</i> , (1994))	N (<i>INVITTOX Protocol No.120</i>)	Raw FL20 (mg/ml) and mean FL20 (mg/ml) T0, T4 summarised in graphs.
Assessment of Initial Damage and Recovery Following Exposure of MDCK Cells to an Irritant (Clothier <i>et al.</i> , 1999)	In-house	In-house. No chemical coding	In-house	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP (historical data from ECETOC, (1992))	N- combined FL/AB assay	Mean FL15 (mg/ml) ± SD; concentration response curves T0

Study name	Study management	Chemical selection, chemical coding?	Process of data collection and statistical analysis	Avail. of standard prediction model	Quality of data (<i>in vivo</i> and <i>in vitro</i> GLP compliant)	Avail. of standardised protocol*	Data format in publication
Ocular Irritancy Assessment of Cosmetics Formulations and Ingredients: Fluorescein Leakage Test. (Zanvit <i>et al.</i> , 1999)	Independent	Independent. Chemical coding	Independent	Y	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP	N (<i>INVITTOX Protocol No.120</i>)	Correlation graphs for <i>in vitro</i> and <i>in vivo</i> classifications. Mean FL20 (mg/ml) values w/o SD or SEM for 9 chemicals only. T0 or T4
Comparative Evaluation of Five <i>In Vitro</i> Tests for Assessing the Eye Irritation Potential of Hair-care Products (Jones <i>et al.</i> , 2001)	In-house	In-house. Formulations coded in publication.	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP; <i>In vivo</i> : GLP not stated	N	Mean FL% w/o SD or SEM T0, T72
The Prediction of Human Skin Responses by using the Combined <i>In Vitro</i> Fluorescein Leakage/Alamar Blue (Resazurin) Assay. (Clothier <i>et al.</i> , 2002)	In-house	Independent (provided by various companies). Chemical coding	In-house	Y	<i>In vitro</i> : non-GLP; <i>In vivo</i> : human occluded patch test 1h, 24h	N -combined FL/AB assay	FL% summarised in graphs T0

FRAME In-house data

Study name	Study management	Chemical selection, chemical coding?	Process of data collection and statistical analysis	Avail. of standard prediction model	Quality of data (<i>in vivo</i> and <i>in vitro</i> GLP compliant)	Avail. of standardised protocol*	Data format in publication
Company # 4 (FRAME, circa 1991)	In-house	Independent. Chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP	N	Mean FL20 (mg/ml) ± SEM T0, T4, T24, T48, T72
FRAME- Report on Comparison of 40 Cosmetic and Domestic Formulations Supplied by Company # 8 and Evaluated by 3 <i>In Vitro</i> Cytotoxicity Tests (FRAME, 1992)	In-house	Independent. Chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP	INVITTOX Protocol No.71	Raw data- FL20, FL50 (mg/ml), T0
Company # 5 (FRAME, 1992)	In-house	Independent. Chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP	INVITTOX Protocol No. 71	Raw data- FL20, FL50 (mg/ml), T0, T72
Final Report on Testing of 12 Mild Surfactants supplied by Company # 5 for Cytotoxicity Testing at the FAL (FRAME, 1992)	In-house	Independent. Chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP	INVITTOX Protocol No. 82	Raw data - FL20 (mMI) T0, T4, T24, T48, T72
Company # 5 (FRAME, 1993)	In-house	Independent. Chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP	INVITTOX Protocol No. 82	Raw data - FL% T0, T4, T24, T72

Study name	Study management	Chemical selection, chemical coding?	Process of data collection and statistical analysis	Avail. of standard prediction model	Quality of data (<i>in vivo</i> and <i>in vitro</i> GLP compliant)	Avail. of standardised protocol*	Data format in publication
Five Company # 5 Baby Products (FRAME, 1994)	In-house	Independent. Chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP	INVITTOX Protocol No. 82	Raw data - FL%, T0, T4, T24, T48, T72
Results from the Fixed Dose Fluorescein Leakage <i>In Vitro</i> Cytotoxicity Test on 4 Company # 5 Test Sample Formulations (FRAME, 1994)	In-house	Independent. Chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP	INVITTOX Protocol No. 82	Raw data - FL% T0, T4, T24, T48, T72
Sainsburys, Effects of 6 Coded Chemicals on MDCK Cells (FRAME, 1998)	In-house	Independent. Chemical coding	In-house; no statistical analysis	N	<i>In vitro</i> : non-GLP	N	Raw data - FL20 (mg/ml) T0, T72

*. INVITTOX Protocols detailed in parentheses were assigned based on the authors of this BRDs knowledge of the protocol. Many FL assay protocols described in the literature were accepted as INVITTOX Protocols at a later date. GLP= good laboratory practice; SD= standard deviation; SEM= standard error of the mean; T= time-point; w/o= without

2. Test Definition (Module 1)

2.1. Rationale for the proposed test method

2.1.1. Intended uses/ purpose

The FL assay was designed to measure the chemical-induced loss of trans-membrane impermeability of a confluent epithelial monolayer. Damage to inter-cellular adhesion molecules is an important event in chemical-induced eye irritation, and can be modelled using the FL assay. The FL assay was specifically developed to detect potentially mild and moderate irritants to the human eye, which are often cosmetic products.

The FL assay is performed on a confluent monolayer of non-replicating cells which are grown on permeable inserts. Madin-Darby Canine Kidney (MDCK) renal tubular epithelial cells are typically used as they form tight junctions and desmosomes similar to those found on the apical side of conjunctivae and corneal epithelia (INVITTOX Protocol No. 71). Test materials are applied to the cells growing on the apical side of the insert for a short exposure, e.g. one minute or 15 minutes. The test material is then removed and fluorescein dye added to the insert; the amount of dye that passes through the monolayer and insert within a given time is recorded. Due to the short exposure periods, the FL assay generally measures the effects of relatively high chemical concentrations. FL assay data differs from many other *in vitro* cytotoxicity assays which measure the effects of relatively low chemical concentrations and longer exposures on cell viability and replication rates. An advantage of the short incubation period used in the FL assay is that water-based ingredients and formulations can be tested neat if they can be easily removed after the short exposure period. This allows more direct comparisons of the FL assay results with the chemical effects *in vivo*, which is the endpoint assessed by regulatory authorities. As the FL assay can be repeated on the same cells for up to 72h, delayed effects and recovery can be measured in addition to immediate acute effects. Recovery is an important part of a chemical's toxicity profile which is also assessed by the *in vivo* ocular irritation test.

Throughout Europe and the USA, the FL assay is used in industry, as a screening step to detect potential eye irritants, in the early developmental phase of product ingredients and formulations. Within a test battery, the FL assay is particularly useful for comparing and distinguishing mild chemicals/formulations with very similar eye irritation potencies (INVITTOX Protocol No. 86). A FL assay protocol is used routinely by Company # 3, for predicting the eye irritation potential of surfactant-based cosmetic products. Surfactants are surface-active agents which exert immediate damage at the site of contact, including damage to cell junctions, which is measured by the FL assay. The data generated using the FL assay are used by the company for in-house assessments and do not constitute any basis for regulatory testing.

There are four principal FL protocols which have been accepted as INVITTOX Protocols; 'The Fluorescein Leakage Test' (INVITTOX Protocol No. 71), 'Fixed Dose Procedure for the Fluorescein Leakage Test' (INVITTOX Protocol No. 82), 'Trans-epithelial Permeability Assay' (INVITTOX Protocol No. 86), 'The Fluorescein Leakage Test -SOP of Company # 4 (INVITTOX Protocol No. 120) (Annex I). With the exception of INVITTOX Protocol No. 82, these protocols have featured, or provided the basis for similar protocols, in large-scale international validation and/or evaluation studies; The EC/HO International validation study on alternatives to the Draize eye irritation test (Balls *et al.*, 1995), CTFA Evaluation of alternatives program: an evaluation of *in vitro*

alternatives to the Draize primary eye irritation test (Gettings *et al.*, 1996), the European Cosmetic Toiletry and Perfumery Association (COLIPA) study on alternatives to the Draize rabbit eye irritation test (Brantom *et al.*, 1997), the Interagency Regulatory Alternatives Group (IRAG) Working Group 3: Cell-function based assays (IRAG) (Botham *et al.*, 1997). These studies assessed many *in vitro* assays, including the FL assay, to determine their abilities and applicabilities for replacing the current *in vivo* method for predicting chemical-induced *in vivo* eye irritation potential in humans. The studies evaluated *in vitro* assays which measured acute effects, rather than chronic effects or recovery.

2.1.2. Regulatory rationale and applicability

No FL assay protocol has been validated as a regulatory test for predicting *in vivo* ocular irritation. Scientific consensus is that no single *in vitro* cell-based test can replace the current *in vivo* test for eye irritation testing. The FL assay is recommended for use within a test battery of other *in vitro* tests for eye irritation, to enable the range of potencies and mechanisms of chemicals *in vivo* to be covered.

The validity of the FL assay for predicting chemical-induced *in vivo* eye irritation potential, led to it undergoing prevalidation (Southee, 1998). The prevalidation process took place in three phases; Phase I for protocol refinement, Phase II for protocol transfer, and Phase III for protocol predictivity. Phase I was successful in determining a single and appropriate FL assay protocol for testing in the later phases. A number of problems prevented the chosen FL assay protocol entering validation although it was concluded that the assay showed some relationship to the *in vivo* MMAS data with good inter-laboratory variability. On the basis of this study, the tested protocol was accepted as INVITTOX Protocol No. 120.

2.1.3. Scientific basis for the test (mechanistic)

The FL assay was developed by Tchao (1988) as a model for detecting materials that are potentially irritating to the eye. *In vivo*, the tight junctions and desmosomes of the corneal epithelium prevent solutes and foreign materials moving into the cornea. Solute in the cornea can induce water to move by osmosis into the cornea, thus causing oedema. Loss of trans-epithelial impermeability, due to damaged tight junctions and desmosomes, is one of the early events in chemical-induced ocular irritation. A confluent layer of MDCK cells consists of inter-cellular tight junctions and desmosomes. The confluent monolayer used in the FL assay is non-proliferating which models the non-proliferating state of the *in vivo* corneal epithelium. Whilst desmosomes maintain cell to cell adhesion, tight junctions form between adjacent cells and form a permeability barrier that can prevent the movement of molecules as small as 350MW. Tight junctions are found at the apical surfaces of conjunctiva, corneal and skin epithelia. It is assumed that a significant part of ocular irritation is related to the state and ability of the corneal epithelium to act as a barrier against foreign and potentially irritant materials (Botham *et al.*, 1997). Increasing the permeability of the corneal epithelium *in vivo* has been found to accompany the inflammation and surface damage observed when eye irritation develops (Igarashi, *et al.*, 1989; Ward *et al.*, 1997b).

General FL assay method outline

A confluent layer of MDCK cells is grown on the semi-permeable membrane of an insert; the inserts are placed into the wells of 96 or 24 well plates (figure 2.1.3.1.).

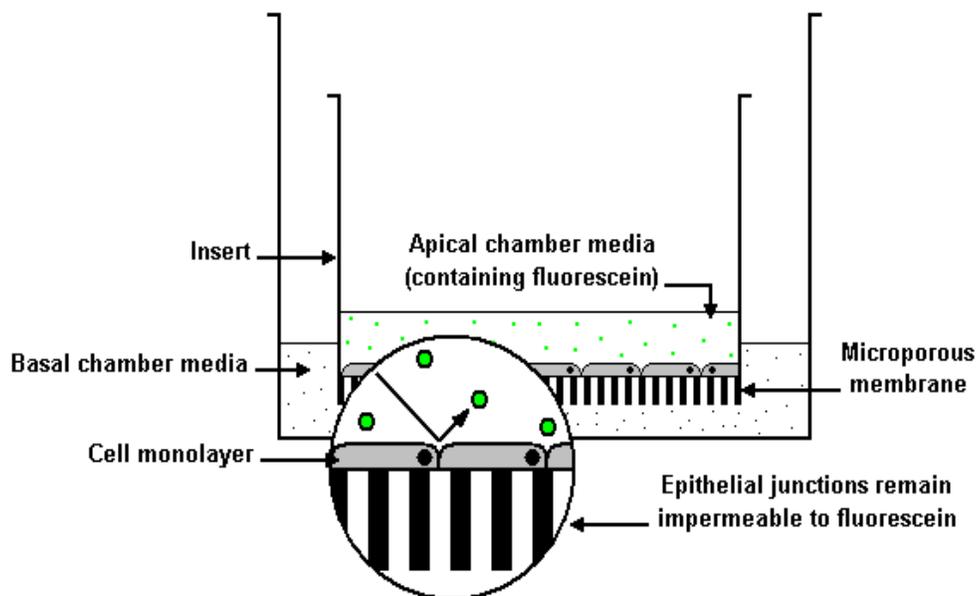


Figure 2.1.3.1. Diagram of MDCK cells grown on an insert membrane for the FL assay (from Wilkinson, 2006).

MDCK cells grown on porous membranes orientate correctly with prominent apical microvilli and cellular adhesion molecules including tight junctions and desmosomal junctions. As the MDCK monolayer forms tight junctions similar to those found in the corneal epithelium, the FL assay is considered an appropriate model for detecting damage to the tight junctions of the cornea. The cells are exposed to the test material for a short period (~one minute), which can be sufficient time for damage to occur to the tight junctions and desmosomes of the monolayer. The test material is then removed and the non-toxic, highly fluorescent sodium-fluorescein dye (Chan and Hayes, 1994) is added to the apical side of the monolayer for a defined period (~30-60 minutes). Sodium-fluorescein has a molecular weight of 376.3, which is too large to pass through the tight junctions that prevent the passage of molecules with molecular weights ≥ 350 . Chemicals that disrupt the tight junctions of the monolayer allow sodium-fluorescein dye to 'leak' through the monolayer and the porous membrane into the well. The amount of sodium-fluorescein that leaks to the basal side of the monolayer, within a specified time period, is measured. Sodium-fluorescein dye can be measured spectrophotometrically at 490nm or spectrofluorometrically at excitation and emission respective wavelengths of 485/530nm.

FL is calculated in reference to readings from a blank control where a confluent monolayer is not treated and a maximum leakage control where the FL through an insert without cells is measured (figure 2.1.3.2.).

$$FL_D = [(A-B)/(C-B)] \times (M_C - M_B) + M_B$$

D = % inhibition.

A = % damage (20% or 50% fluorescein leakage)

B = % fluorescein leakage < A

C = % fluorescein leakage > A

M_C = Concentration (mg/ml) of C

M_B = Concentration (mg/ml) of B

Figure 2.1.3.2. Equation for the calculation of the chemical/formulation concentration causing 20 percent FL (FL₂₀) (from INVITTOX Protocol No. 71).

FL₂₀ is the endpoint typically used in the various FL assay protocols. Specific damage to the cellular adhesion molecules occurs before damage to cell membranes and intracellular contents. Thus, the chemical concentrations inducing damage predominately to the tight junctions are possibly more accurately identified by FL₂₀ rather than FL₅₀, which is more likely to include non-specific cell damage. The amount of FL is proportional to the chemical-induced damage to the tight junctions, desmosomes and cell membranes. The damage to the tight junctions and desmosomes, induced by a test material gives an indication of its eye irritation potential *in vivo*. The tight junctions of MDCK cells can sometimes be restored following initial damage, if returned to fresh culture medium (Clothier *et al.*, 1994). As the sodium-fluorescein dye is non-toxic at the concentration used, the cells can be re-cultured after the FL assay has been performed. The FL assay can then be repeated at later time-points in order to assess recovery following the initial exposure to the test material. The speed at which FL reduces following the chemical exposure, relates to the rate at which the monolayer re-gains trans-epithelial impermeability. Recovery is an important and relevant endpoint to assess as rapid recovery of trans-epithelial impermeability *in vivo* prevents further exposures and secondary insults.

Like other cell based assays, the FL assay is reported not to have a good predictive capacity for test materials that have high acidity or alkalinity, fixative properties and reactivity with medium contents. The ocular irritation potential of materials that are water soluble and/or where the toxic effect is not affected by dilution are generally predicted accurately using the FL assay. In comparison to other *in vitro* tests, a concern unique to the FL assay is that damage to the insert membranes can occur due to the multiple rinsing steps which feature in most protocols. Routine observations of the cells on the insert membranes throughout the exposure and the recovery phases should detect damaged confluent layers and/or insert membranes and prevent mis-interpretation of results.

Table 2.1.3. Summary of the physicochemical properties that the FL assay is (in)capable of modelling.

Physicochemical Property	FL assay capable of testing materials with this physicochemical property?
Alcohol	Y
Fixative	N
Extreme pH	Y
Gases	N
Liquids	Y (if aqueous soluble)
Solid materials	Y (if aqueous soluble, but cannot be tested in its solid form)
Emulsions	Y
Granular materials	Y (if forms stable emulsion)
Suspensions	Impaired***
Coloured materials	Impaired**
Toxicity affected by dilution	N*
Highly viscous materials	Impaired
Volatile materials	Impaired
Reactive chemistries	N
Hydrophobic/lipophilic chemicals	N
Neat concentrations of chemicals	Y
MW > 350	N

* the FL assay is unable to measure the toxicity of chemicals that have basic toxic mechanisms which are affected by dilution.

** the FL assay is able to measure coloured materials which can be fully removed from the insert following the chemical exposure and therefore do not interfere with OD readings.

*** solid materials suspended in liquid have the propensity to precipitate out and the final concentration exposed to cells can be difficult to determine.

The FL assay has predominately been used to test surfactant and surfactant-based materials. The TEP assay (INVITTOX Protocol No. 86) is used in-house at Company # 3 for testing surfactants and surfactant-based formulations only. The FL assay protocol featured in the COLIPA study was used to test only surfactants and surfactant-based formulations as a prediction model (PM) was only available for surfactant-based materials (Zanvit *et al.*, 1999). The TEP assay was evaluated in the CTFA study Phase III, where surfactant-based formulations were tested (Gettings *et al.*, 1996). Surfactants are found in the majority of cosmetic and house-hold products and therefore constitute a wide-ranging and relevant chemical class for ocular irritation testing. Surfactants are generally mild materials, as determined by Draize test scores, which cause immediate damage at the site of contact, including damage to the expression of adhesion molecules. There is limited data regarding the predictive capacity of the FL assay for other chemical classes. Shaw *et al.*, (1990, 1991) reported a good correlation between FL₂₀ results and *in vivo* ocular irritation data for 22 chemicals, which included 13 chemicals also tested in the EC pilot study (anon, 1991). Clothier *et al.*, (1994) stated that the EU risk phrase classifications for the 21 chemicals tested in the EC pilot study were accurately predicted using the Fixed Dose FL assay (INVITTOX Protocol No. 82).

As the cells used in the FL assay are representative of corneal epithelial cells only, the assay is limited in the range of its predictive capacity. Although the FL assay does not model all types of possible corneal ocular irritation measured by the *in vivo* test, i.e. damage to the corneal stroma, it can measure effects that impact on chemical-induced stromal damage. For example, sodium dodecyl sulphate (SDS) is hypothesised to cause corneal opacity by permeating across the corneal epithelium to the stroma (Tchao, 1988).

In summary, the FL assay has many features rendering it a suitable as an *in vitro* model for predicting *in vivo* ocular irritation, e.g., relevant, cell types, chemical concentrations, exposures, and endpoint. The FL assay is particularly advantageous in that it allows effects to be measured which occur prior to, or even independent of cell death. Often, cell death does not occur in the cases of mild irritation and therefore a sensitive assay like the FL assay is essential. The assay is particularly useful as it allows recovery to be measured up to 72h following an initial exposure.

2.1.4. Similarities and differences of modes of action in the test method and the reference species

Most regulatory authorities require eye irritation data from the *in vivo* ocular irritation test developed by Draize *et al.*, (1944), e.g. OECD TG 405 (2002), EU Commission Directive 2004/73/EC (2004). Generally, the albino rabbit is the reference animal for *in vivo* eye irritation testing; healthy young adult male and female albino rabbits are used. The albino rabbit is used due to its relatively large eyes which make adverse effects easier to observe. Rabbits are used despite their ocular physiology known to be considerably different from that of humans.

The Draize test for ocular irritation (Draize *et al.*, 1944) subjectively scores the chemical-induced effects, specific to the cornea, conjunctiva and iris (figure 2.1.4.). The corneal epithelium is responsible for excluding foreign material from the surrounding environment and is therefore an important factor in protecting the eye from irritant responses (Curren and Harbell, 1998). An *in vitro* system which models the barrier function of the corneal epithelium, such as the FL assay is required if the Draize test is to be replaced. In the Draize scoring system, the cornea has the greatest weighting ascribed to the scores, indicating that these effects are the most important in terms of the resulting overall chemical-induced eye irritation; the number of days for the effects to clear from all the different tissues is also scored (figure 2.1.4.).

The conjunctiva is the external layer of the eye which is most vulnerable to the external environment. It is a non-keratinised squamous epithelium with secreting cells which covers the exposed surface of the eyeball and the inner surface of the eyelids (Atkinson, 1993; Curren and Harbell, 1998). It also has glands which provide moisture and secrete the components of the tear film. Foreign materials in the conjunctiva can induce inflammation. Inflammation is noted when the network of blood vessels in the conjunctiva dilate and produce an appearance of increased redness. Oedema can then occur due to changes in the capillary permeability of the blood vessels which allows fluids to leak into the interstitial spaces. Severe oedema can occur to an extent which impairs closure of the eye lid. Mild irritants can cause conjunctivitis without coupled damage to the cornea (Curren and Harbell, 1998). Despite the different degrees of conjunctival damage that can occur, chemical-induced effects have less weighting than

the cornea and iris, in the Draize scoring system. This is primarily due to the reversibility of chemical-induced effects to the conjunctiva.

The iris is situated under the cornea within aqueous humour. It is a vascular structure comprised of loose connective tissue, muscle and pigmented cells (Atkinson, 1993). Through muscular control of pupil dilation it regulates the amount of light that enters the eye and reaches the retina. Irritation can cause the vessels to dilate and leak vascular fluid which can cause oedema. Protein leakage from the vessels can cause aqueous flare which alters the refractive index of the aqueous humour (Atkinson, 1993). Severe damage can cause iris tissue destruction, resulting in the iris being unresponsive to light (Curren and Harbell, 1998).

I. Cornea

A. Opacity-Degree of Density (area which is most dense is taken for reading)

No opacity	0
Scattered or diffuse area-details of iris clearly visible	1
Easily discernible translucent areas, details of iris slightly obscured	2
Opalescent areas, no details of iris visible, size of pupil barely discernible	3
Opaque, iris invisible	4

B. Area of Cornea Involved

One quarter (or less) but not zero	1
Greater than one quarter, less than one-half	2
Greater than one half, less than three quarters	3
Greater than three quarters, up to whole area	4

Score equals A x B x 5 Total maximum = 80

II. Iris

A. Values

Normal	0
Folds above normal, congestion, swelling, circumcorneal injection (any one or all of these or combination of any thereof), iris still reacting to light (sluggish reaction is positive)	1
No reaction to light, haemorrhage; gross destruction (any one or all of these)	2

Score equals A x 5 Total maximum = 10

III. Conjunctivae

A. Redness (refers to palpebral conjunctivae only)

Vessels normal	0
Vessels definitely injected above normal	1
More diffuse, deeper crimson red, individual vessels not easily discernible	2
Diffuse beefy red	3

B. Chemosis

No swelling	0
Any swelling above normal (includes nictating membrane)	1
Obvious swelling with partial eversion of the lids	2
Swelling with lids about half closed	3
Swelling with lids about half closed to completely closed	4

C. Discharge

No discharge	0
Any amount different from normal (does not include small amount observed in inner canthus of normal animals)	1
Discharge with moistening of the lids and hairs just adjacent to the lids	2
Discharge with moistening of the lids and considerable area around the eye	3

Score (A + B + C) x 2 Total maximum = 20

Figure 2.1.4. Weighted scoring system for ocular lesions (from Draize *et al.*, 1944).

Although cell-based *in vitro* models are unable to attain the complexity of the *in vivo* situation, a number of cell lines that are not ocular in origin have been found to produce cytotoxic data, that can be used to predict the *in vivo* eye irritation potential of test materials (Borenfreund and Borrero, 1984).

At an ECVAM Expert Meeting (Scott *et al.*, (*under preparation*)) four different types of eye irritation processes were proposed as; membrane lysis, coagulation, saponification, and reactive chemistries. A list is given of the events that occur in ocular irritation which also incorporates the eye irritation processes proposed (Scott *et al.*, (*under preparation*)) (in bold font); the latter effects are the most severe (table 2.1.4.1.). Loss of trans-epithelial permeability is an important step in the eye irritation process, and it occurs before those mechanisms that were highlighted at the ECVAM Expert Meeting (Scott *et al.*, (*under preparation*)). The extent and duration of loss of trans-epithelial impermeability impact on the mechanisms defined by Scott *et al.*, (*under preparation*).

Table 2.1.4.1. Summary of the events involved in chemical-induced eye irritation *in vivo* which are (not) modelled by the FL assay.

Event involved in chemical-induced eye irritation	Modelled by the FL assay
Chemical interaction with tear film (Klyce and Beuerman, 1988; Hackett and McDonald, 1994)	N
Chemical binding to the conjunctival epithelium (Hackett and McDonald, 1994; Hogan and Zimmerman, 1962)	N
Adhesion molecules compromised (Farquhar and Palade, 1963; van Meer <i>et al.</i> , 1992; Katahira <i>et al.</i> , 1997)	Y
Corneal epithelium damage (<i>Dua et al.</i> , 1994)	Y
* inhibition of receptor-mediated membrane transport (Dearman <i>et al.</i> , 2003)	Y
* compromise of cell membrane integrity of upper corneal epithelium (<i>Dua et al.</i> , 1994, Hackett and McDonald, 1994; Maurer and Parker, 1996)	Y
* cell membrane lysis of all corneal epithelium layers (Hackett and McDonald, 1994)	Y
Hydration of corneal stroma (Hackett and McDonald, 1994).	N
Cross-linking of proteins in corneal stroma (Butler and Hammond, 1980; Eurell <i>et al.</i> , 1991; Chan and Hayes, 1994)	N
<i>Erosion of corneal stroma</i> (Baldwin <i>et al.</i> , 1973; Hackett and McDonald, 1994; Maurer <i>et al.</i> , 1997)	N
<i>Cell damage to corneal epithelium and limbus</i> (Jacobs and Martens, 1990; Wilhelmus, 2001)	N
<i>Dilation and increased lymphatic leakage from scleral vasculature (oedema and erythema)</i> (Hackett and McDonald, 1994)	N
<i>Stimulation of nerve endings, i.e. enhanced blinking, tearing</i> (Chan and Hayes, 1994)	N
<i>Erosion of nerve endings in cornea and sclera</i> (Butler and Hammond 1980; Klyce and Beuerman, 1988; Araki <i>et al.</i> , 1994)	N
Duration of response, i.e. length of time cell responses deteriorate. Duration of response covers the effects of reactive chemicals which can cause coagulation , saponification , that are effects which develop and increase over time. (Hubert, 1992; Maurer and Parker, 1996)	Y
Recovery from response, i.e. length of time for cell responses to return to control levels (Hubert, 1992)	Y

Text in italics indicates irreversible responses.

Based upon these defined effects (table 2.1.4.1.) the FL assay is capable of detecting chemicals that induce the third and fourth elements of irritation described. Due to fundamental differences between the *in vitro* and *in vivo* situation, e.g., *in vitro* the cells are grown as a monolayer, and *in vivo* the cornea consists of 5-7 epithelial layers and

the conjunctiva 2-3 epithelial layers, care must be taken when interpreting the *in vitro* data. *In vitro* cell-based assays cannot model chemical-induced effects to the tear film. *In vivo*, chemical interaction with the tear film modifies the chemical concentration and exposure duration to the cornea. As the cornea is the tissue with the greatest weighting for effects in the Draize scoring system, the effect of the tear film has an important effect on chemical induced eye-irritation. Similar to all *in vitro* methods, systemic effects such as hormones, blood flow, and the immune system on chemical-induced toxicity cannot be replicated.

In addition to the immediate chemical-induced effects measured by the FL assay, it is also capable of measuring reversibility, delayed effects and recovery (Clothier *et al.*, 1994). *In vivo*, the corneal epithelium has demonstrated recovery following exposure to mild irritants (Clothier and Sansom, 1996). Recovery is an important aspect of the eye irritation process that does not always relate to the degree of initial insult (Clothier *et al.*, 1994). Use of the non-toxic (at the concentrations used) sodium-fluorescein dye allows recovery of the impermeable monolayer to be assessed. *In vitro*, recovery has also been observed in the FL assay over a 72h period following initial chemical exposure. Following the initial chemical insult, recovery is assessed by the reduction of FL over time. The ability of the FL assay to measure delayed effects and recovery is a unique property of this cell-based *in vitro* model. Recovery measured by the FL assay allows comparisons to the *in vivo* Draize test which measures 'days to clear.'

A proposal for categorising the principal mechanisms of common classes of test chemicals into the four categories of mechanisms that can cause ocular irritation was made during an ECVAM Expert Meeting for testing strategies (Scott *et al.*, (*under preparation*)) (table 2.1.4.2.).

Table 2.1.4.2. Examples of chemical classes and their effects through the different mechanisms categorised (Scott *et al.*, (*under preparation*)).

Membrane Lysis	Coagulation	Saponification	Reactive Chemistries
surfactants (all)	acids	alkaline materials	peroxides
organic solvents	-concentrated	-concentrated	mustards
ketones	-diluted	-diluted	alkyl halides
alcohols	-derivatives	-derivatives	epoxides
volatile liquids	cationic surfactants		bleaches (oxidisers)
ethers	organic solvents		
polyethers			
esters			
aromatic amines			

The majority of chemical classes exert membrane lysis (table 2.1.4.2.). In some cases, a chemical class can have more than one mechanism of action, i.e. cationic surfactants can cause cell damage via coagulation and membrane lysis (table 2.1.4.2.). Thus, the FL assay, which specifically detects membrane damage, should be capable of detecting such eye irritants regardless of the different chemical classes, but dependent on concentration. The chemical types for which the FL assay, and most cell-based *in vitro* assays, cannot measure potential eye irritation are generally those that are strong,

acids, alkalis, fixatives, and viscous. These chemicals have mechanisms that are not measured by the FL assay, e.g. extensive coagulation, saponification or specific reactive chemistries. Chemicals known to have a high ocular irritation potential, through testing for skin irritation, are normally assumed to be strong eye irritants, without additional *in vivo* testing specifically for eye irritation. The FL assay, designed to detect sub-lethal damage to cell membranes and tight junctions, is an appropriate *in vitro* test for detecting mild ocular irritants. Assuming that the FL assay is able to measure membrane damage induced by all the chemical classes which cause membrane lysis (as defined by Scott *et al.*, (*under preparation*)), the assay will have a large and relevant applicability domain. Overall, the FL assay allows for the testing of eye irritation induced by a wide range of chemicals.

The sodium-fluorescein dye used in the FL assay is non-toxic at the concentrations used and has also been used in humans to assess damage to the corneal epithelium (INVITTOX Protocol No. 71). *In vivo*, chemicals can adversely affect cell adhesion molecules, which then allow fluorescein to penetrate into the stroma. Fluorescein detected in the stroma indicates damaged areas of the epithelium (Igarashi, 1986; Chan and Hayes, 1994). FL not only occurs due to damaged adhesion molecules but also due to membrane damage. In this context, the FL assay does measure one of the mechanisms of eye irritation defined at the ECVAM Expert Meeting (Scott *et al.*, (*under preparation*)) that cannot be distinguished from the FL due to impaired adhesion molecules. In order to determine the type of cell damage that has led to trans-epithelial permeability, results from the FL assay have been compared with results from the Neutral Red Release (NRR) assay which specifically measures membrane damage (Hubbard *et al.*, 1994).

The effects of certain physiological features of the eye such as the tearing system, on resulting toxicity, are difficult to identify. Inherently variable lachrymation rates affect the concentration and duration of chemical exposures in the eye. It is difficult to identify actual *in vivo* exposure durations and concentrations, and replicate them for *in vitro* testing. It is estimated that following an accidental exposure, the foreign material is likely to remain in the eye for approximately 30 seconds before it is removed by blinking, tear formation and rinsing. In comparison to humans, rabbits have a slower blink rate and a less efficient tearing system (from Europeans for Medical Progress, www.curedisease.com: Problems with the Draize Test, accessed 23.08.05), which results in longer exposure durations in the rabbit eye in comparison to humans. The exposure period used in the FL assay is a compromise between reproducing the probable length of time of an *in vivo* accidental chemical exposure, whilst also allowing sufficient time for experimental procedures to be performed reproducibly. In comparison to most other cytotoxicity assays for eye irritation, the short exposure period allows the effects of high chemical concentrations to be measured, thus replicating the *in vivo* test situation in the rabbit. Unlike the *in vivo* situation, solids cannot be tested in the FL assay unless suspended in a liquid vehicle.

A problem associated with the short exposure period of the FL assay is the difficulty of efficiently removing the test materials after the short exposure period. This is particularly true for viscous materials, such as gels and creams which are the type of materials often tested using the FL assay. Due to the short exposure period, the mild materials often need to be tested neat in order to produce a response which can be measured. Therefore the problems associated with viscous materials cannot be reduced by dilution. Test materials can also bind to the insert membrane, thus making

their removal very difficult. Chemical binding to the insert membrane, is more common for cationic materials, such as benzalkonium chloride, which are attracted to the positively charged membrane (Balls *et al.*, 1995). Increased washing steps to remove the test material from the insert can also lead to insert membrane damage and thus erroneous results. Alternatively, if test materials are not removed fully and/or efficiently, they can potentially physically block the passage of the sodium-fluorescein through the insert, which would cause chemical effects to be under-estimated. In general, additional uncontrolled exposure time is a greater proportion of the short exposure period of the FL assay, than with assays with longer exposures. This leads to greater variability in results, and low assay reproducibility. As the FL assay can be repeated at multiple time-points, erroneous results due to ineffective removal of the test material would be more likely detected in comparison to cell viability assays which use single time-points. The efficient removal of test agents from the eye is also a concern of the *in vivo* test.

2.2. Test method protocols

The major components of the test method protocol(s) featured in the literature and in-house studies are shown in chronological order and/or associated studies (table 2.2.). Full protocol information is provided in Appendix II

Table 2.2. Test method protocols

INVITTOX Protocols

Study/Company/ Organisation	Fluorescein Leakage Test -INVITTOX Protocol No. 71	Fixed Dose FLT -INVITTOX Protocol No. 82	Trans-epithelial permeability (TEP) Assay -INVITTOX Protocol No. 86
Protocol Used/Basis	Based on Tchao (1988)	Based on FLT INVITTOX Protocol No. 71	Validated in-house Company # 3
Cell Type/Strain	MDCK CB997 (ECACC: 84121903)	MDCK CB997 (ECACC: 84121903)	MDCK NBL-2 (ATCC: CCL 34)
Seeding Density (cells/ml)	4x10 ⁵ cells/ml (400µl/insert)	4x10 ⁵ cells/ml (400µl/insert)	5x10 ⁵ cells/ml (200µl/insert)
Passage Range defined?	N	N –but cells have been used upto passage 40	Cells used after three passages post-thawing; no upper limit stated
Time to attain confluency? (h)	96	96	48
Insert Type(s) and Pore Size(s)	Millicell-HA 12mm (0.45µm pore size; Anocell 10 inserts can also be used)	Anopore™ 10; 0.2µm or 0.02µm pore size	Costar 6.5mm Transwell tissue culture inserts; 0.45µm pore size
Duration of Exposure (mins)	1 or 15	1	15
Exposure at RT?	Y	not stated	N -incubator
Type(s) of Materials tested?	N/A	Chemicals from Comm. of European Commun. collaborative study on the evaluation of alternative methods to the eye irritation test (anon, 1991)	In-house testing of surfactant-based materials
Fluorescein Concentration delivered to monolayer (mg/ml)	0.1mg/ml Na-fluorescein in HBSS (0.01% (w/v)) (0.02% Na-fluorescein in HBSS for Anocell inserts)	0.02% Na-fluorescein in HBSS	0.02% Na-fluorescein in HBSS
Time allowed for Fluorescein Leakage (mins)	30 RT	60 RT	30 RT
Maximum Leakage Control (insert without cells) (Y/N)	Y	Y	N –material concentrations attain a plateau of maximum leakage
Negative Control	HBSS	HBSS	HBSS
Positive Control	100mg/ml Brij 35	130mg/ml glacial acetic acid	None -to be selected by Company # 3

Endpoint (e.g. FL10, FL20, %FL)	FL20, FL50 (mg/ml) T0; if FL20 is not reached the max FL% should be quoted alongside concentration causing leakage	FL% T0	EC50 % T0
Recovery Time-point(s) Following Initial Exposure (h)	N	4, 24, 48 and 72 (under-consideration)	N
Prediction Model	N	Y -50mg/ml differentiated irritants from non-irritants	EC50 <1.8% fail; EC50 ≥2.2% pass. EC50 1.8% -2.2% =borderline.

Study/Company/ Organisation	FLT -SOP Company # 4 INVITTOX Protocol No. 120 (developed as consequence of ECVAM Prevalidation Study, 1998)
Protocol Used/Basis	Based on FLT INVITTOX Protocol No. 71
Cell Type/Strain	MDCK NBL-2 (ATCC: CCL34)
Seeding Density (cells/ml)	4x10 ⁵ cells/ml (500µl/insert= 2x10 ⁵ cells/insert)
Passage Range defined?	3-30.
Time to attain confluency? (h)	96
Insert Type(s) and Pore Size(s)	Millicell-HA 12mm (0.45µm pore size)
Duration of Exposure (mins)	15
Exposure at RT?	Y
Type(s) of Materials tested?	surfactants and surfactant-based formulations
Fluorescein Concentration delivered to monolayer (mg/ml)	0.01mg/ml Na-fluorescein in HBSS (0.001% (w/v))
Time allowed for Fluorescein Leakage (mins)	30 RT
Maximum Leakage Control (insert without cells) (Y/N)	Y
Negative Control	HBSS
Positive Control	0.16mg/ml SDS in HBSS
Endpoint (e.g. FL10, FL20, %FL)	FL10, FL20 (mg/ml) T0, T4. Even if FL10 is attained, max FL% should be quoted with concentration causing FL.
Recovery Time- point(s) Following Initial Exposure (h)	4

Prediction Model	PM from COLIPA study for surfactants only; FL20 T4 >100mg/ml =non irritant/slight, MMAS < 15; 20-100mg/ml = moderate, MMAS 15-30; <20mg/ml= irritant/severe, > 30 MMAS
-------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Studies Reported

Study/Company/ Organisation	Trans-epithelial Permeability of Fluorescein <i>In Vitro</i> as an Assay to Determine Eye Irritants (Tchao, 1988) Poster	Loss of Trans-epithelial Impermeability of a Confluent Layer of Madin-Darby Canine Kidney (MDCK) Cells as a Determinant of Ocular Irritancy Potential (Shaw <i>et al.</i> , 1990)	Predicting Ocular Irritancy and Recovery from Injury using Madin-Darby Canine Kidney Cells (Shaw <i>et al.</i> , 1991)
Protocol Used/Basis	N	-	-
Cell Type/Strain	MDCK	MDCK CB997 (ECACC: 84121903)	MDCK
Seeding Density (cells/ml)	1.5x10 ⁵ cells/insert	10 ⁵ cells/insert (in 400µl medium)	Cells grown to confluence
Passage Range defined?	N	not stated	not stated
Time to attain confluency? (h)	72	96	96
Insert Type(s) and Pore Size(s)	Millicell with HATF (surfactant-free) membrane, 12mm diameter	Anocell 10	Anotec 10mm porous tissue culture insert
Duration of Exposure (mins)	15	1	1
Exposure at RT?	24°C	not stated	not stated
Type(s) of Materials tested?	surfactants	Chemicals with range of mechanistic activities and potencies; 13 taken from EC pilot study	16 chemicals with various mechanisms and potencies.
Fluorescein Concentration delivered to monolayer (mg/ml)	0.02% Na-fluorescein in HBSS	0.02% (w/v) fluorescein in HBSS	0.02% (w/v) fluorescein in HBSS
Time allowed for Fluorescein Leakage (mins)	30, 24°C	60 RT	60 RT
Maximum Leakage Control (insert without cells) (Y/N)	Not stated	Y	Y
Negative Control	HBSS	distilled water	distilled water
Positive Control	N/A	not stated	not stated
Endpoint (e.g. FL10, FL20, %FL)	Preliminary experiments -increases in FL observed	FL20 and FL50 (mg/ml) T0	FL20 (mg/ml) T0
Recovery Time-point(s) Following Initial Exposure (h)	N	N	72
Prediction Model	N	N	N

Study/Company/ Organisation	Company # 4 (FRAME, circa 1991)	Company # 4 In- house Fluorescein Leakage Test SOP (1992)	Human Corneal Epithelial Primary Cultures and Cell Lines with Extended Life Span: <i>In Vitro</i> Model for Ocular Studies (Kahn <i>et al.</i> , 1993)
Protocol Used/Basis	Shaw <i>et al.</i> , (1990)	Company # 4 (1992)	Stated measurements were according to Tchao (1988)
Cell Type/Strain	MDCK CB997 (ECACC: 84121903)	MDCK CB997 (ECACC: 84121903)	Human corneal epithelial cell line
Seeding Density (cells/ml)	1x10 ⁶ cells/cm ² insert	4x10 ⁵ cells/ml (500µl/insert= 2x10 ⁵ cells/insert)	2.4x10 ⁵ cells/cm ²
Passage Range defined?	Shaw <i>et al.</i> , (1990)	3-30.	1-5
Time to attain confluency? (h)	72-96	96	Not stated
Insert Type(s) and Pore Size(s)	Anotec 10	Millicell-HA 12mm, 0.45µm pore size	Cellagen disks CD24
Duration of Exposure (mins)	1	15	N/A –assay used to determine the impermeability of layer
Exposure at RT?	Shaw <i>et al.</i> , (1990)	not stated	Not stated
Type(s) of Materials tested?	4 surfactants and 3 Company # 4 products	N/A	N/A
Fluorescein Concentration delivered to monolayer (mg/ml)	200mg/L	10µg/ml Na-fluorescein in HBSS	0.02% Na-fluorescein in PBS
Time allowed for Fluorescein Leakage (mins)	60 RT	30 RT	30
Maximum Leakage Control (insert without cells) (Y/N)	Y	Y	Not stated
Negative Control	not stated	HBSS	N/A
Positive Control	acetic acid	0.16mg/ml SDS in HBSS	N/A
Endpoint (e.g. FL10, FL20, %FL)	FL20, FL50 (mg/ml)	FL20 (units not stated) T0, T4	FR%
Recovery Time- point(s) Following Initial Exposure (h)	72	4, 24, 48, 72	N/A
Prediction Model	N	N	N

Study/Company/ Organisation	Development of a Fixed Dose Approach for The Fluorescein Leakage Test. (Clothier <i>et al.</i>, 1994)	Investigations of the MDCK Permeability Assay as an <i>In Vitro</i> Test of Ocular Irritancy (Gautheron <i>et al.</i>, 1994)	Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i>, 1994)
Protocol Used/Basis	Method according to Shaw <i>et al.</i> , (1991)	Tchao (1988)	Shaw <i>et al.</i> , (1990)
Cell Type/Strain	MDCK (method according to Shaw <i>et al.</i> , (1991))	MDCK (Flow Laboratories)	Shaw <i>et al.</i> , (1990)
Seeding Density (cells/ml)	Method according to Shaw <i>et al.</i> , (1991)	1.8-2 x10 ⁵ cells/insert	Shaw <i>et al.</i> , (1990)
Passage Range defined?	Method according to Shaw <i>et al.</i> , (1991)	not stated	Shaw <i>et al.</i> , (1990)
Time to attain confluency? (h)	'based on exposure of a confluent layer of MDCK cells'	96	Shaw <i>et al.</i> , (1990)
Insert Type(s) and Pore Size(s)	Method according to Shaw <i>et al.</i> , (1991)	Millipore HA	Shaw <i>et al.</i> , (1990)
Duration of Exposure (mins)	Method according to Shaw <i>et al.</i> , (1991)	15	1
Exposure at RT?	Method according to Shaw <i>et al.</i> , (1991)	Y	Shaw <i>et al.</i> , (1990)
Type(s) of Materials tested?	21 chemicals from the EC/HO study	42 chemicals with range of chemical structures and irritancy potential	6 commercially available surfactants, 11 formulations (4 baby shampoos and 7 bath products)
Fluorescein Concentration delivered to monolayer (mg/ml)	Method according to Shaw <i>et al.</i> , (1991)	0.2mg/ml Na-fluorescein in HBSS (pH adjusted to pH 7.4)	Shaw <i>et al.</i> , (1990)
Time allowed for Fluorescein Leakage (mins)	60 RT	30 RT	Shaw <i>et al.</i> , (1990)
Maximum Leakage Control (insert without cells) (Y/N)	Method according to Shaw <i>et al.</i> , (1991)	Y	Shaw <i>et al.</i> , (1990)
Negative Control	Method according to Shaw <i>et al.</i> , (1991)	HBSS pH adjusted to 7.4	Shaw <i>et al.</i> , (1990)
Positive Control	Method according to Shaw <i>et al.</i> , (1991)	Not stated	Shaw <i>et al.</i> , (1990)
Endpoint (e.g. FL10, FL20, %FL)	FL% T0	FL20, FL50 (mg/ml) T0	FL50 (mg/ml) T0; FL% T0
Recovery Time-point(s) Following Initial Exposure (h)	72	N/A	24, 48

<p>Prediction Model</p>	<p>50mg/ml hypothesised to be the cut-off point to distinguish R36/R41 chemicals from NI if FL20% was taken to indicate significant toxicity.</p>	<p>N</p>	<p>N</p>
--------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------	----------	----------

Study/Company/ Organisation	The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test (Balls <i>et al.</i>, 1995)	The Evaluation of Pesticide Ingredients and Formulations <i>In Vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i>, 1995)	Effects of Surfactant Re-treatment <i>In Vitro</i>: A Method to Evaluate Changes in Cell Junctions and Cell Viability. (Clothier and Sansom, 1996)
Protocol Used/Basis	According to Tchao (1988)	INVITTOX Protocol No. 71	INVITTOX Protocol No. 71 (combined with AB assay)
Cell Type/Strain	MDCK (strain not stated)	MDCK (INVITTOX Protocol No. 71)	MDCK CB997
Seeding Density (cells/ml)	not stated	1x10 ⁷ cells/ml	60µl of 2x10 ⁵ cells/ml added to 7 or 8 wells of each strip
Passage Range defined?	not stated	INVITTOX Protocol No. 71	passage range 13<24 used
Time to attain confluency? (h)	not stated	INVITTOX Protocol No. 71	48-72
Insert Type(s) and Pore Size(s)	Millipore HA	Anopore 10	Anopore membrane, 0.2µm pore size
Duration of Exposure (mins)	1	1	1
Exposure at RT?	not stated	INVITTOX Protocol No. 71	INVITTOX Protocol No. 71 and No. 80
Type(s) of Materials tested?	60 chemicals ranging in mechanisms and potency for which historical <i>in vivo</i> data was available- data primarily from ECETOC database (1992)	Pesticide formulations, vehicles w/o pesticide ingredients, and pure pesticides.	CAPB
Fluorescein Concentration delivered to monolayer (mg/ml)	not stated	200mg/l Na-fluorescein in PBS	0.01% Na-fluorescein in a 1:10 dilution of AB in HBSS
Time allowed for Fluorescein Leakage (mins)	not stated	60 (RT unknown)	60 (incubator)
Maximum Leakage Control (insert without cells) (Y/N)	Y	Y	Y
Negative Control	not stated	not stated	HBSS
Positive Control	not stated	acetic acid	INVITTOX Protocol No. 71 and No. 80
Endpoint (e.g. FL10, FL20, %FL)	FL20 (mg/ml)	FL20 (mg/ml) T0, T72	FL% T0 AB% T0
Recovery Time-point(s) Following Initial Exposure (h)	N/A	72	(initial), then 1, 24, 72 after 1st and 2nd treatments

<p>Prediction Model</p>	<p>(not applied to published results) FL20 values: <100mg/ml= R36 or R41; >750mg/ml= NI; 100-750mg/ml= R41 if no recovery after 72h, R36 if recovery after 72h</p>	<p>N</p>	<p>N</p>
--------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------	----------

Study/Company/ Organisation	CTFA Evaluation of Alternatives Program: An Evaluation of <i>In Vitro</i> Alternatives to the Draize Primary Eye Irritation Test: Phase III (Gettings <i>et al.</i>, 1996)	Evaluation of a Human Corneal Epithelial Cell Line as an <i>In Vitro</i> Model for Predicting Ocular Irritation (Kruszewski <i>et al.</i>, 1997)	Evaluation of Chemically Induced Toxicity Using an <i>In Vitro</i> Human Corneal Epithelium (Ward <i>et al.</i>, 1997b)
Protocol Used/Basis	Based on Tchao (1988)	none	none
Cell Type/Strain	MDCK NBL-2 (ATCC: CCL 34)	HCE-T	HCE-T
Seeding Density (cells/ml)	1.5x10 ⁵ cells/insert	not stated	1.5-2x10 ⁵ cells/insert
Passage Range defined?	not stated	N	not stated but cell line only attains approx 20 passages
Time to attain confluency? (h)	48	168	168
Insert Type(s) and Pore Size(s)	Costar 6.5mm Transwell tissue culture inserts	Cellagen, 14mm collagen-membrane	Cellagen, 14mm collagen-membrane
Duration of Exposure (mins)	15	5	5
Exposure at RT?	not stated	37°C	37°C
Type(s) of Materials tested?	23 surfactant-based formulations	Surfactant formulations and twenty chemicals with various properties	Benzalkonium chloride, SDS, ethanol, isopropanol
Fluorescein Concentration delivered to monolayer (mg/ml)	0.02% Na-fluorescein in HBSS	0.02% (w/v) Na-fluorescein in high calcium KGM medium w/o growth supplements	0.02% (w/v) Na-fluorescein
Time allowed for Fluorescein Leakage (mins)	30 RT	30 Incubator	30 Incubator
Maximum Leakage Control (insert without cells) (Y/N)	not stated	Y	Y
Negative Control	HBSS -used as non-treated control	high calcium KGM	high calcium KGM
Positive Control	Not stated	Not stated	Not stated
Endpoint (e.g. FL10, FL20, %FL)	EC50 (%) T0	TEP assay relative FR85 (%)	TEP assay relative FR85 (%)
Recovery Time-point(s) Following Initial Exposure (h)	N	24	48
Prediction Model	Regression modelling of <i>in vitro</i> endpoint and <i>in vivo</i> data to enable prediction of MAS values for any formulation	N	N

Study/Company/ Organisation	IRAG Working Group 3: Cell Function-based Assays (Botham <i>et al.</i>, 1997) LAB A	IRAG Working Group 3: Cell Function-based Assays (Botham <i>et al.</i>, 1997) LAB B	Evaluation of Tissue Culture Insert Membrane Compatibility in the Fluorescein Leakage Assay (Ward <i>et al.</i>, 1997a)
Protocol Used/Basis	Based on Tchao (1988)	Based on Martin and Stott (1992)	Method modified from Shaw <i>et al.</i> , (1990)
Cell Type/Strain	MDCK (ECACC: 84121903)	MDCK (ECACC: 84121903)	MDCK CB997 (ECACC: 84121903)
Seeding Density (cells/ml)	not stated	not stated	2x10 ⁵ cells/insert
Passage Range defined?	not stated	not stated	Method modified from Shaw <i>et al.</i> , (1990)
Time to attain confluency? (h)	48-96	48-96	72
Insert Type(s) and Pore Size(s)	Millicell-HA	Costar Transwell tissue culture insert	Anopore with aluminium oxide membrane (Nunc), 0.2µm, 0.02µm pore sizes; Millicell-CM insert with PTFE Biopore membrane coated with human placental type IV collagen, 0.4µm pore size; Millicell-HA insert with membrane composed of mixed cellulose esters 0.45µm pore size (Millipore)
Duration of Exposure (mins)	15	15	1
Exposure at RT?	not stated	not stated	RT
Type(s) of Materials tested?	42 (14 surfactants, 16 alcohols, 12 other chemicals MAS range, 0-98)	25 surfactant-based personal care products (MAS 0-40) performed in modified Draize (anaesthetic) test and 28 shampoos with undefined modification to Draize	6 known surfactants
Fluorescein Concentration delivered to monolayer (mg/ml)	0.02% (w/v) Na- fluorescein in HBSS	0.02% (w/v) Na- fluorescein in HBSS	0.01% Na-fluorescein in HBSS
Time allowed for Fluorescein Leakage (mins)	30 RT	30 RT	30 RT
Maximum Leakage Control (insert without cells) (Y/N)	not stated	not stated	Y
Negative Control	not stated	not stated	HBSS
Positive Control	not stated	not stated	Method according to Shaw <i>et al.</i> , (1990)

Endpoint (e.g. FL10, FL20, %FL)	FL20 and FL50 (mg/ml) T0	FL50 (mg/ml) T0	FL20 (mg/ml) T0
Recovery Time-point(s) Following Initial Exposure (h)	N	N	72
Prediction Model	not stated	not stated	N

Study/Company/ Organisation	A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom <i>et al.</i>, 1997)	Ocular Irritancy Assessment of Cosmetics Formulations and Ingredients: Fluorescein Leakage Test. (Zanvit <i>et al.</i>, 1999) DATA FROM COLIPA	FLT- SOP Company # 4 (SOP used in the ECVAM Prevalidation Study 1996-1998) – submitted by Company # 4
Protocol Used/Basis	Protocol of Cottin <i>et al.</i> , (1992)	Company # 4 Test Protocol (1992)	Based on FLT Protocol INVITTOX Protocol No. 71
Cell Type/Strain	MDCK NBL-2	MDCK NBL-2	MDCK NBL-2
Seeding Density (cells/ml)	see entry for Zanvit <i>et al.</i> , (1999)	2x10 ⁵ cells/(not written but assume) insert	4x10 ⁵ cells/ml (500µl per insert= 2x10 ⁵ cells/insert)
Passage Range defined?	see entry for Zanvit <i>et al.</i> , (1999)	N -presence of tight junctions verified using immunofluorescence with ZO-1 antibody.	not stated
Time to attain confluency? (h)	see entry for Zanvit <i>et al.</i> , (1999)	96	4
Insert Type(s) and Pore Size(s)	see entry for Zanvit <i>et al.</i> , (1999)	Millipore HA mixed cellulose ester membranes	Millicell-HA 12mm, 0.4µM pore size
Duration of Exposure (mins)	15	15	15
Exposure at RT?	see entry for Zanvit <i>et al.</i> , (1999)	not stated	Y
Type(s) of Materials tested?	4 surfactants (diff concentrations) and 23 surfactant-based formulations soluble in HBSS; 30 materials tested by FRAME, 33 materials by Company # 4 -differed per laboratory	4 surfactants (diff concentrations) and 23 surfactant-based formulations soluble in HBSS; 30 materials tested by FRAME, 33 materials by Company # 4 -differed per laboratory	surfactants
Fluorescein Concentration delivered to monolayer (mg/ml)	see entry for Zanvit <i>et al.</i> , (1999)	10µg/ml Na-fluorescein	0.01mg/ml in HBSS (0.001% (w/v) Solution Na-fluorescein
Time allowed for Fluorescein Leakage (mins)	see entry for Zanvit <i>et al.</i> , (1999)	30 Incubator	30
Maximum Leakage Control (insert without cells) (Y/N)	see entry for Zanvit <i>et al.</i> , (1999)	Y	yes
Negative Control	see entry for Zanvit <i>et al.</i> , (1999)	not stated	HBSS
Positive Control	see entry for Zanvit <i>et al.</i> , (1999)	not stated	0.16mg/ml SDS

<p>Endpoint (e.g. FL10, FL20, %FL)</p>	<p>FL20 (mg/ml) T4</p>	<p>FL20 (mg/ml) at T0 or T4 calculated as specified in INVITTOX Protocol No. 71. If FL20 not reached, maximum FL% attained is noted along with the test concentration</p>	<p>FL10, FL20 (mg/ml) after 30mins and 4h since contact of test material. Even if FL10 is attained, max% FL should be quoted along with mg/ml conc. causing the FL</p>
<p>Recovery Time-point(s) Following Initial Exposure (h)</p>	<p>4</p>	<p>4</p>	<p>4 H</p>
<p>Prediction Model</p>	<p>PM for FL20 T4 values only. Classification-based PM for surfactant-based materials only, developed by Company # 4 (historical data of 43 surfactant ingredients and formulations for which <i>in vivo</i> data was available). FL20 (mg/ml) >100, non irritant/slight, < 15 MMAS; 20-100, moderate, 15-30 MMAS; <20, irritant/severe, >20 MMAS</p>	<p>PM for FL20 T4 values only. Classification-based PM for surfactant-based materials only, developed by Company # 4 (historical data of 43 surfactant ingredients and formulations for which <i>in vivo</i> data was available). FL20 (mg/ml) >100, non irritant/slight, < 15 MMAS; 20-100, moderate, 15-30 MMAS; <20, irritant/severe, >20 MMAS</p>	<p>PM as used in the COLIPA study; >100mg/ml =non irritant/slight, MMAS < 15; 20-100mg/ml = moderate, MMAS 15-30; <20mg/ml= irritant/severe, > 30 MMAS</p>

Study/Company/ Organisation	Fluorescein Leakage Test: a Useful Tool in Ocular Safety Assessment (Cottin and Zanvit, 1997)	Fluorescein Leakage Test: a useful tool in ocular safety assessment Toxicology in Vitro 11(1997)3 99-405 M. COTTIN – submitted by Company # 4
Protocol Used/Basis	N/A	FLT- SOP Company # 4
Cell Type/Strain	MDCK (ATCC: CCL 34)	MDCK NBL-2
Seeding Density (cells/ml)	2x10 ⁵ cells/insert	4x10⁵ cells/ml (500µl per insert= 2x10⁵ cells/insert)
Passage Range defined?	N	between 3rd and 30th after thawing
Time to attain confluency? (h)	96	4
Insert Type(s) and Pore Size(s)	Millicell-HA, 0.45µm pore size	Millicell-HA 12mm, 0.45µM pore size
Duration of Exposure (mins)	15	15
Exposure at RT?	Y	Y
Type(s) of Materials tested?	Surfactant and surfactant-based formulations	surfactants
Fluorescein Concentration delivered to monolayer (mg/ml)	10µg/ml Na-fluorescein in HBSS	0.01mg/ml in HBSS (0.001% (w/v) Solution Na-fluorescein
Time allowed for Fluorescein Leakage (mins)	30 RT	30 RT
Maximum Leakage Control (insert without cells) (Y/N)	Y	yes
Negative Control	HBSS	HBSS
Positive Control	not stated	0.16mg/ml SDS
Endpoint (e.g. FL10, FL20, %FL)	FL10, FL20 (mg/ml) T0, T4	FL20 (mg/ml) after 30mins, 4h and 72 h since contact of test material.
Recovery Time-point(s) Following Initial Exposure (h)	4, 24, 48, 72	72H
Prediction Model	FL20 <20mg/ml =irritant; FL20 ≥20mg/ml and <100mg/ml =moderately irritant; FL20 ≥100mg/ml = slightly irritant	No

Study/Company/ Organisation	Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase II) (Southee, 1998)	Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	Assessment of Initial Damage and Recovery Following Exposure of MDCK Cells to an Irritant (Clothier <i>et al.</i> , 1999)
Protocol Used/Basis	Prevalidation of the FL Assay based on INVITTOX Protocol No. 71	Prevalidation of the FL Assay based on INVITTOX Protocol No. 71	Based on INVITTOX Protocol No. 71 (combined with AB assay –INVITTOX Protocol No. 80).
Cell Type/Strain	MDCK NBL-2 (ECACC: 85011435)	MDCK NBL-2 (ECACC:85011435)	MDCK
Seeding Density (cells/ml)	2x10 ⁵ cells/ml	2x10 ⁵ cells/ml	8x10 ⁴ cells/insert
Passage Range defined?	3<30	3<30	<100
Time to attain confluency? (h)	96	96	according to INVITTOX Protocol No. 71
Insert Type(s) and Pore Size(s)	Millicell-HA 12mm diameter, 0.45µm pore size,	Millicell-HA 12mm diameter, 0.45µm pore size,	Anopore membrane, 0.2µm pore size
Duration of Exposure (mins)	15	15	5
Exposure at RT?	Y	Y	not stated
Type(s) of Materials tested?	5% triton x-100, CTAB, Company # 3 Baby Shampoo, glycerol, ammonium nitrate	Mild surfactants relevant to cosmetic testing with <i>in vivo</i> data available from BIBRA after obtaining from Company # 1 and Sigma.	3 known irritants
Fluorescein Concentration delivered to monolayer (mg/ml)	0.01% (w/v) fluorescein in HBSS with Ca ⁺⁺ and Mg ⁺⁺	0.01% (w/v) Na-fluorescein in HBSS with Ca ⁺⁺ and Mg ⁺⁺	0.01% Na-fluorescein in HBSS
Time allowed for Fluorescein Leakage (mins)	30±2 Incubator	30±2 Incubator	60 Incubator
Maximum Leakage Control (insert without cells) (Y/N)	Y	Y	not stated
Negative Control	HBSS	HBSS	HBSS
Positive Control	0.16mg/ml SDS in HBSS	0.16mg/ml SDS in HBSS	not stated
Endpoint (e.g. FL10, FL20, %FL)	FL20 (mg/ml) T0, T4	FL20 (mg/ml) T0, T4	FL15 (mg/ml) (in order to compare results with human corneal model endpoint of fluorescein retention 85%) T1.5
Recovery Time-point(s) Following Initial Exposure (h)	4	4	24, 48, 72, 96

<p>Prediction Model</p>	<p>Based upon COLIPA PM using FL20 T4 results:> 100mg/ml = non-irritant/slight; 20-100mg/ml = moderate; <20 mg/ml = irritating/severe.</p>	<p>Based upon COLIPA PM using FL20 T4 results:> 100mg/ml =non-irritant/slight; 20-100mg/ml =moderate; <20 mg/ml =irritating/severe</p>	<p>N/A</p>
--------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------

Study/Company/ Organisation	Comparative Evaluation of Five <i>In Vitro</i> Tests for Assessing the Eye Irritation Potential of Hair-care Products (Jones <i>et al.</i> , 2001)	The Prediction of Human Skin Responses by using the Combined <i>In Vitro</i> Fluorescein Leakage/Alamar Blue (Resazurin) Assay. (Clothier <i>et al.</i> , 2002)
Protocol Used/Basis	Modification of Tchao (1988)	Adaptation of Tchao (1988)
Cell Type/Strain	MDCK (obtained from R Tchao)	MDCK CB997
Seeding Density (cells/ml)	2x10 ⁵ cells/insert (500µl/insert)	5x10 ⁵ cells/ml, (200µl/insert)
Passage Range defined?	2-30.	1<30
Time to attain confluency? (h)	not stated (poss. 7 days)	72
Insert Type(s) and Pore Size(s)	Millicell HA, 0.45µm pore size	Millicell-HA 12mm, 0.45µm pore size; Anopore membranes 0.2µm pore size
Duration of Exposure (mins)	10secs shampoos, 30secs conditioners	1
Exposure at RT?	Y	not stated
Type(s) of Materials tested?	Shampoos and conditioners	handwash formulations, laundry detergent formulations and moisturisers
Fluorescein Concentration delivered to monolayer (mg/ml)	0.01% Na-fluorescein (w/v) in HBSS with Ca ²⁺ and Mg ²⁺	0.01% Na-fluorescein in HBSS
Time allowed for Fluorescein Leakage (mins)	30 RT	60 (55) Incubator
Maximum Leakage Control (insert without cells) (Y/N)	Y	Y
Negative Control	not stated	HBSS
Positive Control	not stated	1mg/ml CAPB (non-preserved 50% activity) or 0.16mg/ml SDS
Endpoint (e.g. FL10, FL20, %FL)	FL% T0, T72	FL%, AB% T0
Recovery Time-point(s) Following Initial Exposure (h)	24, 48, 72	4, 24, 48, 72

<p>Prediction Model</p>	<p>Results compared to the FL result for benchmark shampoo and conditioner. Material acceptable if same as benchmark; 'further investigation' defined by $\pm 20\%$ from benchmark value</p>	<p>Generated from the results for the handwash products, and evaluated with results from laundry powdered cleaners and the moisturisers. Criteria given for FL and AB results induced by various concentrations at different time-points for each level of <i>in vivo</i> score 0-4 in 0.5 increments.</p>
--------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Frame In-house Studies

Study/Company/ Organisation	Company # 4 (FRAME, circa 1991)	FRAME- Report on Comparison of 40 Cosmetic and Domestic Formulations Supplied by Company # 8 and Evaluated by 3 <i>In Vitro</i> Cytotoxicity Tests (FRAME, 1992)	Company # 5 Chemicals (FRAME, 1992)
Protocol Used/Basis	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Cell Type/Strain	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Seeding Density (cells/ml)	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Passage Range defined?	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Time to attain confluency? (h)	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Insert Type(s) and Pore Size(s)	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Duration of Exposure (mins)	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Exposure at RT?	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Type(s) of Materials tested?	4 surfactants and 3 finished products	40 cosmetics and detergent formulations	5 cosmetic ingredients and final product
Fluorescein Concentration delivered to monolayer (mg/ml)	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Time allowed for Fluorescein Leakage (mins)	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Maximum Leakage Control (insert without cells) (Y/N)	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Negative Control	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Positive Control	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.	INVITTOX Protocol No. 71.
Endpoint (e.g. FL10, FL20, %FL)	FL20. FL50 (mg/ml)	FL20, FL50 (mg/ml)	FL20, FL50 (mg/ml)
Recovery Time- point(s) Following Initial Exposure (h)	4, 24, 48, 72	N/A	72
Prediction Model	N	N	N

Study/Company/ Organisation	Final Report on Testing of 12 Mild Surfactants supplied by Company # 5 for Cytotoxicity Testing at the FAL (FRAME, 1992)	Company # 5 (FRAME, 1993)	Five Company # 5 Products (FRAME, 1994)
Protocol Used/Basis	Shaw <i>et al.</i> , (1990, 1991)/ INVITTOX Protocol No. 82	INVITTOX Protocol No. 82.	Shaw <i>et al.</i> , (1990, 1991)
Cell Type/Strain	MDCK (assume CB997)	INVITTOX Protocol No. 82.	MDCK (assume CB997)
Seeding Density (cells/ml)	4x10 ⁵ cells/insert (400µl/insert)	INVITTOX Protocol No. 82.	4x10 ⁵ cells/insert (400µl/insert)
Passage Range defined?	not stated	INVITTOX Protocol No. 82.	15-20
Time to attain confluency? (h)	96	INVITTOX Protocol No. 82.	96
Insert Type(s) and Pore Size(s)	Anocell, 0.2µm pore size	INVITTOX Protocol No. 82.	Anocell, 0.2µm pore size
Duration of Exposure (mins)	1	INVITTOX Protocol No. 82.	1
Exposure at RT?	INVITTOX Protocol No. 82.	INVITTOX Protocol No. 82.	INVITTOX Protocol No. 82.
Type(s) of Materials tested?	12 mild surfactants	Chemicals remain coded	5 baby care products-water in oil and oil in water emulsions
Fluorescein Concentration delivered to monolayer (mg/ml)	not stated	INVITTOX Protocol No. 82.	not stated
Time allowed for Fluorescein Leakage (mins)	30	INVITTOX Protocol No. 82.	INVITTOX Protocol No. 82.
Maximum Leakage Control (insert without cells) (Y/N)	Y	INVITTOX Protocol No. 82.	INVITTOX Protocol No. 82.
Negative Control	INVITTOX Protocol No. 82.	INVITTOX Protocol No. 82.	INVITTOX Protocol No. 82.
Positive Control	130mg/ml acetic acid	INVITTOX Protocol No. 82.	130mg/ml acetic acid
Endpoint (e.g. FL10, FL20, %FL)	FL%	FL%	FL%
Recovery Time-point(s) Following Initial Exposure (h)	4, 24, 48, 72	24, 48, 72	24, 48
Prediction Model	N/A	N	N/A

Study/Company/ Organisation	Results from the Fixed Dose fluorescein leakage <i>In Vitro</i> Cytotoxicity test on 4 test Sample Formulations from Company # 5 (FRAME, 1994)	Company # 6, Effects of 6 Coded Chemicals on MDCK Cells (FRAME, 1998)
Protocol Used/Basis	Shaw <i>et al.</i> , (1990, 1991)	not stated
Cell Type/Strain	MDCK CB997	MDCK NBL-2 (ECACC)
Seeding Density (cells/ml)	4x10 ⁵ cells/insert (400µl/insert)	2x10 ⁵ cells/insert (400µl/insert)
Passage Range defined?	N	not stated
Time to attain confluency? (h)	96	96
Insert Type(s) and Pore Size(s)	Anocell, 0.2µm pore size	Anopore or Millipore
Duration of Exposure (mins)	2	1 or 15
Exposure at RT?	not stated	not stated
Type(s) of Materials tested?	Company # 5 samples	Company # 6 coded samples
Fluorescein Concentration delivered to monolayer (mg/ml)	not stated	0.01% (w/v) fluorescein solution
Time allowed for Fluorescein Leakage (mins)	60 RT	60 RT
Maximum Leakage Control (insert without cells) (Y/N)	Y	Y
Negative Control	Y	HBSS
Positive Control	130mg/ml acetic acid	130mg/ml acetic acid
Endpoint (e.g. FL10, FL20, %FL)	FL%	FL20 (mg/ml)
Recovery Time- point(s) Following Initial Exposure (h)	4, 24, 48, 72	4, 72
Prediction Model	N/A	N/A

Red font indicates protocol information received from companies directly. CAPB= cocamidopropylbetaine; CTAB =cetyl trimethyl ammonium bromide; ECACC =European Collection of Cell Cultures; FLT =fluorescein leakage test; GLP= good laboratory practice; HCE-T =human corneal epithelium-transfected; NI= non-irritant; RT =room temperature; SD =standard deviation; SEM =standard error of the mean; T=time-point; WP =well plate(s).

2.2.1. Description of protocol components and rationale for differences, if available

The following sections detail the various protocol components of the FL assay and provide rationales for the protocol differences. Details of the ECVAM Prevalidation study (Southee, 1998) which compared various protocol elements are given where appropriate.

i. Cell type

Two MDCK cell-strains exist and feature equally in the FL assay INVITTOX protocols; INVITTOX Protocols No. 71 and No. 82 employ the MDCK cell strain CB997, whilst INVITTOX Protocols No. 86 and No. 120 use the NBL-2 cell strain. The NBL-2 cells were more frequently cited in the USA publications whilst the CB997 cells were commonly used in Europe-based studies.

INVITTOX Protocol No. 120 was developed following the ECVAM sponsored FL assay Prevalidation study (Southee, 1998). Phase I of the ECVAM Prevalidation study undertook a number of experiments for protocol refinement. The protocol variables investigated were; MDCK cell strain, growth medium and insert membrane. The aim of the study was to determine which combination produced maximum barrier function (0% leakage).

As NBL-2 cells were reported to form a better barrier than CB997 cells, the optimum growth medium for this cell strain was investigated. Both MDCK cell strains were grown in the following medium combinations:

- Minimum Essential Medium (MEM) + 10% Foetal Bovine Serum (FBS) + 1% L-glutamine
- Dulbecco's Modified Eagles Medium (DMEM)/F12 (1:1) with 15mM Hepes + 10% heat inactivated FBS + 1% L-glutamine.

MEM supplemented with 10% FBS and 1% glutamine was found to be the optimum growth medium for the NBL-2 cell strain; the different types of media used in the various INVITTOX protocols supports this finding. INVITTOX Protocol No. 71 and No. 82 culture MDCK CB997 cells in DMEM/Nutrient Mix F12 (1x concentrate with L-glutamine and 15mM Hepes), supplied by Gibco (UK). INVITTOX Protocol No. 86 cultures MDCK NBL-2 cells in Eagle's MEM made up in Earle's BSS (supplier not stated). In INVITTOX Protocol No. 120, MDCK NBL-2 cells are cultured in modified Eagle's medium with Earle's salts (2x concentrate with L-glutamine and without phenol red) supplied by Gibco (USA). Use of different media according to the cell strain cultured was supported by the literature (Clothier *et al.*, 1994; Shaw *et al.*, 1990; 1991; Zanvit *et al.*, 1999). All protocols supplement the medium with 10% FCS/FBS.

The FL Assay Prevalidation study (Southee, 1998) tested the growth of both MDCK cell strains on Millipore-HA, Costar Transwell and Nunc Anopore inserts to determine which promoted optimum growth and barrier function. Cells were treated with 0.16mg/ml SDS and the amount of FL(%) at T0 and T4 was recorded. The combination giving the most consistent and complete barrier function was MDCK NBL-2 cells grown on Millipore-HA inserts and cultured in MEM supplemented medium. Both cell strains grown on

Anopore membranes and treated with SDS had a tendency to lift and peel away from the membrane.

The TEP assay has been performed using a human corneal epithelial cell line developed by Gillette Medical Evaluation Laboratories (Kruszewski *et al.*, 1997; Ward *et al.*, 1997b). Human corneal epithelium primary cultures transfected with the SV40 large T antigen plasmid pRSV-T (Ward *et al.*, 1997b) produced a cell line (HCE-T) that expresses human corneal epithelia features for approximately 24 passages (Kruszewski *et al.*, 1997). The advantages of using the HCE-T cell line, for predicting damage to the human corneal epithelium, is that it forms a similar number of layers as found in the *in vivo* situation and the cells are of human origin. A human corneal epithelial cell line is also used at the FAL. In comparison to the HCE-T cell line, the J-HCET cell line used at FAL expresses similar properties to normal human corneal epithelial cells up to approximately 100 passages (Araki-Sasaki *et al.*, 1995).

Bridging Study:

In order to replicate the *in vivo* situation more closely, the SV40 transfected Japanese human corneal epithelial cell line (J-HCET) has been used at the FAL. These immortalised human corneal epithelial cells stratify 4-6 layers when grown on collagen membranes at air-liquid interfaces. In comparison to MDCK cells, J-HCET cells grew less satisfactorily on the inserts and generally took a longer time to form confluent layers (Cheah; BMedSci, 2000).

The 'J-HCET protocol' was used to predict the effects of repeated exposures to low doses of surfactants on the corneal epithelium barrier. The J-HCET cell model which consisted of stratified cell layers was less sensitive than the MDCK cell model which formed a monolayer (Clothier *et al.*, 1999). Various protocol modifications were made to obtain reproducible results with the J-HCET cell line; J-HCET cells were grown on 0.4µm Nunc polycarbonate membrane inserts whereas MDCK cells were grown on 0.2µm Anopore inserts. Fluorescein Isothiocyanate Dextran (FD) with molecular weights of 4.4kD (FD-4) and 9.5kD (FD-9) were used in conjunction with J-HCET cells as sodium-fluorescein gave variable results and a greater rate of FL than with MDCK cells, suggesting that the intra-cellular spaces of the tight junctions were larger in the J-HCET cell line. The FL assay in the J-HCET model ranked the four surfactants in agreement with *in vivo* rankings (Cheah; BMedSci, 2000). Further work was deemed necessary in order to determine the use of this cell model for predicting *in vivo* ocular irritation (Cheah; BMedSci, 2000).

More recent studies have shown that in the correct culture conditions the J-HCET cells can generate an equally tight junction complex and para-cellular cleft as found with MDCK cells, i.e. a 6% maximum of the no cells control can be applied as for the MDCK cells (Moore *et al.*, 2005; Wilkinson and Clothier, 2005).

ii. Cell seeding

Seeding density is important to ensure that the cells attain a confluent monolayer before the FL assay is performed. A confluent layer is needed to enable tight junctions to form between the cells and produce an impermeable layer to the sodium-fluorescein dye.

iii. Cell passage number

The cell passage number is important to ensure that the cells used for the FL assay function similarly. INVITTOX Protocol No. 120 reports that the MDCK cells should be within the passage range 3-30 from thawing. Many other publications regarding the FL assay have also stated that experiments were performed with cells within this passage range. Cells within this passage range have similar functionality which aids assay results to be reproducible.

iv. Chemical solvent

All protocols use HBSS or deionised water rather than medium as the chemical solvent. Some protocols stipulate that the HBSS used is without phenol red. Phenol red could potentially remain in the wells and interfere with the sodium-fluorescein optical density (OD) readings by the spectrofluorimeter. HBSS is used rather than medium to reduce the effects of the medium components interfering with the test material, and also to reduce the possibility of the medium components binding to, or disrupting the insert membrane. Using HBSS as the solvent reduces the likelihood of variation due to unpredictable properties and effects of the medium. Mineral oil is used for insoluble materials.

INVITTOX Protocol No. 120 states that the HBSS should also contain calcium. Calcium is important to maintain the tight junctions. INVITTOX Protocol No. 71 does not stipulate that HBSS should contain calcium as it was not so commonly available at the time the protocol was developed. The author of INVITTOX Protocol No. 71 now states that HBSS preferably should contain calcium, although as the exposure period for INVITTOX Protocol No. 71 is shorter than INVITTOX Protocol No. 120, the absence of calcium from HBSS would have a lesser adverse effect. Also, some residue of cell culture medium which contains calcium is likely to remain in the well during the chemical exposure.

v. Washing steps

In the literature, washing steps are often performed before the test chemical is added to the cells. Of the INVITTOX protocols, only INVITTOX Protocol No. 86 uses a washing procedure before adding the test chemical. With increased experience with the FL assay it became apparent that a washing step before the chemical exposure increased assay reproducibility by reducing the variable effect of test chemical interactions with medium components (personal comm. R Clothier). HBSS is commonly used as the washing solution although distilled water features in some protocols. The number of times the washing is repeated varies between protocols from 1 to 5. The advantage of multiple washing steps to increase method reproducibility, is offset by the greater chance of damage to the cell monolayer and/or insert membrane.

Washing steps after the test material is removed and before sodium-fluorescein is added to the cell monolayer feature consistently in the INVITTOX Protocols. HBSS is commonly used as the washing solution for one or two rinses, and PBS or distilled water less so. INVITTOX Protocol No. 86 features ten rinses. Efficient washing is necessary to ensure the test material is fully removed and that uncontrolled prolonged exposures do not occur, which could impact on the reproducibility and predictivity of the assay. If the test chemical remains in the insert, it has the potential to physically block the sodium-fluorescein dye from leaking through the cell monolayer and insert membrane,

thus leading to an under-estimation of the chemical's toxicity. Also, if the chemical remains in the well when sodium-fluorescein is added, unknown prolonged exposures and uncontrolled interactions between the chemical and the sodium-fluorescein dye can occur that can impact on the reproducibility and predictivity of the assay, e.g. binding of the two solutions or binding between solutions and the membrane.

vi. Duration of chemical exposure

The duration of the test material exposure is usually one minute or 15 minutes. INVITTOX Protocol No. 71 uses both exposures according to the potency of material being tested; although no guidance is given for determining which material exposure period to use, the one minute exposure is commonly reported in the literature for this protocol. INVITTOX Protocol No. 82 uses a one minute chemical exposure period. This protocol differs from others as it determines the amount of FL (%) caused by exposure to a fixed chemical concentration. INVITTOX Protocol No. 86 and INVITTOX Protocol No. 120 use a 15 minute chemical exposure period as the NBL-2 cell strain is known to form a greater number of tight junctions, and therefore form a more impermeable monolayer than the CB997 cell strain used in INVITTOX Protocols No. 71 and No. 82.

vii. Sodium-fluorescein dye concentration

The sodium-fluorescein dye is used by all INVITTOX protocols and with the exception of studies conducted using human corneal cells, features consistently in the literature. The concentration of the sodium-fluorescein dye varies according to each protocol; INVITTOX Protocol No. 71 and No. 120 use 0.01% sodium-fluorescein whereas INVITTOX Protocol No. 82 and No. 86 use 0.02% sodium-fluorescein. INVITTOX Protocol No. 71 states that 0.02% sodium-fluorescein maybe used when Anocell inserts are used instead of Millicell-HA inserts. Anocell inserts have a smaller pore size than Millicell-HA inserts, and therefore a higher concentration of the sodium-fluorescein dye is necessary in order to obtain sufficient ODs to distinguish different levels of damage to the monolayer. All protocols use Millicell-HA inserts with the exception of INVITTOX Protocol No. 86 which uses Costar Transwell inserts.

viii. Insert type

The various insert types used for the FL assay were known to alter the results considerably (Ward *et al.*, 1997a). Inserts principally vary in the following parameters; number of pores, pore size, formation of the o-ring, membrane surface charge, presence and type of coating (e.g. Laminin Type I or Collagen IV). These differences can affect cell attachment, formation of tight junctions, and impact on the sensitivity of cell responses to various chemical exposures. The main properties of four different inserts commonly used in the FL assay protocols are shown (table 2.2.1.1.)

Table 2.2.1.1. Insert properties (modified from Ward *et al.*, 1997a)

Insert Name	Material	Thickness (µm)	Pore Size (µm)
Anopore	Aluminium oxide	45	0.2
Anopore	Aluminium oxide	45	0.02
Millicell-HA	Mixed cellulose esters	80-150	0.45
Millicell-CM	Stretched polytetrafluorethylene (Biopore) (coated with human placental type IV Collagen)	50 (insert alone)	0.4

The insert membranes promote differentiation of the cells. The epithelial monolayer grown on the membranes acts as a barrier to apical and basal media. Anopore membranes (Anocell inserts) possess a very flat surface to which cells attach via small cytoplasmic extensions. The Millicell-HA membranes have an uneven, matted surface to which the cells project downwards (Zanvit *et al.*, 1999). Cells grown on the Anopore inserts form a greater number of tight junctions and can be more sensitive to toxic effects, than cells grown on the Millipore-HA inserts (Atkinson, (1995), unpublished observations). This difference is probably related to cell attachment to the insert membranes as MDCK cells have greater attachment to Millicell-HA membranes in comparison to Anopore membranes. The various sensitivities of the cells grown on different membranes is overcome by using a 15 minute toxicant exposure for cells cultured on the Millicell-HA insert membranes (Zanvit *et al.*, 1999) which contrasts to the one minute exposure used for cells grown on the Anopore membrane. The original FL assay developed by Tchao (1988) used a 0.45µm pore-sized insert membrane and a 15 minute exposure. Shaw *et al.*, (1990) modified this protocol to use 0.02µm pore-sized insert membranes with a one minute exposure. The one minute exposure period models the approximate length of time a material is likely to remain in the eye *in vivo*, following an accidental exposure in humans.

The two inserts most commonly cited in the FL assay publications were the Millicell-HA insert and the Anocell insert. Ward *et al.*, (1997a) performed a bridging study to determine the effects of insert membranes (featured in table 2.2.1.1.) on cell growth and FL assay sensitivity. Only the Millicell-CM membrane required a coating solution which consisted of human placental type IV collagen. This insert was not cited in any of the INVITTOX Protocols. Cell attachment to the membranes was good, with FL through the monolayer less than 10% for all the inserts tested. The following surfactants were tested in each insert; SDS, tween 20, triton x-100, benzalkonium chloride, cetrimide, CAPB. The majority of test materials caused different results according to the insert used. The cationic surfactants benzalkonium chloride and cetrimide produced similar results regardless of insert type, e.g., 1mg/ml of both test chemicals induced an amount of FL which remained the same or decreased at higher test concentrations. The amount of FL induced by the threshold concentration of 1mg/ml varied between inserts (Ward *et al.*, 1997a). The Millicell-HA insert produced a wide range of FL₂₀ results in comparison to the other inserts tested (Ward *et al.*, 1997a). In the absence of a monolayer, the cationic surfactants bind to insert membranes and/or blocked pores to different degrees. As FL cannot be determined to be solely due to the insert or cell monolayer, experiments using a no cell control help to interpret the effects of these chemicals on the cells.

The FL assay data from the EC/HO study (Balls *et al.*, 1995) which used Millicell-HA inserts, was compared to data for the same chemicals tested on Anopore membranes (Ward *et al.*, 1997a). The results from the Anopore membranes were found to give a better correlation to Draize MMAS scores than the Millipore membranes. The regulatory bodies generally required MAS scores.

In comparison to the Millipore membrane, an advantage of the Anopore membrane is that it becomes opaque when wet. This allows damage to the membrane and/or cell monolayer, which may occur due to the multiple washing steps, to be easily observed. The FL Prevalidation study tested CB997 and NBL-2 MDCK cell strains in combination with Millipore (0.45µm pore size), Anopore (0.2µm pore size), and Costar inserts (0.4µm pore size) (Southee, 1998). The Anopore membrane was more robust than the Millicell-HA insert (Southee, 1998). Where damage to the membrane is visible, results from the assay are discarded. Despite the advantages of Anopore membranes, the FL Assay Prevalidation study observed that both MDCK cell strains had a tendency to peel-off the Anopore membrane and cause barrier permeability.

Test materials can impede FL through the insert membrane via a number of interactions, e.g. attraction of charges between the test chemical and insert membranes or cells, or a physical blockage due to the molecular weight of a chemical. Chemical binding to the membrane, is more common for cationics, e.g. benzalkonium chloride, which are attracted to the positively charged membrane (Balls *et al.*, 1995). Negatively charged cell surface proteins can also attract positively charged surfactants. Chemical binding to the insert membrane or cell surface, increases the chemical exposure period but can also physically reduce FL, since the sodium-fluorescein dye can bind to the cationic surfactant bound to the insert membrane. This can be readily monitored by exposing the membrane to the top concentration of the chemical tested and then adding sodium-fluorescein dye at the normal concentration for the standard time. If binding of the sodium-fluorescein dye occurs, the insert membrane appears yellow after the test material has been washed-off. It is essential to know the binding properties of the test materials in order not to over-estimate the toxic potential of those materials that remain exposed to cells longer than the defined exposure period. In contrast, materials that block the pores and impede FL will cause toxicity to be under-estimated. Viscous materials are difficult to remove from the inserts, which pose the problem of continued exposure and/or physically blocking the membrane pores and preventing FL; both effects will produce erroneous results. The different properties of the insert membranes determine which test materials they will interact with, e.g. acetic acid reacts with some Millicell-HA membranes but not the Anopore membrane (INVITTOX Protocol No. 120).

ix. Duration for FL

Following test chemical/formulation exposure and removal, the subsequent time given for FL to occur varies according to each protocol. A 30 minute period at room temperature (INVITTOX Protocol No. 71, INVITTOX Protocol No. 86, INVITTOX Protocol No. 120) is commonly used. Other protocols employ a 60 minute period, which can be either at room temperature (INVITTOX Protocol No. 82) or incubated at 36°C (Clothier *et al.*, 1997; Clothier *et al.*, 2002). The incubator is used with a FL period of 60 minutes since it is possible that a prolonged exposure to room temperature could damage the cells, and impact on FL due to the stress of external conditions. The incubator is also used where the laboratories are interested in combining the FL assay with a cell activity assay e.g., Resazurin/Alamar Blue (AB) assay. There is no evidence

that a 30 minute exposure at room temperature damages the cells. An advantage of using a 60 minute period to allow for FL is that higher amounts of leakage occur. Thus, a 60 minute period for FL will potentially allow for a greater degree of discrimination between FL assay results, than the 30 minute period. A greater range of values is particularly important for the Fixed Dose FL assay (INVITTOX Protocol No. 82) where only one concentration is tested. This will allow the toxicity of a number of test materials at a set concentration to be compared. The advantage of the 30 minute period for FL is that the assay is quicker to perform. Both 30 minute and 60 minute periods are sufficient to produce reproducible FL assay results.

x. Time-point FL is measured

The time-points used to measure FL following the chemical exposure varies between protocols. The most common time-point for measuring acute effects is immediately following the chemical exposure. As some protocols allow 30 minutes or 60 minutes for FL to occur, the time of measuring the amount of FL varies by 30 minutes. The other common time-point for measuring acute effects is at 4h (INVITTOX Protocol No. 120). Additionally, the cell monolayers have a greater tendency to 'peel-off' the membrane when the FL assay is performed immediately following the exposure, which is less likely to occur when the assay is performed at 4h. Subsequently, the 4h time-point is generally used when the results are entered into a PM to predict human eye irritation (Southee, 1998; Zanvit *et al.*, 1999). Zanvit *et al.*, (1999) also stated that the FL measured 4h following the chemical exposure produced results with a better predictive capacity for surfactants.

Additional time-points used to assess recovery and delayed effects are 24h, 48h, 72h (INVITTOX Protocol No. 86 and No. 120). Measurements can be made at 24h intervals over a total of 72h; the 72h time-point features frequently in the literature (Clothier *et al.*, 1994; Clothier *et al.*, 1999; Clothier *et al.*, 2002; Cottin and Zanvit, 1997). Variability can occur at 72h and this can be due to the effects of repeated FL assays or the differential time for recovery between replicates. A measurement for recovery at the 72h time-point was incorporated into the PM used by the Fixed Dose FL assay (INVITTOX Protocol No. 82) (Clothier *et al.*, 1994). The EC/HO study also proposed a PM that included 72h results which could be used to distinguish R36 and R41 classified ocular irritants.

xi. Positive Controls

The various FL Assay INVITTOX Protocols use different acceptance criteria and positive controls. INVITTOX Protocol No. 71 uses 100mg/ml Brij 35 which should cause approximately 30% FL, although results are accepted if the positive control produces 20% to 40% FL. INVITTOX Protocol No. 120 uses a positive control of 130mg/ml glacial acetic acid, which should produce approximately 50% damage to the cell layer; results are also accepted if the positive control produces FL results in the range of 30-70% (Southee, 1998). INVITTOX Protocol No. 82, uses 0.16mg/ml SDS which must induce less than 50% leakage in order for the results to be accepted. In the literature, SDS is the most common chemical used as a positive control but the range of FL which it must induce for results to be accepted varies between the publications. INVITTOX Protocol No. 86 does not state which chemical is to be used as a positive control. However, this protocol featured in the CTFA study Phase III where a Company # 3 baby shampoo was used as a positive control (Gettings *et al.*, 1996). Results were only accepted if the test concentrations of Company # 3 baby shampoo, reached a plateau of 100% FL over

more than one concentration. This acceptance criteria controls only for the correct concentration range being tested and does not aid test reproducibility.

With the exception of INVITTOX Protocol No. 120, all INVITTOX protocols state a range for which the positive controls must induce a certain amount of FL. Assuming that suitable chemicals have been chosen as positive controls, INVITTOX Protocol No. 71, with the smallest range for the positive control, would be expected to produce the most reproducible results for the test chemicals.

xii. FL endpoint measured

The endpoint to be measured as stated in the various INVITTOX Protocols differs, but are typically; FL₁₀ (INVITTOX Protocol No. 120), FL₂₀ (INVITTOX Protocol No. 71, No. 82, No. 120), FL₅₀ (INVITTOX Protocol No. 71), EC₅₀ (INVITTOX Protocol No. 86) and FL% (INVITTOX Protocol No. 71). The FL assay has also been performed on a human corneal cell line where the chemical concentration giving rise to 85% fluorescein retention (FR₈₅) was recorded which equates to FL₁₅ (Clothier *et al.*, 1999).

FL₂₀ features most predominately in the literature. FL₂₀ is used in order to determine the concentrations that might cause damage to the tight junctions rather than non-specific cellular damage which is more likely to be measured at the FL₅₀ endpoint. The high chemical concentrations required to induce FL₅₀ values within a short exposure could possibly cause severe damage to the cells and even cell loss.

Prediction Models

Prediction models (PMs) are necessary to interpret the FL assay results in terms of potential ocular irritancy to humans. In various publications, PMs have been defined for all four INVITTOX Protocols; only INVITTOX Protocols No. 86 and No. 120 actually feature PMs. A PM for INVITTOX Protocol No. 71 data was devised following the results of the EC/HO study (Balls *et al.*, 1995); the PM does not appear in INVITTOX Protocol No. 71 (table 2.2.1.2.).

Table 2.2.1.2. PM for data generated using INVITTOX Protocol No. 71 as presented in the EC/HO Study publication (Balls *et al.*, 1995).

FL₂₀ (mg/ml)	EU Classification
>750mg/ml	Not Classified
>100 ≤750mg/ml and recovery after 72h	R36
>100 ≤750mg/ml and further deterioration after 72h	R41

There was no definitive PM reported for the method in INVITTOX Protocol No. 82 (table 2.2.1.3.).

Table 2.2.1.3. The PM for INVITTOX Protocol No. 82 (taken from the publication of Clothier *et al.*, (1994)).

Chemical tested at 50mg/ml	EU Classification
causes <FL20 (%)	Not Classified
causes ≥FL20 (%) with consideration of recovery or further deterioration at 72h	R36/R41

NB. Clothier *et al.*, (1994) state that the fixed concentration tested can differ according to prior knowledge of the types of test materials.

INVITTOX Protocol No. 120 contains the following PM (table 2.2.1.4.).

Table 2.2.1.4. PM from INVITTOX Protocol No. 120

FL ₂₀ (mg/ml)	EU Classification
>100 mg/ml	non-irritant/slight
20 -100 mg/ml	moderate
<20 mg/ml	irritant/severe

Table 2.2.1.5. shows the PM which features in INVITTOX Protocol No. 86.

Table 2.2.1.5 PM from INVITTOX Protocol No. 86 (TEP assay)

EC ₅₀ (%)	TEP Assay Rating
≤1.8%	fail
>1.8% <2.2%	borderline/undetermined
≥ 2.2%	pass

Briefly, the PMs used with the INVITTOX Protocols vary in the following ways:

- number of classifications
- correlation with standard classification system (i.e. EU risk phrases)
- size of the range of values used to distinguish different irritancy classes
- use of recovery data
- types of materials PM is applicable for*

* although not explicitly stated in INVITTOX Protocol No. 120, the PM was devised based on the results of surfactant-based formulations and only applied to these types of test materials in the studies reported in the literature (Zanvit *et al.*, 1999).

Both INVITTOX Protocol No. 71 and No. 82 have PMs reported in the literature which predicted EU classifications. INVITTOX Protocol No. 120 is comprised of three classes which distinguish two levels of irritancy, but not according to a standard classification system. INVITTOX Protocol No. 86 features a PM which is only able to distinguish irritants and non-irritants but not according to a standard classification system. The different PMs have been challenged to various levels according to the types of materials tested to generate the data which were entered into the PMs.

2.2.2. Proposed critical components of the protocol that impact on reproducibility and/or predictive capacity of the assay

Of the protocol steps listed in the previous section, the following elements were considered 'critical' in terms of the reproducibility and predictive capacity of the FL assay protocols (figure 2.2.2.1.):

- Cell type, cell strain -in medium with a sufficient calcium concentration to ensure tight junction formation and integrity.
- Cell seeding -to ensure confluence of the cells at the times of exposure.
- Confluence and tight junction integrity checked for each insert before FL assay –fluorescein leakage must be below 6%.
- Passage number range to ensure even and reproducible tight junction formation.
- Insert and membrane type:
 - affects cell growth
 - chemical binding can be membrane type specific
- Test chemicals should be prepared in calcium-containing HBSS or medium without serum to avoid serum protein binding.
- Test chemical exposure period duration –differs according to cell type and cell strain but should enable immediate chemical effects to be measured
- Ambient temperature of exposure is room temperature (23-25°C) or at 37°C in the CO₂ incubator. The lower room temperature has been found to have little effect on the rate of FL, but the temperature can be of relevance when longer than 1-5 minute chemical exposures are employed.
- Time given for FL to occur to ensure fluorescence OD values are within a range to produce a dose-response curve
- Endpoint FL₂₀, FL₅₀, FL%, calculated as % of non-cell control inserts treated with top concentration of test chemical set at 100%.
- Time-point FL is measured, i.e. immediately, or at 4h, then 24h, and/or 48h and/or 72h for recovery.
- Prediction model

Figure 2.2.2.1. Critical elements of FL assay protocols

The rationales for these critical protocol steps were stated in the previous section (section 2.2.1):

2.2.3. List of studies with similar protocols (no protocol differences, or no impact of protocol differences on predictive capacity)

Various FL assay protocols have been grouped according to the number of features that they have in common.

Group 1 INVITTOX Protocol No. 71: These studies used the same cell strain, test material exposure duration, and FL duration:

- Loss of Trans-epithelial Impermeability of a confluent layer of Madin-Darby Canine Kidney (MDCK) Cells as a Determinant of Ocular Irritancy Potential (Shaw *et al.*, 1990)
- Predicting Ocular Irritancy and Recovery from Injury using Madin-Darby Canine Kidney Cells (Shaw *et al.*, 1991)
- Development of a Fixed Dose Approach for The Fluorescein Leakage Test. (Clothier *et al.*, 1994) (**INVITTOX Protocol No. 82**)
- Evaluation of Tissue Culture Insert Membrane Compatibility in the Fluorescein Leakage Assay (Ward *et al.*, 1997a)
- The Evaluation of Pesticide Ingredients and Formulations *In Vitro* and Correlations with *In Vivo* Data (Clothier *et al.*, 1995)
- The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test (Balls *et al.*, 1995)
- FAL In-house studies

The study of Tchao *et al.*, (1988) could not be included in this group because it was not documented which cell strain was used in the protocol.

The study of Jones *et al.*, (1988) could not be included in this group because the cells used were obtained from Tchao and the cell strain was not documented. In addition, the chemical exposure duration and the endpoint measured using this protocol also differed to those listed in this section.

The inserts used in the two studies performed by Shaw *et al.*, (1990; 1991) were the same, although marketed under different names; Anocell 10 (Shaw *et al.*, 1990) and Anotec 10mm (Shaw *et al.*, 1991). The endpoints differed between the Shaw *et al.*, (1990; 1991) studies and the Fixed Dose FL assay of Clothier *et al.*, (1994). The Fixed Dose FL assay tested a set concentration of 50mg/ml for each test material. Materials that attained FL₂₀ (%) or above at this concentration were classified as potential irritants; those that induced less-than FL₂₀ were classed as non-irritants. No PM was used in the studies by Shaw *et al.*, (1990; 1991), although a good correlation between the ranking of the FL₂₀ results with *in vivo* EU risk phrase classifications were reported. The predictive capacity of the protocols for *in vivo* classifications were similar as comparisons of the results from Shaw *et al.*, (1990) with the Fixed Dose FL assay data (Clothier *et al.*, 1994) generally indicated the same chemicals to be either irritants or non-irritants. For example, 50mg/ml DMSO did not cause any FL in the Fixed Dose FL assay and was ranked as one of the least toxic chemicals tested in the study of Shaw *et al.*, (1990).

Group 2- INVITTOX Protocol No. 120: These studies had the same protocol elements in terms of cell strain, test material exposure duration, insert type, duration for FL:

- A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom *et al.*, 1997)

- Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Southee, 1998)
- Ocular Irritancy Assessment of Cosmetics Formulations and Ingredients: Fluorescein Leakage Test. (Zanvit *et al.*, 1999)

The protocols featured in the following publications could not be grouped with INVITTOX Protocol No. 120 and similar protocols' as the FL assay was performed immediately after the chemical exposure:

- Gautheron *et al.*, (1994a.) Investigations of the MDCK Permeability Assay as an *in vitro* test of Ocular Irritancy
- Fluorescein Leakage Test: a Useful Tool in Ocular Safety Assessment (Cottin and Zanvit, 1997)

Group 3- INVITTOX Protocol No. 86: Data from the CTFA study Phase III was also submitted to the IRAG study, thus identical protocols are featured.

- CTFA Evaluation of Alternatives Program: An Evaluation of *In Vitro* Alternatives to the Draize Primary Eye Irritation Test, Phase III (also submitted to IRAG Working Group 3: Cell Function-based Assays (Botham *et al.*, 1997) Lab B)
- Company # 3

Group 4- INVITTOX Protocol No. 71 combined with the Alamar Blue assay

- Effects of Surfactant Re-treatment *In Vitro*: A Method to Evaluate Changes in Cell Junctions and Cell Viability (Clothier and Sansom, 1996)
- The Prediction of Human Skin Responses by using the Combined *In Vitro* Fluorescein Leakage/Alamar Blue (Resazurin) Assay. (Clothier *et al.*, 2002)
- Assessment of Initial Damage and Recovery Following Exposure of MDCK Cells to an Irritant (Clothier *et al.*, 1999)

The protocols featured in the following publications could not be grouped with other INVITTOX Protocols due to many differences but most significantly, different cell types:

- Human Corneal Epithelial Primary Cultures and Cell Lines with Extended Life Span: *In Vitro* Model for Ocular Studies (Kahn *et al.*, 1993)
- Evaluation of a Human Corneal Epithelial Cell Line as an *In Vitro* Model for Predicting Ocular Irritation (Kruszewski *et al.*, 1997)
- Evaluation of Chemically Induced Toxicity Using an *In Vitro* Human Corneal Epithelium (Ward *et al.*, 1997b)

2.2.4. Known applicability and limitations of the assay (including ranges of irritancy, types of substances, technical limitations)

The FL assay was designed to measure chemical-induced permeability of a monolayer in order to model loss of impermeability of the corneal and conjunctival epithelia. The FL assay was specifically developed to detect potentially mild and moderate irritants to the human eye, which are often cosmetic products. The technical limitations of the assay and the applicability domain are discussed below.

There are certain technical limitations specific to MDCK cell culture. The tight junctions that block the passage of the sodium-fluorescein dye through the monolayer are increasingly compromised with increasing cell passage number. Incomplete formation of the tight junctions results in increased FL in the non-treated control. Therefore, a defined permissible maximal leakage in the non-treated controls is important. Heat inactivated-FCS can be difficult to obtain commercially, and the process of heat-inactivating serum can cause protein precipitations to form that require filtration; this process in-house can lead to inter-laboratory and intra-laboratory variations in the quality of the serum batch to batch. Protocols do not generally give details on the method for heat-inactivating serum which could reduce variability and improve tight junction competence. As with all *in vitro* assays there is the potential for the cells to become transformed over time, thus it is vital that passage number ranges for the assays are stated. However, the risk is that cell stocks are depleted quickly, as with the Gillette human corneal cells that require a low passage number.

An advantageous feature of the FL assay is the short exposure which is pertinent to the *in vivo* scenario, where an accidental exposure is likely to be blinked and washed out by tearing in approximately one minute. Due to the short exposure period, high concentrations of test materials can be tested which are relevant to human accidental exposures. High chemical concentrations can react with the insert membranes, and is particularly noticeable for corrosive compounds. However, one can assume that any chemical causing membrane corrosivity would also have a significant ocular irritancy potential. There are some examples of chemicals at high concentration that react with the insert membrane and compromise fluorescein leakage through the membrane. This can be monitored by exposing an insert with no cells to the highest chemical concentration tested.

A problem associated with some test materials is the difficulty in efficiently removing them from the insert after a short incubation period; the incubation periods featured in the literature range from ten seconds to 15 minutes. This is particularly applicable for viscous materials, such as gels and creams that are commonly formulated to have low irritation potentials and for which the FL assay was designed to measure. A ten second exposure is less likely to produce reproducible results than a 15 minute exposure because if the test material is not removed fully or efficiently, the impact is a greater proportion of a ten second period in comparison to a 15 minute period. Thus, the reproducibility of the assay is reflected by the ability to control the real/effective exposure period.

The majority of FL assay protocols featured in the literature tested materials solubilised in either HBSS or distilled water. Chemicals that are not soluble in HBSS or distilled water can only be used if they form a stable suspension or emulsion. However, emulsions and suspensions will not be as homogeneous as a solubilised material and

different areas of the cell monolayer will receive different concentrations of the test material which will cause variable results. It is not easy to establish that an emulsion or suspension is uniform or stable, although the short exposure duration provides less time for the emulsion or the suspension to degrade. Therefore it is important to specify the total time between emulsion or suspension preparation and end of the exposure time. Materials that are not soluble or do not form stable aqueous emulsions can be solubilised or suspended in mineral oil. Highly volatile materials such as acetylformaldehyde need to be solubilised in mineral oil to reduce evaporation and ensure that they remain in contact with the cells for the specified exposure period. INVITTOX Protocols No. 71 and No.120 state that products can be tested neat. Formulations that are tested neat are typically creams and gels that are already diluted in a vehicle for direct human exposure, and have low potency. These can be difficult to remove from the insert after the short exposure period. Solids cannot be reliably tested as the concentration in contact with the cells cannot be assumed to be equal to that placed on the cell monolayer, nor uniformly distributed. Also, solids cannot be easily removed from the cell surface following the short exposure period (Balls *et al.*, 1995).

Like other cell based assays, the FL assay does not have a good predictive capacity for test materials that have fixative properties and/or reactivity with medium contents. The FL assay is useful for testing only mild to moderate irritant chemicals within a defined range where it can measure mechanistic damage caused to the adhesion molecules. Testing surfactants requires the ionic properties of the chemicals to be taken into consideration when interpreting results. For example, the pH of cell culture medium can affect whether an amphoteric surfactant will display anionic or cationic properties (Cheah; BMedSci, 2000). Materials that are water soluble and/or the toxic effect is not affected by dilution are generally predicted accurately using the FL assay. It is important that the basic toxic mechanism is not affected by dilution as the formation of micelles by surfactants can unpredictably alter cellular responses over a concentration range; this impacts on the predictive capacity of *in vitro* assays. Micelles are surfactant molecules grouped into spheres which reduce the surfactant surface area available for exposure to the cells. At a set concentration, referred to as the 'critical micelle concentration' (CMC), surfactants form micelles. Thus, it is possible that a higher surfactant concentration produces less irritancy than a lower concentration *in vitro* and *in vivo*, as micelles have formed at the higher concentration. As the FL assay can test neat or high test material concentrations, the unpredictable effects of dilution on the CMC could remain undiscovered. The effect of CMC is a particular problem for the Fixed Dose FL assay, as only one concentration is measured and the effects of micelle formation on irritancy potential will not be easily detected. In most of the Fixed Dose FL assay studies conducted at FAL, a series of fixed doses are tested in order to discriminate between micellar and non-micellar effects.

The sodium-fluorescein dye is solubilised in HBSS as phenol red and bovine serum found in the medium could potentially interfere with the OD readings. The human corneal model uses high calcium KGM medium. As KGM medium is without serum, this medium type is less likely to interfere with test results than a full serum-medium used for MDCK cells. However, the KGM medium does have a buffering capacity greater than the salt solutions used in other FL assays that can compromise the effects of acids or alkalis.

The inserts used for the FL assay were not designed specifically for the FL assay but as filters. Insert manufacturers have issued warnings regarding potential membrane

incompatibilities with test chemicals being filtered. It is important to know of any interaction between the insert membrane and test material in order to be able to confidently interpret FL assay results. For example, test materials can bind to the insert membrane, thus making their removal very difficult. Chemical binding to the insert membrane, is more common for cationic materials, such as benzalkonium chloride, which are attracted to the positively charged membrane (Balls *et al.*, 1995). The different inserts available are known to affect the results as acetic acid is incompatible with the Millicell-HA insert but often used as a positive control for the Anopore membrane (Balls *et al.*, 1995). Negatively charged cell surface proteins could also attract positively charged surfactants. Increasing the number of washing steps to remove the test material from the insert membrane can also lead to insert membrane damage and therefore erroneous results. Routine observations of the monolayer on the inserts throughout the exposure and the recovery phases should detect damaged monolayers and/or insert membranes and prevent erroneous results being accepted. Anopore membranes are opaque when wet and so damage to the monolayer and the membrane can be easily observed and the results from the assay discarded. Alternatively, if test materials with high molecular weights are not removed fully and/or efficiently, they can physically block the passage of the sodium-fluorescein dye through the insert, which could cause chemical effects to be under-estimated. In general, additional uncontrolled exposure time is a greater proportion for the FL assay with a short exposure period (~one minute), than with the assays with longer exposures (~15 minutes). This leads to greater variability in results, and low assay reproducibility. As the FL assay can be repeated at multiple time-points, erroneous results due to ineffective removal of the test material would more likely be detected in comparison to cell viability assays which use single time-points. The efficient removal of test agents from the eye is also a concern of the *in vivo* test.

Like many *in vitro* assays for eye irritancy, there are a number of endpoints that are measured *in vivo* that presently cannot be measured or replicated using the FL assay. The effect of lachrymation which reduces toxic effects by diluting and removing the test material are difficult to quantify and replicate *in vitro*. However, the results gained using the Draize test in rabbits eyes are not wholly representative of the human situation due to the species variation in lachrymation rates. Pathological conditions such as chemosis which are scored in the Draize test cannot be measured using an *in vitro* test.

In contrast to most cell-based *in vitro* assays for eye irritation, reversibility and recovery of effects can be measured using the FL assay for up to 72h following the initial exposure. For assessing recovery, the inserts are removed to new wells containing fresh medium following the chemical exposure. Transferring the filter inserts can cause bubbles to form under the basal side of the insert membrane which can impede contact between the cells and the medium. Bubbles can also reduce the passage of fluorescein from the apical side of the insert membrane into the solution below the insert membrane and distort results (Balls *et al.*, 1995). Repeated assays on the same population of cells are more likely to cause damage to the monolayer.

Attempts to study the effects of repeated chemical exposures have proved unsuccessful following the second exposure due to cumulative cell damage which destroys the monolayer (Clothier and Sansom, 1996).

A combined FL/AB assay has been developed to try to distinguish the specific damage caused to the tight junctions from cell membrane damage. A disadvantage specific to

the combined FL/AB assay is that the AB assay results can be erroneous due to the length of time taken to perform the protocol steps. AB reduction is affected by temperature (Clohier, personal communication). The combined assay is a lengthy procedure which involves temperature changes of the cell culture. In addition, high cell density and prolonged culture times allow reversal of the AB reduction process which can lead to an over-estimation of toxicity (Larson *et al.*, 1997). Therefore the AB assay results may not correctly represent the chemical-induced damage to cell viability. This would cause FL results to be read incorrectly leading to over-estimations or under-estimations of chemical effects on tight junctions.

In conclusion, the FL assay is better suited to measure high concentrations of test materials that have low to mid-range toxicity and are soluble in water or HBSS. Test materials that are difficult to remove from the inserts due to viscosity or binding to the membrane are not accurately measured. Materials that have their basic toxic mechanism affected by dilution are not accurately measured by the FL assay.

2.2.5. Others

i. Combined assays

The FL assay has been used in combination with other assays, principally cell viability assays. These assay combinations allow the specific damage caused to the tight junctions to be distinguished from general cell loss and non-specific cell damage affecting viability and/or membrane functions.

A combined FL/AB assay was used to predict human skin responses to a set of test materials comprised of surfactant-based handwashes, laundry detergents and moisturisers (Clothier *et al.*, 2002). The AB assay measures the reduction of resazurin to resorufin by a number of enzymatic pathways, and the rate of reduction can be used as a measure of cell viability (Page *et al.*, 1993; O'Brien *et al.*, 2000). Using a combined FL/AB assay, chemical effects on cell viability and barrier integrity can be measured. A significant loss of resorufin production and an increase in FL can indicate loss of cell viability leading to cell death, whilst FL with only small decreases in resorufin production indicates reversible barrier damage to the cell monolayer. Details of this study are given in brief.

Following a one minute exposure, a combined solution of sodium-fluorescein dye and AB was placed onto the MDCK monolayer, and incubated for 60 minutes. AB was also placed into the well to allow apical and basal uptake. Following incubation, the solution in the wells was measured for FL. The solution from the insert was then added to that in the corresponding well and measured for resorufin production. Repeat assays with fluorescence readings were performed 4h, 24h, 48h and 72h after the initial test material exposure. The results from the two assays were considerably different at the various time-points. For example, 24h after a 25mg/ml exposure to 'Kids' antibacterial moisturising handwash' cellular activity measured by the AB assay was restored. The epithelium impermeability, as measured by the FL assay, was not restored until the 48h time-point.

A PM was generated using FL/AB data from the hand-wash formulations. FL/AB data were compared to human patch test data from experiments which were performed during the same time-period as the FL assay. The FL/AB data for the laundry detergents and moisturisers were entered into the PM to predict the human patch data. The PM was under-predictive for two of ten handwashes and over-predictive for two of six moisturisers and eight out of ten laundry powders. The *in vivo* patch test data to which the *in vitro* data was compared was also notably variable (Clothier *et al.*, 2002).

The combined FL/AB assay was also used to study recovery following repeat exposures to surfactants (Clothier and Sansom, 1996). The effect of repeated surfactant exposures were of interest as many surfactant-based products, such as cosmetics, are designed for repeated use. Adverse reactions are potentially more likely to be observed following multiple exposures rather than a single acute exposure. Therefore, the ability of the FL/AB assay to measure the effects of repeated exposures to the surfactant CAPB was assessed. MDCK cells were grown to confluence and the FL/AB assay performed prior to test material exposure. The cells were exposed to CAPB for one minute and the FL/AB assay carried out immediately and then at 24h and 72h following exposure. After the 72h FL/AB assay, the cells were re-treated with one of four different surfactants for one minute. The FL/AB assay was then performed at 1h, 24h and 72h

following the initial exposure. After the second exposure, cells exposed to a pre-treatment of CAPB did not recover as well as those that only received HBSS. In general, the lower concentrations of the second surfactant treatment were found to promote recovery as determined by decreased FL and increased AB reduction (Clothier and Sansom, 1996).

The combined FL/AB assay was used with the J-HCET cell line (Araki-Sasaki *et al.*, 1995) to predict the effects of repeated exposures to low doses of surfactants on the corneal epithelium barrier. The J-HCET cells were grown in low concentrations of surfactant containing medium prior to toxicant exposure (Cheah; BMedSci, 2000). Pre-treatment of a low dose surfactant was hypothesised to reduce the sensitivity of the cells to the following acute dose as measured by the FL assay. These experimental conditions reflected the *in vivo* scenario of wearing eye drops or being repeatedly exposed to contact lens solutions. Cells were treated with low doses of either CAPB or benzalkonium chloride and then treated with either SDS, tween 20, CAPB or benzalkonium chloride (Cheah; BMedSci, 2000). A five minute chemical exposure was used in combination with a FL₁₅ endpoint (Cheah; BMedSci, 2000). FL/AB assays following this treatment and prior to the acute exposure showed that the cell junctions were not affected by these low doses. FD was used in conjunction with J-HCET cells as sodium-fluorescein dye produced variable results and a greater rate of FL than with MDCK cells. For this study, FDs with molecular weights of 4.4kD (FD-4) and 9.5kD (FD-9) were used to assess the size of the tight junctions. FD-4 at a concentration of 200µM showed the smallest variation and optimum FL OD values. Throughout the experiments, AB values did not differ considerably, indicating that measured FL was due to damaged tight junctions rather than cell damage. The FL/AB assay was repeated at 24h and 48h after the second treatment, and recovery as indicated by reduced FL, was observed by 48h for all treated cells. In agreement with *in vivo* data, the FL assay ranked the four surfactants accordingly, with the cationic being the most toxic followed by the anionic, amphoteric and non-ionic surfactants (Cheah; BMedSci, 2000). Further work was deemed necessary in order to determine the use of this cell model for predicting the *in vivo* situation (Cheah; BMedSci, 1997).

The FL assay has also been used in combination with the Neutral Red Uptake (NRU) assay which was used to distinguish tight junction damage from loss of MDCK cell viability caused by ethylenebis(oxyethylenenitrilo)tetraacetic acid (EGTA) (Carter; BMedSci, 1994). EGTA was tested as it was hypothesised to be a more appropriate positive control than the acetic acid which was then routinely used. EGTA chelates extracellular calcium which is required for tight junction function (Carter; BmedSci, 1994). A combined NR dye/ sodium-fluorescein dye solution was added to the inserts and the protocols modified to allow a 30 minute incubation with the sodium-fluorescein dye and a 60 minute incubation with the NR dye. Following a one hour incubation, the dyes were removed, cells washed, and a fixative added to each well. OD values were not significantly different for the FL assay performed alone or in combination with the NRU assay. The addition of the fixative for the NRU assay prevented repeated FL assays being performed on the same cells so recovery could not be assessed. EGTA was found to require more than a one minute exposure to cause damage to the tight junctions and was therefore concluded not to be an appropriate positive control (Carter; BMedSci, 1994).

ii. Storage of seeded MDCK cells

The FL assay has featured in a number of large-scale validation studies (Balls *et al.*, 1995; Gettings *et al.*, 1996; Zanvit *et al.*, 1999). This had led to investigations to study the feasibility of storing pre-prepared plates seeded with MDCK cells. The blood substitute HypoThermasoTM was successfully used to store human epidermal cells which were viable and retained functional integrity at 4°C. The ability of MDCK cells to remain viable and form tight junctions following 60 hours of storage in HypoThermasoTM was investigated. Five surfactants representing a range of ionics and cytotoxic potencies were tested using INVITTOX Protocol No. 71 and INVITTOX Protocol No. 82. A 60h exposure of MDCK cells to HypoThermasoTM at 4°C did not affect MDCK cell growth rate nor their ability to form tight junctions. In comparison, cells stored in medium for 60h at 4°C detached from the membrane when returned to 37°C. FL rates were comparable for both cold-stored and normally plated MDCK cells following treatments with the five surfactants for both acute exposure and recovery time points.

Annex I Standard FL assay protocols: a. The Fluorescein Leakage Test, INVITTOX Protocol No. 71; b. Fluorescein Leakage Test - SOP of Company # 4, INVITTOX Protocol No. 120; c. Trans-epithelial Permeability (TEP) Assay, INVITTOX Protocol No. 86; d. Fixed Dose Procedure for The Fluorescein Leakage Test, INVITTOX Protocol No. 82 ; e. Company # 3 Robotic Trans-Epithelial Permeability (TEP) Assay ; f. ECVAM Prevalidation Study Phase II Protocol; g. ECVAM Prevalidation Study Phase III Protocol

Appendix II Complete list of protocol steps

3. Within-laboratory reproducibility (Module 2)

Fourteen FL assay data sets were appropriate for within-laboratory reproducibility analyses. The mean, SD and CV were calculated for all the chemicals and formulations tested in these data sets (refer Annex II). The relevant information for these studies is given in table 3.1.

3.1 Table presenting the results and relevant information for each study

INVITTOX Protocol No. 71

Study	Variability Reported	No. of Test Substances	Type of Test Substances	No. of Laboratories	No. of Operators	No. of Experiments	No. of replicates (wells)	Data Format
Report on Comparison of 40 Cosmetic and Domestic Formulations Supplied by Company # 8 and Evaluated by 3 <i>In Vitro</i> Cytotoxicity Tests (FRAME, 1992)	No	40	surfactant-based formulations	1	3	4	3	FL20, FL50 (mg/ml) T0
Final Report on Testing of 12 Mild Surfactants supplied by Company # 5 for Cytotoxicity Testing at the FAL (FRAME, 1992)	n/a	12	mild surfactants	1	Not stated	n ≤ 4	3	FL20 (mM) T0, T24, T48, T72
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994)	No	10	surfactants	1	3	n ≤ 5	3	FL20, FL50 (mg/ml) T0 and T72
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994) –re-runs based on activity	No	10	surfactants	1	3	n ≤ 5	3	FL20 FL50 (mg/ml) T0 and T72
The Evaluation of Pesticide Ingredients and Formulations <i>In Vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995)	No	4	pesticides pure ingredients	1	1	4	3	FL20 (mg/ml) T0 and T72

INVITTOX Protocol No. 120

Study	Variability Reported	No. of Test Substances	Type of Test Substances	No. of Laboratories	No. of Operators	No. of Experiments	No. of replicates (wells)	Data Format
A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom <i>et al.</i> , 1997) FRAME	No	30	Surfactants and surfactant-based formulations soluble in HBSS	1	Not stated	4	3	FL20 (mg/ml) T0 and T72
ECVAM Prevalidation Study -Phase II (Southee, 1998)	A hierarchical ANOVA noted laboratory variation (results not reported).	5	4 chemicals and 1 formulation	4	FRAME =1, ECVAM =2, Company # 3=1, Company #7=not stated	n≤5	3	FL20 (mg/ml) T0 and T4
ECVAM Prevalidation Study -Positive control data (Southee, 1998) Phase II	No	1	0.16mg/ml SDS	4	FRAME =1 (2 occasions 1 other person), ECVAM =2, Company # 3=1, Company # 7=not stated	n≤2 3	3	FL20 (%) T0 and T4
ECVAM Prevalidation Study -Phase III (Southee, 1998)	No	10	surfactants	3	FRAME =1 (2 occasions 1 other person), ECVAM=2 Company # 3 =1	n≤5	3	FL20 (mg/ml) T0 and T4
ECVAM Prevalidation Study -Positive control data (Southee, 1998) Phase III	No	1	0.16mg/ml SDS	3	FRAME =1, ECVAM =unknown, Company # 3 =1	n≤1 9	3	FL20 (%) T0 and T4

INVITTOX Protocol No. 86

Study	Variability Reported	No. of Test Substances	Type of Test Substances	No. of Laboratories	No. of Operators	No. of Experiments	No. of replicates (wells)	Data Format
CTFA Phase III TEP assay data (submitted by CTFA)	No	23	surfactant-based formulations	1	Not stated	n≤4	Not stated	EC50 (%) T0
TEP assay Formulation data (submitted by Company # 3)	No	41	formulations; some surfactant-based	1	Not stated	n≤3	Not stated	EC50 (%) T0

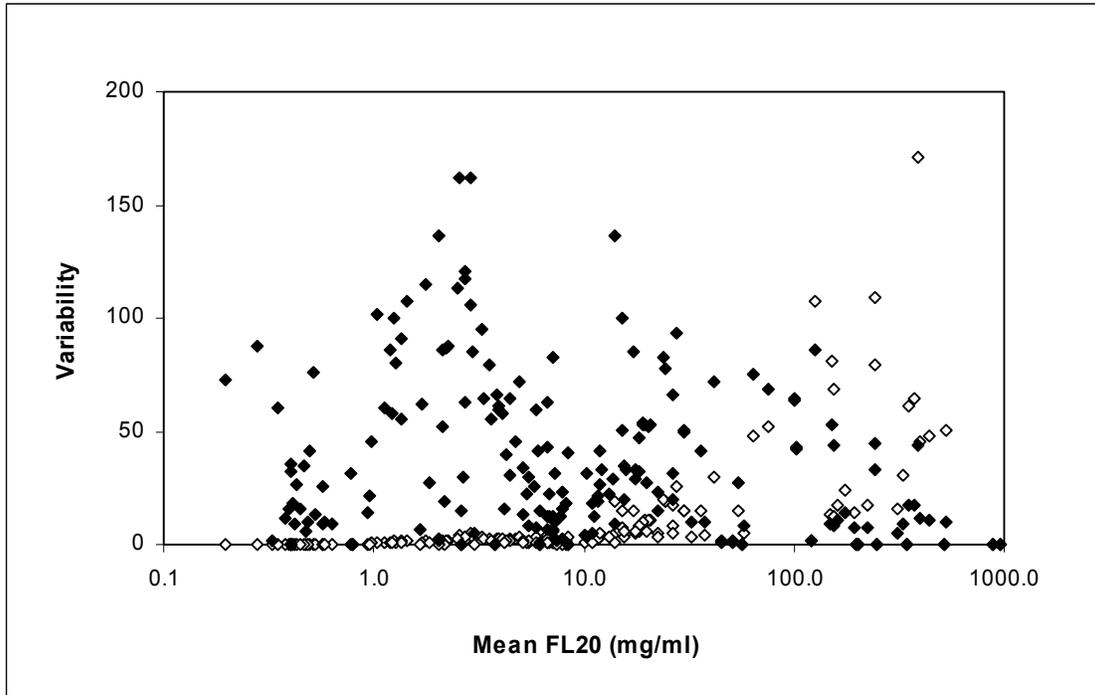
INVITTOX Protocol No. 82

Study	Variability Reported	No. of Test Substances	Type of Test Substances	No. of Laboratories	No. of Operators	No. of Experiments	No. of replicates (wells)	Data Format
Company # 5 test concentration X (FRAME, 1993)	No	12	surfactants for cosmetics	1	Not stated	n≤ 4	3	FL (%)T0, T24, T48, T72
Company # 5 test concentration XI (FRAME, 1993)	No	12	surfactants for cosmetics	1	Not stated	n≤ 4	3	FL (%)T0, T24, T48, T72

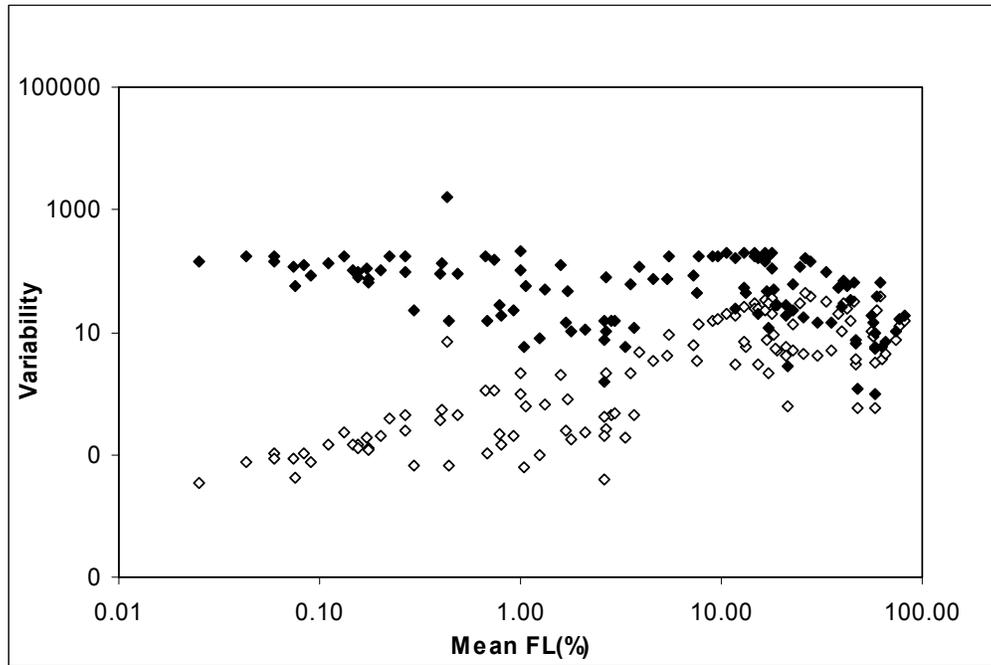
3.2. Compilation of Results

3.2.1. Statistical approach(es) used: description & rationale for the approach used

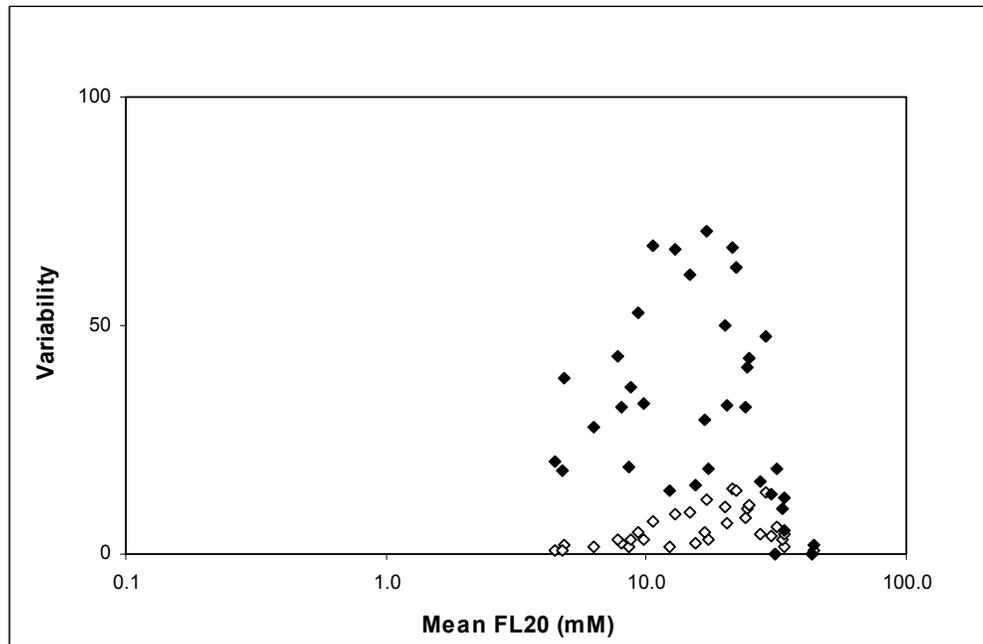
Data were analysed according to ECVAM guidance. The SD and the CV were plotted against mean FL assay values for the different endpoints, FL (mg/ml), FL (%) (figure 3.2.1.a-b). The aim was to determine which measure of variability (i.e. S.D or CV) produced the most consistent level of variability over the entire range of mean FL assay values. The most consistent measure of variability would then be used in further analyses to assess protocol within-laboratory variability.



a.



b.



c.

Figure 3.2.1. \diamond SD and \blacklozenge CV values plotted against the mean FL assay results; a. FL₂₀ (mg/ml); b. FL (%) ($n \geq 2$); c. FL₂₀ (mM). Materials which produced greater-than or less-than values in any experimental run were not plotted. Data for these graphs were taken from the studies listed in table 3.1. The raw data are visible in the Annex file titled 'Intralaboratory II' and are arranged according to the units of measurement, e.g. %, mM, mg/ml.

Figure 3.2.1.a. predominately contains data produced by measuring FL immediately following a one minute chemical exposure, and to a lesser extent from FL assays measured 72h and then 4h, 24h and 48h following the initial test material exposure. The mean FL₂₀ (mg/ml) values covered a larger range than the mean FL (%) values.

Figure 3.2.1.b. consisted predominately of data produced by measuring FL immediately following a one minute exposure and 4h exposure, and to a lesser but equal extent from FL assays measured at 24h, 48h and 72h following the initial test material exposure. For FL (%) data, the range of values covered by the SD and CV values were similar although the CV values were higher values in comparison to the SD values.

Figure 3.2.1.c. consisted of data produced from a single study where only mild surfactants were tested. Data are shown for FL measured immediately following a one minute exposure and then at 4h, 24h, 48h and 72h following the initial test material exposure. The CV values were higher in comparison to the SD values and covered a wider range.

For the larger data sets, the SD values increased as the mean FL₂₀ (mg/ml) and FL% values increased. The CV measurement, which exhibited random distribution across the range of FL₂₀ (mg/ml) values and FL% values, was selected for analyses.

3.2.2. Results and discussion

Table 3.2.2. shows the summary results for each study featured in table 3.1.. For each test material, greater-than values were not included in the calculations of mean, SD and CV. In some cases a single value was used to determine the mean as the greater-than values generated by some experimental repeats were not included in the calculation. For each study, the overall mean, overall median, mean SD, median SD, mean CV and median CV were calculated from the mean, SD and CV for each test material. For test materials that only produced greater-than values, no mean, SD or CV was calculated for that test material. For studies that contained some test materials with no mean, mean SD or mean CV values, the overall mean, overall median, mean SD, median SD, mean CV and median CV values could not be calculated. Annex II shows the raw data used to produce the summary values in table 3.2.2., and the mean, mean SD and mean CV for each test material where calculated.

Table 3.2.2. Summary table of the overall reproducibility per study. There were 25 data sets (compared to 14 in table 3.1) as some studies contained multiple data sets (e.g. ECVAM Prevalidation study (Southee, 1998)).

INVITTOX Protocol No. 71

Study	No. of Test Substances	No. of Experiments	Data Format	Overall Mean**	Mean SD*	Mean CV (%)*	Overall Median**	Median SD*	Median CV (%)*
Report on Comparison of 40 Cosmetic and Domestic Formulations Supplied by Company # 8 and Evaluated by 3 <i>In Vitro</i> Cytotoxicity Tests (FRAME, 1992)	40	4	FL20 (mg/ml) T0	24.8	-	-	5.2	-	-
Company #5 Chemicals (FRAME, 1992)	6	n≤ 5	FL20 (mg/ml) T0 and T72	-	-	-	-	-	-
Final Report on Testing of 12 Mild Surfactants supplied by Company # 5 for Cytotoxicity Testing at the FAL (FRAME, 1992)	12	n≤ 4	FL20 (mM) T0, T4, T24, T48	-	-	-	-	-	-
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994)	10	n≤ 5	FL20 (mg/ml) T0 and T72	5.1	2.8	63.2	5.5	2.6	58.6
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994) –re-runs based on activity	10	n≤ 5	FL20 (mg/ml) T0 and T72	16.2	9.1	56.5	11.2	5.9	56.8

Study	No. of Test Substances	No. of Experiments	Data Format	Overall Mean**	Mean SD*	Mean CV (%)*	Overall Median**	Median SD*	Median CV(%)*
The Evaluation of Pesticide Ingredients and Formulations <i>In vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995)	4	4	FL20 (mg/ml) T0 and T72	-	-	-	-	-	-

INVITTOX Protocol No. 82

Study	No. of Test Substances	No. of Experiments	Data Format	Overall Mean**	Mean SD*	Mean CV (%)*	Overall Median**	Median SD*	Median CV(%)*
Company # 5 test concentration X (FRAME, 1993)	12	n≤ 4	FL(%) T0, T24, T48, T72	6.1	6.57	94.9	0.3	0.3	98
Company # 5 test concentration XI (FRAME, 1993)	12	n≤ 4	FL(%) T0, T24, T48, T72	26.2	-	-	16.8	-	-

INVITTOX Protocol No. 120

Study	No. of Test Substances	No. of Experiments	Data Format	Overall Mean**	Mean SD*	Mean CV (%)*	Overall Median**	Median SD*	Median CV(%)*
A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom <i>et al.</i> , 1997) FRAME	30	4	FL20 (mg/ml) T4	-	-	-	-	-	-
ECVAM Prevalidation Study (Southee, 1998) Phase II –FRAME Laboratory	5	n≤5	FL20 (mg/ml) T0 and T4	129.9	35.6	36.9	19.7	7.9	42.2
ECVAM Prevalidation Study (Southee, 1998) Phase II –Company # 7 Laboratory	5	n≤5	FL20 (mg/ml) T0 and T4	250.3	-	-	24.1	-	-
ECVAM Prevalidation Study (Southee, 1998) Phase II –ECVAM Laboratory	5	n≤5	FL20 (mg/ml) T0 and T4	149.5	-	-	13.9	-	-
ECVAM Prevalidation Study (Southee, 1998) Phase II – Company # 3 Laboratory	5	n≤5	FL20 (mg/ml) T0 and T4	-	-	-	-	-	-

Study	No. of Test Substances	No. of Experiments	Data Format	Overall Mean**	Mean SD*	Mean CV (%)*	Overall Median**	Median SD*	Median CV(%)*
ECVAM Prevalidation Study (Southee, 1998) Phase II –FRAME Laboratory	1 -Positive Control Data	n=3	FL (%) T0 and T4	79.7	13.9	17.4	79.7	13.9	17.4
ECVAM Prevalidation Study (Southee, 1998) Phase II – Company # 7 Laboratory	1 -Positive Control Data	n=6	FL (%) T0 and T4	11.4	4.9	73.6	11.4	4.9	73.6
ECVAM Prevalidation Study (Southee, 1998) Phase II –ECVAM Laboratory	1 -Positive Control Data	n=23	FL (%) T0 and T4	12.8	7.7	67.3	12.8	7.7	67.3
ECVAM Prevalidation Study (Southee, 1998) Phase II – Company # 3 Laboratory	1 -Positive Control Data	n=5	FL (%) T0 and T4	17.1	5.7	35.3	17.1	5.7	35.3
ECVAM Prevalidation Study (Southee, 1998) Phase III –FRAME Laboratory	1 -Positive Control Data	n≤19	FL (%) T0 and T4	4.3	2.7	128.0	4.3	2.7	128.0
ECVAM Prevalidation Study (Southee, 1998) Phase III –ECVAM Laboratory	1 -Positive Control Data	n≤15	FL (%) T0 and T4	9.3	1.9	35.4	9.3	1.9	35.4
ECVAM Prevalidation Study (Southee, 1998) Phase III – Company # 3 Laboratory	1 -Positive Control Data	n≤5	FL (%) T0 and T4	10.5	3.8	603.0	10.5	3.8	603.0

Study	No. of Test Substances	No. of Experiments	Data Format	Overall Mean**	Mean SD*	Mean CV (%)*	Overall Median**	Median SD*	Median CV (%)*
ECVAM Prevalidation Study (Southee, 1998) Phase III –FRAME Laboratory	10	n≤3	FL20 (mg/ml) T0 and T4	50.6	6.7	28.5	1.9	0.6	16.4
ECVAM Prevalidation Study (Southee, 1998) Phase III –ECVAM Laboratory	10	n≤5	FL20 (mg/ml) T0 and T4	28.0	-	-	7.6	-	-
ECVAM Prevalidation Study (Southee, 1998) Phase III – Company # 3 Laboratory	10	n=2	FL20 (mg/ml) T0 and T4	-	-	-	-	-	-

INVITTOX Protocol No. 86

Study	No. of Test Substances	No. of Experiments	Data Format	Overall Mean**	Mean SD*	Mean CV (%)*	Overall Median**	Median SD*	Median CV (%)*
CTFA Phase III TEP assay data (submitted by CTFA)	23	n≤4	EC50 (%) T0	3.7	0.4	13.5	1.9	0.2	14.3
TEP assay Formulation data (submitted by Company # 3)	41	n≤3	EC50 (%) T0	2.8	0.2	8.1	2.1	0.1	7.3

* calculated from the SD and CV calculated for each test material

** calculated from the mean value for each test material

i. Protocols

The FL₂₀ (mg/ml) data were predominately produced by the FAL using INVITTOX Protocol No. 71 or a similar protocol with slight modifications. INVITTOX Protocol No. 71 was based on the protocol of Tchao (1988). Data from the Hubbard *et al.*, (1994) paper used the protocol of Shaw *et al.*, (1990). This was one of the first FL assay protocols published and was very similar to the first FL assay protocol developed by Tchao (1988). An advantage of the data provided by the FAL is that it contained information regarding the time and date of the individual experiments. This information enabled analyses of operator and time variability to be carried-out (Section 3.2.3.).

FL₂₀ (mg/ml) data were also produced using INVITTOX Protocol No. 120, or a slightly modified protocol, which was performed by a number of different laboratories as part of the ECVAM Prevalidation study (Southee, 1998). Phase II of the ECVAM Prevalidation study used a slightly different protocol to the Phase III protocol that was later adopted as INVITTOX Protocol No. 120. Raw positive control data for 0.16mg/ml SDS were also available from the ECVAM Prevalidation study Phase II and Phase III for which the endpoint was FL%.

There were fewer data for INVITTOX Protocol No. 86 which were produced by Company # 3. There were even fewer data for a protocol similar to INVITTOX Protocol No. 82; results were reported for 12 surfactants that were tested at two different concentrations.

A summary of all the differences between the protocols discussed in this section, including those not known to be significant, are given in table 3.2.2.1.

Table 3.2.2.1. The main protocol differences between the FL assay protocols for which data were available to assess intra-laboratory variability.

Protocol Step	Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase II) (Southee, 1998)	Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	FLT- SOP Company # 4 INVITTOX Protocol No. 120 (developed as consequence of ECVAM Prevalidation Study, 1998)	Loss of Trans-epithelial Impermeability of a Confluent Layer of Madin-Darby Canine Kidney (MDCK) Cells as a Determinant of Ocular Irritancy Potential (Shaw <i>et al.</i> , 1990)	Fluorescein Leakage Test. INVITTOX Protocol No. 71	Trans-epithelial Permeability of Fluorescein <i>In Vitro</i> as an Assay to Determine Eye Irritants (Tchao, 1988)
Cell Type/Strain	MDCK NBL-2 (ECACC: 85011435)	MDCK NBL-2 (ECACC: 85011435)	MDCK NBL-2 (ATCC: CCL34)	MDCK CB997 (ECACC: 84121903)	MDCK CB997 (ECACC: 84121903)	MDCK
Seeding Density	2x10 ⁵ cells/ml	2x10 ⁵ cells/ml	4x10 ⁵ cells/ml (500µl/insert= 2x10 ⁵ cells/insert)	10 ⁵ cells/insert (in 400µl medium)	4x10 ⁵ cells/ml (400µl/insert)	1.5x10 ⁵ cells/insert
Medium Type	MEM	MEM	MEM w/o phenol red	DMEM/Ham's F12 (1:1)	DMEM x1 (L-glutamine and 15mM HEPES) /Ham's F12 (1:1)	MEM
Insert Type's, Pore Size's, coating	Millicel-HA, 12mm diameter (0.45µm pore size)	Millicel-HA, 12mm diameter (0.45µm pore size)	Millicel-HA, 12mm diameter (0.45µm pore size)	Anocell 10	Millicel-HA, 12mm diameter (0.45µm pore size), Anocell 10 inserts can also be used	Millicell with HATF (surfactant-free) membrane
Rinsing before Material Incubation (solution and volume)	Insert washed with 500µl HBSS with Ca ⁺⁺ , Mg ⁺⁺ in insert and well. Prior to treatment, medium from under insert is removed, replaced with HBSS Ca ⁺⁺ , Mg ⁺⁺	Insert washed with 500µl HBSS with Ca ⁺⁺ , Mg ⁺⁺ in insert and well. Prior to treatment, medium from under insert is removed, replaced with HBSS w/o Ca ⁺⁺ and Mg ⁺⁺	N	500µl distilled water	N	HBSS
Material Exposure (mins)	15	15	15	1	1 or 15	15

Protocol Step	Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase II) (Southee, 1998)	Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	FLT- SOP Company # 4 INVITTOX Protocol No. 120 (developed as consequence of ECVAM Prevalidation Study, 1998)	Loss of Trans-epithelial Impermeability of a Confluent Layer of Madin-Darby Canine Kidney (MDCK) Cells as a Determinant of Ocular Irritancy Potential (Shaw <i>et al.</i>, 1990)	Fluorescein Leakage Test. INVITTOX Protocol No. 71	Trans-epithelial Permeability of Fluorescein In Vitro as an Assay to Determine Eye Irritants (Tchao, 1988)
Material Volume (µl)	200µl	200µl	500µl	100µl	200µl	500µl
Solvent	HBSS containing Ca ⁺⁺ and Mg ⁺⁺ or light mineral oil. Prepared just before use in case of emulsion and/or suspensions	HBSS with Ca ⁺⁺ and Mg ⁺⁺ , or light mineral oil. Prepared just before use in case of emulsion and/or suspensions	HBSS	tested neat or in distilled water	HBSS or mineral oil	HBSS
Fluorescein Addition Outline	Treated inserts transferred to new 24WP containing 500µl HBSS with Ca ⁺⁺ , Mg ⁺⁺ /well. 200µl 0.01% fluorescein added to each insert.	Treated inserts transferred to new 24WP containing 500µl HBSS with Ca ⁺⁺ , Mg ⁺⁺ /well. 200µl 0.01% fluorescein added to each insert.	Place washed inserts into new 24WP with 500 µl warm HBSS, 500µl 0.01% Na-fluorescein in HBSS put into each insert	Following washing step, inserts removed to fresh wells containing 500µl warm Dulbecco's Ca ²⁺ PBS. 500µl 0.02% w/v fluorescein in Ca ²⁺ PBS added and plate incubated at RT	Inserts checked for damage and moved to 24WP with 400µl HBSS; 400µl 0.01% Na-fluorescein in HBSS added to insert	Fluorescein added to each insert which is placed in fresh 24WP containing 500µl buffer
Fluorescein Volume added (µl)	200µl	200µl	500µl	500µl	400µl	500µl
Fluorescein Concentration delivered to monolayer (mg/ml)	0.01% (w/v) Na-fluorescein in HBSS with Ca ²⁺ and Mg ⁺⁺	0.01% (w/v) Na-fluorescein in HBSS with Ca ⁺⁺ and Mg ⁺⁺	0.01mg/ml Na-fluorescein in HBSS (0.001% (w/v))	0.02% (w/v) Na-fluorescein in HBSS	0.1mg/ml Na-fluorescein HBSS (0.01% (w/v)) (0.02% Na-fluorescein HBSS for Anocell inserts)	0.02% Na-fluorescein in HBSS

Protocol Step	Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase II) (Southee, 1998)	Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	FLT- SOP Company # 4 INVITTOX Protocol No. 120 (developed as consequence of ECVAM Prevalidation Study, 1998)	Loss of Trans-epithelial Impermeability of a Confluent Layer of Madin-Darby Canine Kidney (MDCK) Cells as a Determinant of Ocular Irritancy Potential (Shaw <i>et al.</i>, 1990)	Fluorescein Leakage Test. INVITTOX Protocol No. 71	Trans-epithelial Permeability of Fluorescein <i>In Vitro</i> as an Assay to Determine Eye Irritants (Tchao, 1988)
Time allowed for Fluorescein Leakage (mins)	30±2 Incubator	30±2 Incubator	30 RT	60 RT	30 RT	30 24°C
Vehicle Control	HBSS	HBSS	HBSS	medium	HBSS	HBSS
Endpoint	FL20 (mg/ml) T0, T4	FL20 (mg/ml) T0, T4	FL20, FL10 (mg/ml) T0, T4	FL20, FL50 (mg/ml) T0	FL20, FL50 (mg/ml) T0	EC50 (%) T0
Optical Density Wavelengths (excitation/emission)	Plate gently shaken for 1min, then 200µl samples taken from the plate and transferred to 96WP 485-490nm/650nm (excitation/emission)	Plate gently shaken for 1min then 200µl samples taken from the plate and transferred to 96WP 485-490nm/650nm (excitation/emission)	485nm/530nm (excitation/emission)	492nm/620nm (excitation/emission)	485nm/530nm (excitation/emission)	490nm

No study has been reported which compared the within-laboratory variability for INVITTOX Protocol No. 71 or for INVITTOX Protocol No. 120. From the FL assay data available to this BRD, the CV values were plotted against the mean values for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120, as an approach to compare their relative reproducibilities. This analysis has limitations as different types of materials have been tested in the two protocols. Data from INVITTOX Protocol No. 71 and from similar protocols (i.e. Shaw *et al.*, (1990) were plotted; data from INVITTOX Protocol No. 120 and from similar protocols (i.e. ECVAM Prevalidation Study) were also plotted (figure 3.2.2.1.).

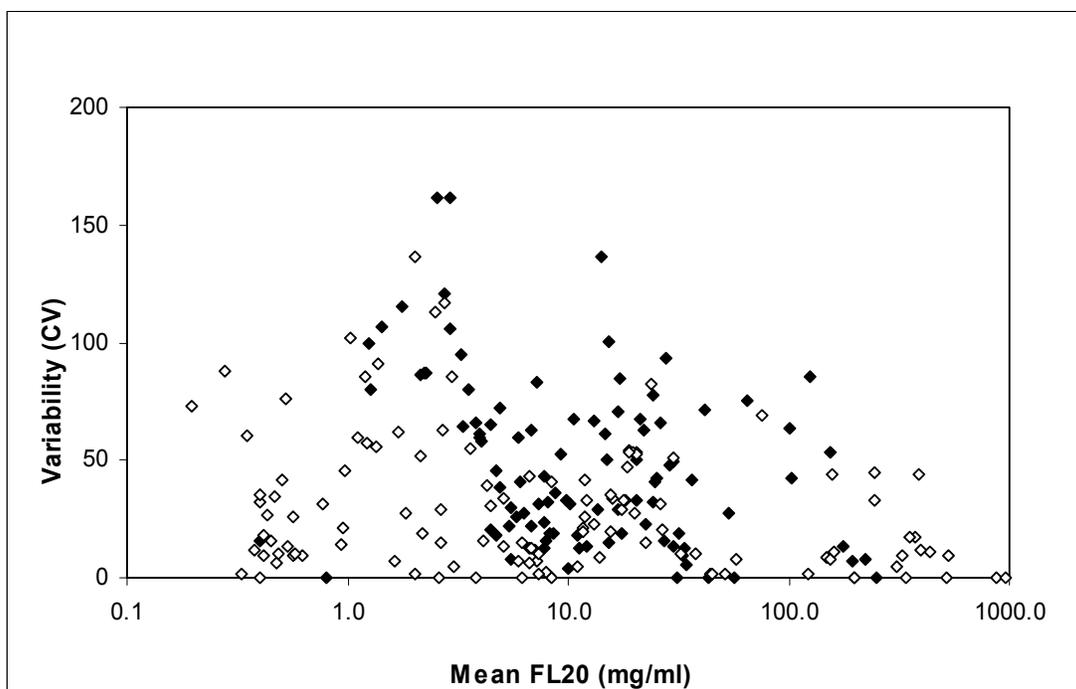


Figure 3.2.2.1. The CV values for INVITTOX Protocol No. 71 \blacklozenge , and INVITTOX Protocol No. 120 \diamond plotted against FL assay results for all available raw data. Materials that had greater-than or less-than FL_{20} (mg/ml) values were not plotted ($n \geq 2$).

There were a similar number of data points for both INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120. INVITTOX Protocol No. 71 was employed in a greater number of smaller studies (i.e., fewer test materials) in comparison to INVITTOX Protocol No. 120 which was carried out in fewer but larger-scale studies where many more materials were tested, i.e. ECVAM Prevalidation study (Southee, 1998) and the COLIPA Eye Irritation International Study (Brantom *et al.*, 1997). In addition, data from INVITTOX Protocol No. 71 was produced exclusively from the FAL, whilst INVITTOX Protocol No. 120 was carried out by a number of different laboratories. The data plotted in figure 3.2.2.2. show that a wider range of FL_{20} (mg/ml) values were produced by the materials tested using INVITTOX Protocol No. 120 and that intra-laboratory variability was comparable for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120 over the same range of FL_{20} (mg/ml) values.

TEP assay data submitted by Company # 3, and from the CTFA study Phase III were plotted (figure 3.2.2.2.). The CV was plotted against the mean EC_{50} values for the test formulations. The data were from an in-house protocol of a robotic TEP assay. The

assay was essentially the same as INVITTOX Protocol No. 86 which was developed by Company # 3. The CTFA Phase III compositions were known, whilst only an outline of the compositions for 17 of the 41 formulations tested in-house by Company # 3 was provided.

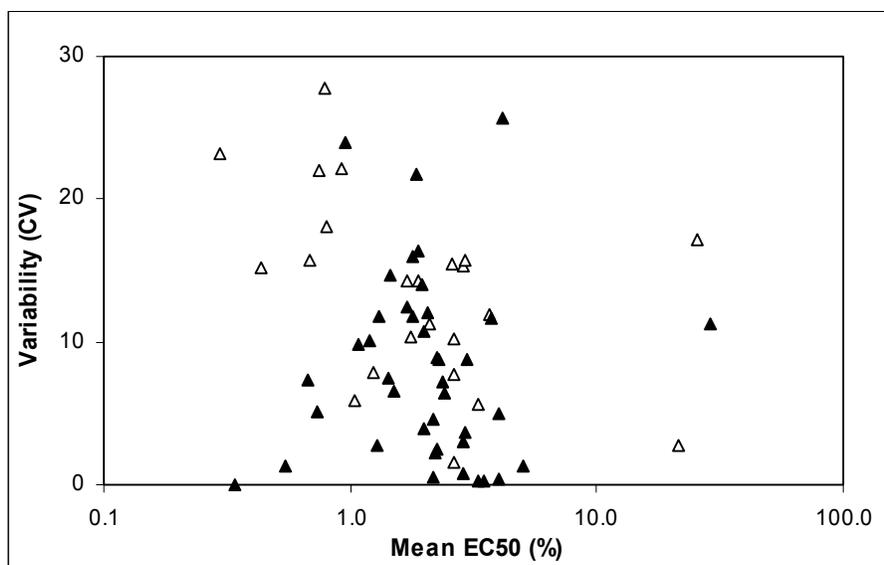
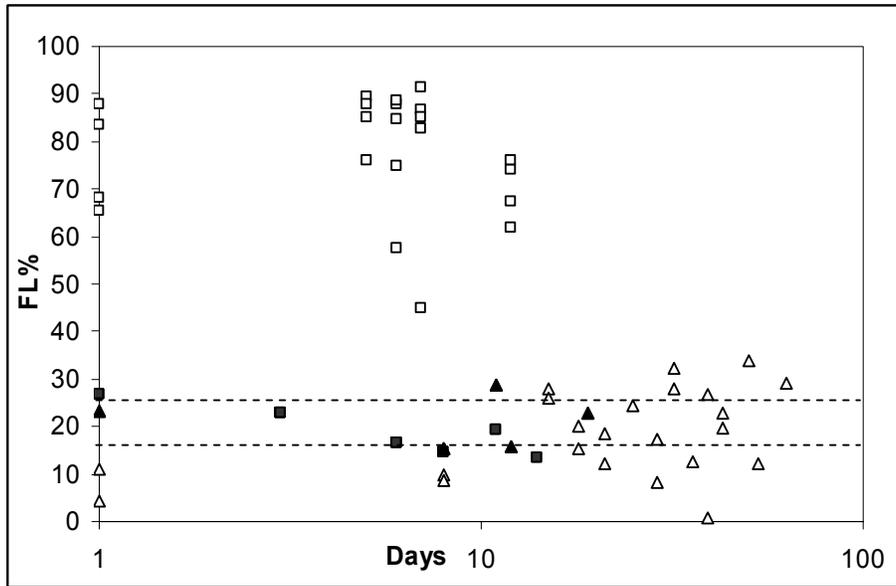


Figure 3.2.2.2. The CV values for TEP assay data plotted against the mean TEP assay values ($n=2$) for Company # 3 formulations ▲, and CTFA Phase III formulations △.

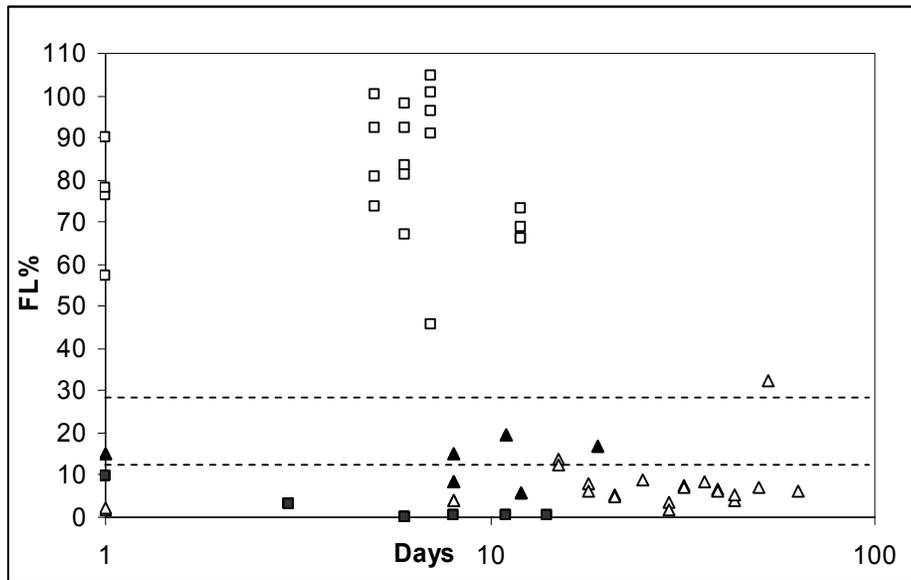
The core range of mean EC_{50} values was relatively small in comparison to the data sets for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120. Additionally, the CV values produced by the data set were smaller in comparison to INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120. This could be due to the use of EC_{50} values rather than FL_{20} values. FL_{20} values are likely to produce more variable results than EC_{50} values as the concentrations which cause 20% effects, tend to be more variable as they are at the start of a dose response curve, whilst EC_{50} concentrations are more likely to be in the linear part of the dose-response curve. It is difficult to directly compare the level of variability produced by the various FL assay protocols due to the different format of the data, although the CV values did seem generally lower for the TEP assay. This finding is indicative of an established system that is in routine use.

Positive Control Data over time, from the ECVAM Prevalidation study (Southee, 1998)

The ECVAM Prevalidation study (Phase II and Phase III) (Southee, 1998) was one of few studies featuring a FL assay protocol to have well documented positive control data. The test protocol was similar to INVITTOX Protocol No. 120. The level of FL (%) that resulted from the positive control (0.16mg/ml SDS) was plotted against time, to determine variability (figure 3.2.2.3. a-b). The number of data points varied per laboratory, as they differed in their use of the positive control. The FAL and ECVAM performed a positive control per chemical/plate, whilst Company # 7 and Company # 3 used the same positive control data for all chemicals performed within that run.



a.



b.

Figure 3.2.2.3. The effect of time on the ECVAM Prevalidation Phase II SLS positive control FL% values for the four participating laboratories at a.) T0; b.) T4; □, FAL; ■, Company # 7; ▲, Company # 3; △, ECVAM. ----- Pre-defined range in which the positive control values should fall for the FL assay results from the test chemicals to be accepted. A version of INVITTOX Protocol No. 120 was used.

For all laboratories at both time-points, the FL positive control results did not vary with time (day) (figure 3.2.2.3.). Where two or more positive control values were generated for a single day, (FAL, ECVAM) the degree of variation between the two points showed no increase or decrease in relation to time. Although the FAL performed the experiments within a shorter period of time than the other laboratories, the range of FL assay values was greater for the FAL at T0 and T4. In this case, variation between

experimental runs did not increase in proportion to the time taken to perform the experiments. In comparison to T0 data, the range of FL% values for the FAL at T4 increased slightly whilst the range of FL% values decreased slightly for the other three laboratories.

Similar analyses were performed using the ECVAM Prevalidation study Phase III positive control data (results not shown). The range of data was similar to that observed in figure 3.2.2.3.b. FAL produced positive control data within a range similar to the other laboratories, but all results were outside the defined acceptable values. Despite slight protocol modifications and an increase in the range of acceptability to 15-30%, there was only a slight improvement in the number of assays from all laboratories with acceptable positive control results. Although many of the FL values were outside the acceptance range, three out of four laboratories produced data within the range of 0-30% whilst the data from the FAL ranged from 50-100%.

ii. Test Materials

The number and types of materials tested in the protocols varied considerably. The number of materials and the range of *in vivo* irritancy levels for each of the studies are shown for those studies reporting *in vivo* data (table 3.2.2.2.). The table provides information regarding the range of *in vivo* ocular irritation covered by the studies but does not include the majority of data analysed in this section as they were without corresponding *in vivo* data.

Table 3.2.2.2. Relevant *in vitro* and *in vivo* information for the types of test materials and potencies.

Study	Types of materials tested	Range of FL ₂₀ values	Type of <i>in vivo</i> data	Range of <i>in vivo</i> scores
The Evaluation of Pesticide Ingredients and Formulations <i>In Vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995)	2 pesticide chemicals and 2 formulations	1- >250mg/ml	Draize (MAS or MMAS) scores upto 96h	0-57
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase II) (Southee, 1998)	4 cosmetic ingredients and 1 formulation	0.006mg/ml -neat	MAS	0-44
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	10 mild surfactants	0.3- 470.5mg/ml	MMAS	0-37
CTFA Evaluation of Alternatives Program: Phase III, TEP assay (Gettings <i>et al.</i> , 1996)	25 surfactant-based formulations	0.24 – 30.3 (EC50(%))	MAS	2.3-43
Ocular Irritancy Assessment of Cosmetics Formulations and Ingredients: Fluorescein Leakage Test. (Zanvit <i>et al.</i> , 1999) COLIPA DATA	11 cosmetic surfactants and 23 (Company # 4) 20 (FAL) surfactant-based formulations	0.2- >1072mg/ml	MMAS	0.67-108

INVITTOX Protocol No. 71 raw data were predominately available for surfactants and surfactant-based formulations; there were data for only two pure chemicals which were

pesticides. These data represent the fact that the FL assay has been predominately used by the cosmetics industry for product development.

Raw INVITTOX Protocol No. 120 data were available for fewer different types of materials; the majority were surfactants. INVITTOX Protocol No. 120 was tested in the ECVAM Prevalidation study Phase III; ten mild surfactants were tested in four different laboratories. The CVs for the surfactants were plotted against the mean FL₂₀ (mg/ml) data T4 (figure 3.2.2.4.).

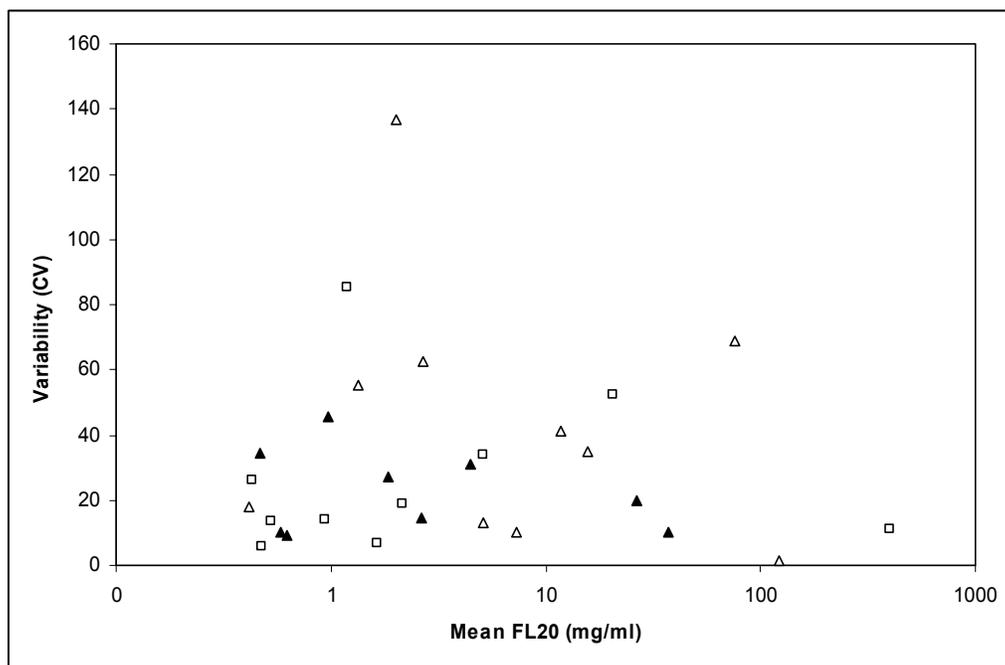


Figure 3.2.2.4. The CV values plotted against mean FL₂₀ (mg/ml) T4 values for the ten surfactants tested in three laboratories; □, FAL; ▲, Company # 3; △, ECVAM (n≥2). Greater-than values were not included in the calculation of the means and were not plotted.

The Company # 4 SOP (1992), which formed the basis for INVITTOX Protocol No. 120, was tested in the COLIPA Eye irritation international validation program (Phase I) (Brantom *et al.*, 1997); only surfactants and surfactant-based solutions soluble in HBSS were tested. The compositions of these test materials are provided (Formulation Annex B). The data from the COLIPA study were included in the analyses for INVITTOX Protocol No. 120. The test materials were tested by Company # 4 and the FAL but data were only available from the FAL. The CVs for the test materials were plotted against the mean FL₂₀ (mg/ml) data (figure 3.2.2.5.).

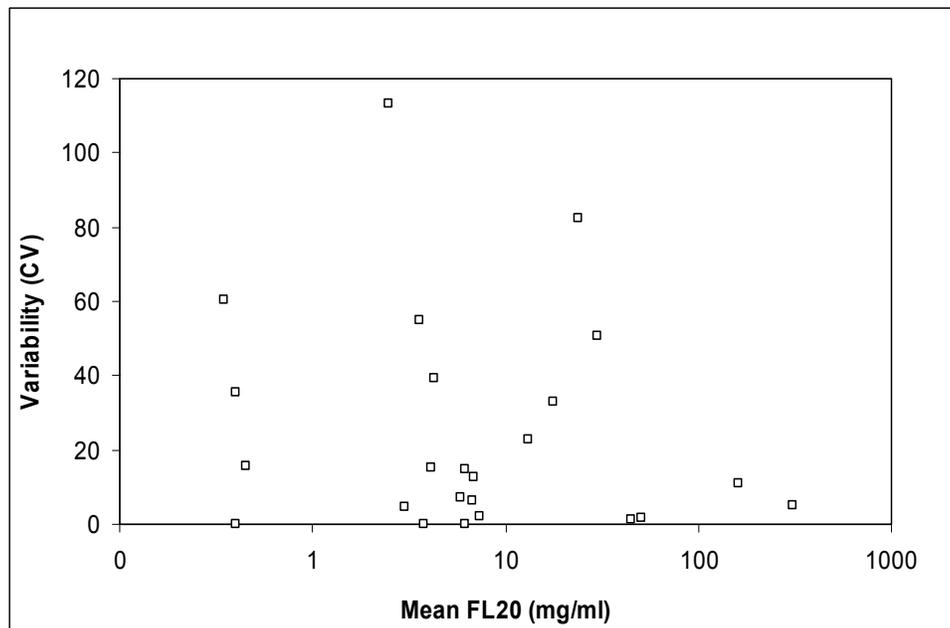
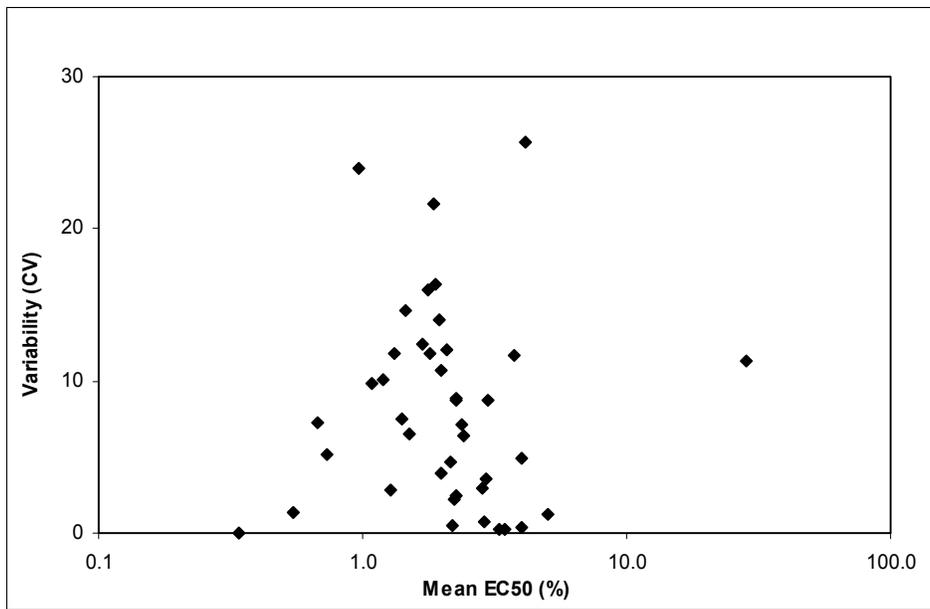
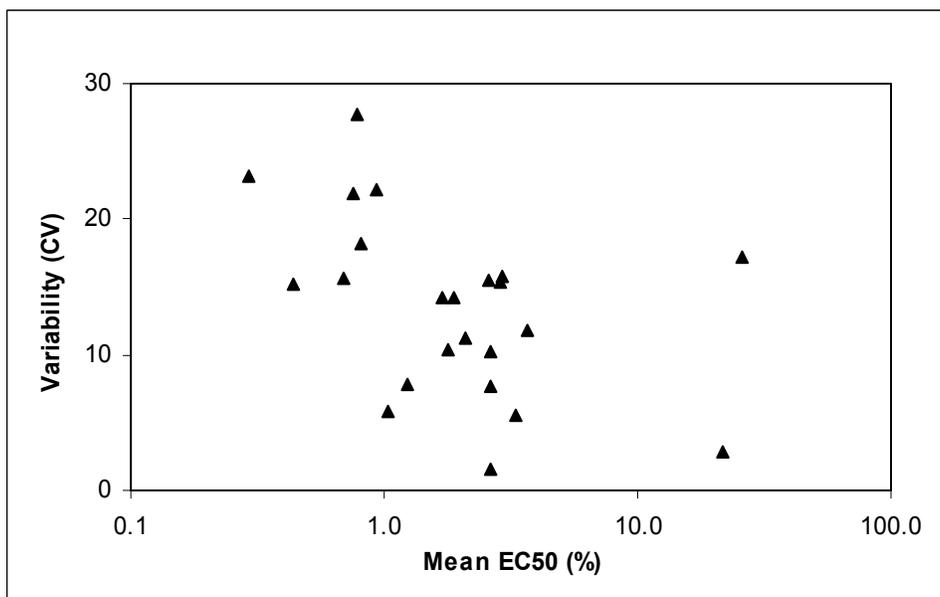


Figure 3.2.2.5. The CVs plotted against the mean T4 FL₂₀ (mg/ml) values for surfactants and surfactant-based formulations tested by the FAL as part of the COLIPA study (n≥2). Greater-than values were not included in the calculation of the mean and were not plotted. The Company # 4 SOP (1992), which formed the basis for INVITTOX Protocol No. 120, was used. .

The compositions of the formulations for the TEP assay data submitted by Company # 3 were only provided for 17 formulations (Formulation Annex C). The *in vivo* Draize data for these 17 formulations indicated that the products were mostly 'mild' and 'moderate' irritants and relatively few were 'severe' irritants. The formulations used in Phase III of the CTFA study were generic surfactant-based formulations representative of those found in the cosmetics industry. The compositions of the CTFA Phase III formulations were available (Formulation Annex A), and they were known to be predominately non-irritants and severe irritants. The CVs were plotted against the mean TEP assay results for formulations tested by Company # 3 (figure 3.2.2.6.a.) and the CTFA study Phase III formulations (figure 3.2.2.6.b.).



a.



b.

Figure 3.2.2.6. The CVs plotted against the mean EC₅₀ (%) values for a. formulations tested by Company # 3 (n≥2); b. CTFA Phase III data (n≥3). INVITTOX Protocol No. 86 was used to generate both data sets.

In comparison to the data sets for INVITTOX Protocol No.71 and INVITTOX Protocol No.120, the data range for INVITTOX Protocol No. 86 was limited. This was probably due to the data supplied here was from Company # 3 which routinely use the TEP assay to test surfactants and surfactant-based formulations only. The core range of mean EC₅₀ values was relatively small (1-10%) in comparison to the data sets for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120 (1-100mg/ml). Additionally, the CV values produced by the data set were relatively small in comparison to the other protocols.

iii. Classifications

The FL assay can produce FL₂₀ values that cover up to six orders of magnitude (table 3.2.2.2.). A potential advantage of such a wide range of FL₂₀ values is that large ranges of values can be used to assign the same classification if there is equal correlation with FL₂₀ values and *in vivo* scores across the entire range of FL₂₀ values. This entails that assay reproducibility does not need to be high in order to obtain reproducible predicted classifications for *in vivo* ocular irritation.

FL assay PMs featured in the CTFA study Phase III (Gettings *et al.*, 1996), COLIPA study (Brantom *et al.*, 1997), and the ECVAM Prevalidation study (Southee, 1998). The PMs were comprised of ranges of FL assay values that were correlated to *in vivo* scores and/or classifications of irritancy. None of the classification PMs that were applied to the FL assay study data were for the EU Risk Phrase Classification system, the GHS classification system, or the EPA classification system. The predicted classifications were assigned to the raw data (i.e. each experimental run) to determine to what extent variable FL assay values affected the resulting predicted classifications (Annex II). The classification systems as featured in the literature were used (rather than the EU, GHS and EPA classification systems).

The PM used in the COLIPA study (Brantom *et al.*, 1997) was defined using historical data for 43 test materials, prior to the testing of COLIPA test materials. The COLIPA PM (table 3.2.2.3.) was also used in the ECVAM Prevalidation study (Southee, 1998).

Table 3.2.2.3. PM from the COLIPA study for a version of INVITTOX Protocol No. 120.

FL20 (mg/ml) T4	Classification	Draize MMAS
>100	Non-irritant/slight	<15
20-100	Moderate	15-30
<20	Irritant/severe	>30

T4= Four hour time-point

There were 30 COLIPA test materials for which raw data were available from the FAL. Of these materials, three had mixed classifications based on the results from the various experimental repeats (n≥2). All three materials with mixed classifications were classified as having both 'moderate' and 'irritant/severe' levels of ocular irritancy. Importantly, the assay and PM appeared able to distinguish irritants from non-irritants. However, these results were for only surfactants and surfactant-based materials that were soluble in HBSS.

The same PM as featured in the COLIPA study was applied to the ECVAM Prevalidation Study Phase III FL₂₀ (mg/ml) T4 data (Southee, 1998) as the test protocols in each study were very similar. The number of mixed classifications from the three different laboratories are presented (table 3.2.2.4.).

Table 3.2.2.4. Variation in classifications assigned according to the COLIPA PM for the three laboratories that participated in the ECVAM Prevalidation study Phase III (n ≥2).

Laboratory	No. of materials tested	No. and combination of mixed classifications
FAL	9	1x moderate and irritant/severe
ECVAM	10	1x moderate and irritant/severe 1x non-irritant/slight and moderate
COMPANY # 3	10	0

Due to the very few materials which produce data leading to mixed classifications, the results suggested that the protocol was reproducible in terms of classifications. The different laboratories showed very similar levels of reproducibility, i.e. the number of mixed classifications for any given test material ranged from 0-1. It is known to be difficult for *in vitro* tests to distinguish different levels of irritancy so it is not unusual that the 2/3 of the mixed classifications were comprised of moderate and irritant/severe classifications. Importantly, the FL assay protocol distinguished irritants and non-irritants. Those materials that were classified as irritants according to the *in vivo* data and not detected as irritants according to the FL assay results would be of greater concern.

In the CTFA study Phase III publication (Gettings *et al.*, 1996), a post-hoc threshold value of ≤2.60 (%) was assigned to the TEP assay data, to try to classify irritants and non-irritants according to the FHSA classification system (Gettings *et al.*, 1996). Based upon this threshold value, the FHSA classifications were assigned to the raw data to determine if for a given test material the various TEP assay experimental runs led to different irritancy classifications (Annex II). Of 23 surfactant-based formulations tested, six produced TEP assay results which led to mixed classifications (n≥3).

No *in vitro* classification system was available for the Company # 8 test formulations that were tested using INVITTOX Protocol No. 71 (data included in figure 3.2.2.1.). As the compositions of the test formulations were unknown it was difficult to confidently apply another PM to this data. In addition, much of the raw Company # 8 data (Annex II) were produced by using different exposure periods, which reduced the number of experimental repeats per test material with the same exposure period.

In general, all of the analyses from the different studies were slightly skewed as for some formulations there were more experimental repeats than others. A greater number of experimental repeats provides greater opportunity for more diverse classifications to be assigned to the same test material. However, a high number of experimental repeats generally only occur in the case of low reproducibility in the prior experimental runs.

3.2.3. Additional analyses of operator and time variability

Ideally, operator and time variability would have been assessed using results from a positive control chemical. However, there were a limited number of available sets of results that had positive control data. Subsequently, the approach taken was based on identifying outlier experimental runs.

The mean, SD and CV were calculated for each test material. As the CV does not take into account the size of the numbers analysed, it is more difficult to attain a low CV value with numbers of lower values than with numbers of higher values. Therefore the following rules were applied for data handling:

- if the mean FL₂₀ was ≤ 20 mg/ml or 20mM, outlier experimental runs were removed in order to attain a CV $\leq 60\%$
- if the mean FL₂₀ was ≥ 20 mg/ml or 20mM, outlier experimental runs were removed in order to attain a CV $\leq 40\%$

The experimental run(s) producing the value that caused a CV greater than 40% or 60% were identified. For this type of analysis, there were a number of criteria that needed to be fulfilled:

- only FL₂₀ mg/ml or FL₂₀ mM data were used
- the number of experimental runs per test material had to be greater than three (data sets that had only three experimental repeats for some test materials were also included if the majority of test materials in the data set had more than three experimental runs).
- for test materials that did not have three experimental repeats, all experimental runs for that test material were removed from the data set and were not included in any analyses.
- the number of experimental runs remaining, after removal of those causing the CV to be $\geq 40\%$, had to be at least three per test material
- 'greater-than' values were changed to actual values, i.e. '>500' became '500.'

For determining 'operator variability' the name of the operator(s) who produced the outlier value(s) was recorded. For determining 'time variability,' the date that produced the outlier value(s) was recorded. The number of times that the various operators and dates were identified as causing a CV greater than either 40% (or 60% depending on the mean FL₂₀ value) were recorded and presented as a proportion of the number of experiments they had performed for the entire data set (table 3.2.3.1.). If the removal of data points did not attain a CV below 40% or 60%, the 'removed' data points were not counted as outliers, but were counted in the total number of experiments that had been performed in the data set. This allowed one to observe if a single operator, or combination of operators, were consistently responsible for test materials that produced CVs above the set threshold.

For test materials that did not attain an acceptable CV value, it was assumed that the test material had inherently variable properties that prevented reproducible results from being attained. In some cases the test material was known to have variable properties, but often the test material was coded so inherent variation could only be assumed by the variable results. Inclusion of this data enabled the robustness of the various FL assay protocols to be fully tested.

Table 3.2.3.1. Results and relevant information for each study

Study	Operator Variability	Time Variability	Operator No.	Exp. No.	Well Replicate No.	No. of Test Substances a., b., c.,	Data Format
A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom <i>et al.</i> , 1997) FRAME	n/a	18.02.94. 1/3 21.02.94. 1/5 28.02.94. 0/5 04.03.94. 0/5	not stated	4	3	a. 5 b. 0 c. 0	FL20 (mg/ml) T0
A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom <i>et al.</i> , 1997) FRAME	n/a	18.02.94. 1/4 21.02.94. 0/5 28.02.94. 0/5 04.03.94. 0/5	not stated	4	3	a. 5 b. 0 c. 0	FL20 (mg/ml) T72
The Evaluation of Pesticide Ingredients and Formulations <i>In vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995)	n/a	13.07.92. 0/4 17.07.92. 0/4 20.07.92. 0/4 20.07.92. 0/4	1	4	3	a. 4 pesticide formulations b. 0 c. 0	FL20 (mg/ml) T0
The Evaluation of Pesticide Ingredients and Formulations <i>In vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995)	n/a	16.07.92. 2/4 17.07.92. 0/4 23.07.92. 0/4 30.07.92. 0/4	1	4	3	a. 4 pesticide formulations b. 0 c. 1	FL20 (mg/ml) T72

Study	Operator Variability	Time Variability	Operator No.	Exp. No.	Well Replicate No.	No. of Test Substances a, b., c.,	Data Format
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994)	n/a	21.09.92. 0/6 28.09.92. 0/4 08.10.92. 0/6 09.10.92. 0/4 13.10.92. 1/6 11.01.93. 0/6 15.11.92. 2/10	not stated	n≤ 5	3	a. 10 surfactants b. 0 c. 1	FL20 (mg/ml) T0
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994) –re-runs based on activity	C 3/28 D 3/8 F 1/1	27.01.93 1/4 01.02.93 2/6 02.02.93 0/4 08.02.93 0/8 05.02.93 0/2 15.02.93 0/1 19.02.93 0/5 19.04.93 0/4 29.03.93 1/2 20.04.93 2/4	3	n≤ 5	3	a. 10 surfactants b. 0 c. 0	FL20 (mg/ml) T0
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994) –re-runs based on activity	C 1/28 D 0/8 F 0/1	27.01.93 0/4 01.02.93 1/6 02.02.93 0/4 08.02.93 0/8 05.02.93 0/2 15.02.93 0/1 19.02.93 0/5 19.04.93 0/4 29.03.93 1/2 20.04.93 2/4	3	n≤ 5	3	a. 10 surfactants b. 0 c. 4	FL20 (mg/ml) T72
Report on Comparison of 40 Cosmetic and Domestic Formulations Supplied by Company # 8 and Evaluated by 3 <i>In Vitro</i> Cytotoxicity Tests (FRAME, 1992)	A 1/5 B 0/9 F 4/64	03.04.92 0/2 10.04.92 0/2 01.05.92. 0/2 03.07.92. 0/1 05.05.92. 0/1 06.03.92. 0/2 07.01.92. 0/2 09.03.92. 0/2 11.05.92. 2/6 12.05.96 0/6 13.03.92. 0/3 15.05.92. 0/2 15.06.92 0/2 17.03.92. 0/4 19.06.92 0/1 21.05.92. 0/4 22.05.92.1/20 23.03.92. 2/4 23.06.92. 0/2 24.04.92. 0/2 27.01.92. 0/1 29.04.92 0/1 29.06.92 0/2 31.03.92. 0/4	3	4	3	a. 40 b. 12 c. 8 d. 6	FL20 (mg/ml) T0

Study	Operator Variability	Time Variability	Operator No.	Exp. No.	Well Replicate No.	No. of Test Substances a, b., c.,	Data Format
Company # 5 test concentration X	n/a	09.02.93. 0/9 16.02.93. 0/11 08.02.93. 0/4 03.02.93. 2/10 06.02.93. 0/9	1	n≤ 4	3	a. 12 b. 2 c. 6	FL20 (mg/ml) T24
Company # 5 test concentration XI	n/a	09.02.93. 2/9 16.02.93. 0/10 08.02.93. 0/3 03.02.93. 0/10 06.02.93. 0/9	1	n≤ 4	3	a. 12 b. 2 c. 3	FL20 (mg/ml) T24
Final Report on Testing of 12 Mild Surfactants supplied by Company # 5 for Cytotoxicity Testing at the FAL (FRAME, 1992)	n/a	n/a –all CVs were below threshold values	2	n≤ 4 for the 5 materials analysed	3	a. 5 b. 0 c.0	FL20, (mM) T0, T24, T48

a. total number of materials in the data set; b. number of test materials that had two or fewer experimental repeats, thus all test material data were removed from the analyses; therefore the final number of test materials analysed is 'No. of substances in data set' minus 'number of test materials that had ≤2 experimental repeats'; c. number of test materials which did not have a CV below the threshold values. Outliers for these test materials were not included in the analyses for operator and time variability. However, this information was included in the counts of total number of runs per date and total number of runs per operator, which enables one to observe if a particular operator and/or dates were consistently responsible for the high variation.

i. Operator Variability

In order to evaluate operator variability, the number of outlier experimental runs per operator was calculated from all the appropriate data sets featured in table 3.2.3.1. (table 3.2.3.2).

Table 3.2.3.2. Operator variability for all the data sets featured in table 3.2.3.1

Data set	Operator					
	A	B	C	D	E	F
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994). T0			3/28	3/8		1/1
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994). T72			1/28	0/8		0/1
Report on Comparison of 40 Cosmetic and Domestic Formulations Supplied by Company # 8 and Evaluated by 3 <i>In Vitro</i> Cytotoxicity Tests (FRAME, 1992)	1/5	0/9	4/64			
Total	1/5	0/9	8/120	3/16	0/0	1/2
%	20%	0%	6.7%	18.8%	n/a	50%
Data sets	1	1	3	1	n/a	2

For each operator, the total number of outlier experimental runs was shown in relation to the total number of runs recorded (table 3.2.3.2). To allow operator variability to be compared, the number of outliers produced by each operator was calculated and presented as a percentage of the total number of runs performed. The results show that operator variability ranged from 0% to 50%. It was difficult to confidently assess the relationship between the number of experiments performed and the level of operator variability due to the limited number of data sets and differing number of experiments in each. However, the data did not indicate a relationship between operator variability and the number of experiments performed, i.e. operator A had a variability of 20% based on 5 experimental runs whilst operator D had a variability of 18.8% based on 16 experimental runs.

The number of studies was also recorded to observe if data from a greater number of studies increased the overall operator variability calculated. There was no correlation between operator variability and number of studies, i.e. variability for operator A and B was calculated using data from one study and were 20% and 0% respectively.

From this limited data, the results showed that operators do vary in their abilities to attain reproducible FL assay results for this protocol. With the exception of operator F (only two experiments performed), the reproducibility ranged from 0-20%. This indicated that this FL assay protocol was reproducible as the maximum rate of outliers produced was approximately 20%. As different data sets were used to assess operator variability, it should be noted that some of the data sets would have contained test materials with properties causing them to be difficult to test using the FL assay, thus leading to variable results impacting on the assessment of operator variability. Where the same sets of

materials were tested by different operators, it did not follow that each operator tested exactly the same set of test materials.

ii. Time Variability

The effect of length of time between experimental runs

As the experimental runs were arranged in chronological order for each chemical, a quick observation of the data indicated that no particular run (i.e. the first or last) was consistently the outlier run (data not shown). Subsequently, one could state that variation in experimental runs was not due to the time they were conducted, i.e. it could have been surmised that over time, more experience with the protocol would be developed so this could have resulted in increased reproducibility over time. The data did not indicate this to be the case, for example, three experimental repeats performed one week or three weeks apart were just as likely to produce outlier results.

The effect of failed experimental runs

It was clear that some experimental runs 'fail' for the majority of chemicals tested on that particular date. The data sets with more than 10 test materials were analysed to determine the number of experimental runs that produced 50% or more outlier results for the materials tested on that particular date (table 3.2.3.3.).

Table 3.2.3.3. Impact of failed experimental runs on overall reproducibility.

Study	No. of dates which produced outliers	No of dates with $\geq 50\%$ outlier results	Did the dates with more than 50% outlier results account for more than 50% of the total number of outliers within the data set?
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994). T0	2	0	n/a
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994). Re-runs based on activity T0	4	1 (only 2 chemicals tested)	No- account for 1 outliers out of a total of 6
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994). Re-runs based on activity T72	1	0	n/a
Report on Comparison of 40 Cosmetic and Domestic Formulations Supplied by Company # 8 and Evaluated by 3 <i>In Vitro</i> Cytotoxicity Tests (FRAME, 1992)	3	1 (only 4 chemicals tested)	No- account for 2 outliers out of a total of 5

a. Did the dates with more than 50% outlier results account for more than 50% of the total outlier dates within the data set?

The analysis of the effect of failed runs was hampered by the lack of large available data sets that were required in order for numerous materials to have been tested per experimental run. There were only two experimental runs that produced $\geq 50\%$ outliers but only four materials were tested in these runs. In general, the number of materials tested in an experimental run was low due to the complexity and time required to carry out the FL assay protocol.

Overall it appeared that 'failed' experimental runs were not responsible for a significant proportion of the total number of outliers within a data set. The majority of outliers were produced for one or two materials within an experimental run containing a high number of chemicals. In conclusion, this analysis indicated that this FL assay protocol was generally reproducible as there were not a great proportion of outliers.

The effect of potential inherent chemical variation

To determine if specific materials were responsible for the majority of outliers within each data set, the number of materials for which there were two or more outliers was recorded (table 3.2.3.4.). For this analysis, materials had to have five or more experimental repeats in order to have sufficient occasions for two outliers to occur, therefore only a few materials from each study qualified for analysis.

Table 3.2.3.4. The number of materials for which two or more outliers were identified in relation to the total number of materials within the data set

Study	No. of chemicals with at least one outlier	No. of chemicals with two of more outliers	No. of chemicals with 2 outliers as a proportion of chemicals with outliers
The Evaluation of Pesticide Ingredients and Formulations <i>In Vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995) T0	0	0	0%
The Evaluation of Pesticide Ingredients and Formulations <i>In Vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995) T72	2	0	0%
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994). T0	3	0	0%
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994). Re-runs based on activity T0	4	2	50%
Use of <i>In Vitro</i> Methodology to Predict Irritancy Potential of Surfactants (Hubbard <i>et al.</i> , 1994). Re-runs based on activity T72	1	0	0%
Final Report on Testing of 12 Mild Surfactants supplied by Company # 5 for Cytotoxicity Testing at the FAL (FRAME, 1992)	0	0	0%

T=time point

Data for 10 materials were analysed and only two test materials produced two or more outliers. More materials require testing, but this data indicated that outliers tended to be random and were not due to potential inherent variation of the materials being tested.

In conclusion, the findings suggested that the effect of time on the reproducibility of the FL assay was random. More data is required to conclusively determine if and how time

could affect FL assay reproducibility. There are a number of inter-related factors that could be responsible for the random distribution of outliers identified in these analyses.

3.3. Additional studies where raw data are not available: attempt to combine the data using weight-of-evidence approaches

Table 3.3. Presents the results and relevant information for each study containing non-raw data.

Study	Variability Reported	No. of Test Substances	Type of Test Substances	No. of Laboratories	No. of Operators	No. of Experiments	No. of replicate (wells)	Data Format
N/A	-	-	-	-	-	-	-	-

The weight-of-evidence approach applied was developed by the authors of this BRD.

I. Protocols

A substantial amount of data was available to allow within-laboratory reproducibility for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120 to be evaluated. Fewer data were available for materials tested using INVITTOX Protocol No. 82 and INVITTOX Protocol No. 86. There were only provisional data for INVITTOX Protocol No. 82; INVITTOX Protocol No. 71 had been used to test formulations at two fixed concentrations only.

Overall, more data were available for materials that were tested using INVITTOX Protocol No. 71, but INVITTOX Protocol No. 120 had been used in more laboratories. Data analysed for both protocols covered similar FL₂₀ (mg/ml) ranges and similar CV ranges. Applying a weight-of-evidence approach, there was greater weighting for INVITTOX Protocol No. 120 as data had been collected from two multi-laboratory studies that used this protocol; the COLIPA study (Brantom *et al.*, 1997) and the ECVAM Prevalidation study (Southee, 1998). A PM was devised before the start of both these studies; both studies also undertook independent statistical analyses. Statistical analyses of the COLIPA study found that reproducibility was good as the predicted classifications for *in vivo* ocular irritation were similar from both laboratories. In comparison, much of the data evaluated here, generated by INVITTOX Protocol No. 71 were produced in-house and did not undergo extensive nor independent statistical analyses.

ii. Test materials

Similar numbers of test materials have been tested using INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120. The variety of test materials used to assess intra-laboratory variability differed according to each INVITTOX Protocol. Chemicals were predominately tested using INVITTOX Protocol No. 120 whilst more formulations were tested using INVITTOX Protocol No. 71. This was because INVITTOX Protocol No. 120 featured in a number of large-scale validation/evaluation studies whilst INVITTOX Protocol No. 71 had been predominately used by the FAL to test formulations provided by industry. There was some INVITTOX Protocol No. 86/ TEP assay data for formulations that had been tested in the CTFA study Phase III (Gettings *et al.*, 1996) and used in-house by Company # 3.

The ICCVAM/NICEATM BRDs for organotypic models for ocular irritation requested data from pure chemicals. The chemicals also had to cover the potency range and mechanisms that the organotypic model was designed to test. Applying the same approach here would cause the data for formulations to have less weighting than the data for chemicals, although it is acknowledged that it is important to determine the predictive capacity of the FL assay for formulations. Using the ICCVAM/NICEATM approach for determining the predictive capacity of organotypic models for ocular irritation, the results from the ECVAM Prevalidation study Phase III (Southee, 1998) would receive greater weighting because pure chemicals were tested. As industry is likely to use the FL assay to test formulations rather than pure chemicals it is also important to know that the FL assay is capable of testing both formulations and chemicals.

INVITTOX Protocol No. 120 also featured in the COLIPA study where it was used to test 33 test materials in two different laboratories. Surfactants and surfactant-based formulations that were soluble in HBSS were tested. The range of potency of the test

materials (according to the *in vivo* data) was predominately from non-irritants to moderate irritants. Overall, greater weighting was given to INVITTOX Protocol No. 120 data which were generated by testing both surfactants and surfactant-based formulations (Formulation Annex B) and had been used in a number of different laboratories.

iii. Classifications

From the collected literature, classification systems were present for the CTFA study Phase III (Gettings *et al.*, 1996), COLIPA study (Brantom *et al.*, 1997) and the ECVAM Prevalidation study Phase III (Southee, 1998). INVITTOX Protocol No. 120 featured in the COLIPA study and the ECVAM Prevalidation study whilst INVITTOX Protocol No. 86 featured in the CTFA study Phase III.

The weight-of-evidence approach to the classifications can be analysed using different criteria. The COLIPA and ECVAM Prevalidation study (Southee, 1998) receive good weighting as the classification PM was defined prior to the testing of the materials. However, the *in vivo* data and *in vivo* classification system used to define the *in vitro* classification system were not transparent as it was only written that the PM was developed using Company # 4 data for 43 surfactants and surfactant-based formulations. In comparison, the CTFA Study established post-hoc threshold values for the TEP assay data, which produced the most accurate classifications possible in respect to the FHSA irritancy classifications. Thus, the CTFA classification model, based on a recognised *in vivo* classification scheme, would receive greater weighting than the PM used in the ECVAM Prevalidation study and the COLIPA study. A greater number and a wider range of test materials were tested using the protocol and PM featured in the COLIPA study and the ECVAM Prevalidation study (INVITTOX Protocol No. 120) in comparison to the number tested in the CTFA study Phase III (INVITTOX Protocol No. 86). The COLIPA PM was also applied to data from a number of different laboratories which contrasts to the one laboratory featured in the CTFA study Phase III.

In conclusion, the data analysed in this section provide greater weighting for the reproducibility of INVITTOX Protocol No. 120 in comparison to the other FL assay INVITTOX Protocols. When variability data (CV values) were compiled for all of the INVITTOX Protocols, variability for INVITTOX Protocol No. 120 and INVITTOX Protocol No. 71 were comparable despite the data for INVITTOX Protocol No. 120 originating from a number of different laboratories. The within-laboratory variability was lower for INVITTOX Protocol No. 86 due to the data originating only from one laboratory and the single type of materials tested.

- | | |
|----------|-----------------------------------------------------------------------------------------------------------|
| Annex II | Raw data used in the statistical analyses for intra-laboratory variability |
| Annex A | CTFA Study Phase III formulation ingredients (from draft HET-CAM BRD: Appendix C2 (ICCVAM/NICEATM, 2004)) |
| Annex B | COLIPA Study test chemicals and formulations compositions (from COLIPA) |
| Annex C | Formulation compositions from Company # 3. |

4. Transferability (Module 3)

4.1. Brief description of study results on transferability and availability of Standardised Operating Procedures (SOPs)

According to 'A modular approach to the ECVAM principles on test validity' (Hartung *et al.*, 2004), the aim of assessing transferability is to determine how much training is necessary to be able to establish the test in a naïve laboratory and reproduce assay predictions. The Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Southee, 1998) is the only study which was known to have investigated protocol transfer of a FL assay protocol (see section 5.2.). Phase II of this study specifically investigated protocol transferability by testing a single protocol in four different laboratories with various levels of experience with the chosen protocol, and the FL assay in general. The Company # 7 laboratory was responsible for refining the test protocol and also participated in testing the materials. The FAL participated in the study and represented a laboratory experienced with the assay in general but not the specific test protocol. Company # 3 also participated in the study; they were familiar with a similar FL assay but their participation also allowed assay transferability to the USA to be determined. ECVAM participated in the study as a naïve laboratory was required to fully assess the completeness of the protocol. Five materials which covered a range of irritancy were tested. Variability was quantified by calculating the CV for the 0% and 100% leakage controls and the positive control (0.16mg/ml SLS) data. The CVs varied considerably between laboratories with the ECVAM laboratory generally producing the most variable results. The laboratories also differed as to which control produced the highest CV. All laboratories produced very few results that had control data within the various pre-defined acceptance ranges of the different control values. The results indicated that there were different levels of intra-laboratory variability in the laboratories and that the protocol did not transfer well. Further refinements to the protocol were made for Phase III of the study which investigated protocol performance, i.e. increasing the ranges of results for the 0% leakage control and the positive control, increased number of washing steps, clarification of protocol steps

For Phase III, ten mild surfactants were tested twice in each laboratory. Analyses included data that did not satisfy the acceptance criteria as all laboratories experienced problems to consistently fulfil the acceptance requirements. Inter-laboratory variability was lowest when the results from ECVAM and Company # 3 (Pearson's correlation coefficient 0.98) were compared and highest when the results from the FAL and Company # 3 (Pearson's correlation coefficient 0.82) were compared. Following the results of this study, this protocol was accepted as INVITTOX Protocol No. 120.

Different FL assay protocols have been tested in large-scale validation and/or evaluation studies. Some studies tested a single protocol in a number of independent laboratories which enabled inter-laboratory variability to be assessed. In the absence of specific studies for protocol transferability, inter-laboratory variability indicates protocol transferability to some extent. None of the studies used any of the standardised INVITTOX protocols, but as a result of these studies, some protocols were later adopted as INVITTOX protocols.

'The EC/HO International validation study on alternatives to the Draize eye irritation test (Balls *et al.*, 1995),' investigated a number of *in vitro* assays for their ability to predict *in*

vivo MMAS scores. It was stated in the publication of Balls *et al.*, (1995) that the FL assay protocol evaluated, was based on the method of Tchao (1988); the protocol featured in this study was later accepted as INVITTOX Protocol No. 71. This protocol was tested in four laboratories; sixty chemicals of high purity were tested. Pearson's correlation coefficients were calculated to compare the results from the different laboratories and values ranged from 0.167 to 0.778 (Balls *et al.*, 1995). All the low correlation co-efficients were obtained when comparisons were made with one particular laboratory which indicated that this laboratory did not follow the protocol in a similar manner to the other three. A general concern of this study was that the protocol was not equally adhered to by all participating laboratories (R Clothier, personal communication), and thus inter-laboratory variation was reflected in the Pearson correlation co-efficients. Certain participating laboratories had used a FL assay method before, and since this was one of the first large multi-laboratory studies it was not fully realised that protocol adherence should be checked during the generation of the data and not at the end of the study. The format of the data submitted by some laboratories showed clearly that the protocol had not been followed. Thus, whilst modern validation studies agree upon the test protocols and undergo training before the study is embarked upon, in the EC/HO study (Balls *et al.*, 1995) the protocol was not tested via initial training or via a training set of chemicals; these are now required for a study to be considered as eligible for a validation study.

'The COLIPA Eye irritation international validation program (Phase I)' (Brantom *et al.*, 1997; Zanvit *et al.*, 1999) was the only other published study known to have tested a single protocol in more than one laboratory. The FL assay protocol's predictive capacity for *in vivo* MMAS values was evaluated. The protocol was performed according to Cottin *et al.*, (1992). Only two laboratories (FAL and Company # 4) tested the protocol and therefore it was difficult to draw definitive conclusions regarding between-laboratory variability. Additionally, both laboratories were experienced with the FL assay protocol used or a similar version which entails that transferability to a naïve laboratory was not assessed. Classifications based on the *in vivo* data were compared to the FL assay predicted classifications using Kappa (k) analysis where a value of one indicated complete agreement between values. Although no direct analyses for inter-laboratory variability were performed, both laboratories were similar in the number of correctly predicted *in vivo* classifications; FRAME K= 0.65±0.3, Company # 4 K= 0.78±0.2 (Zanvit *et al.*, 1999). These results indicated that the protocol was performed similarly in both laboratories although it is known that some information is lost when qualitative classifications are considered rather than quantitative values. From these results, one could conclude that this protocol was successfully transferred to the different laboratories. However, it should be noted that of a range of test materials available for testing in the COLIPA study, the FL assay protocol was only used to test mild surfactants and surfactant-based formulations that were soluble in HBSS. Subsequently, the type of material tested was the same for which the FL assay was specifically designed to test and therefore the assay and PM would be expected to perform well. If other types of materials had been tested, the results from the different laboratories may have varied to greater extents and affected the conclusions regarding protocol transferability. A modified version of this protocol was accepted as INVITTOX Protocol No. 120 following the results of the ECVAM Prevalidation Study (Southee, 1998).

4.2. Facilities and major fixed equipment needed

The fixed equipment required to carry out the FL assay are generally those required for cell culture (table 4.2.)

Table 4.2. The fixed equipment required to carry out the FL assay

Equipment	Use
-laminar flow hood	cell culture only, as the FL assay does not need to be performed using sterile conditions
-incubator (37°C, 5% CO ₂ , 90% humidity)	cell culture and Na-fluorescein dye incubation
- cell counter or counting slide	cell culture and cell seeding into wells
-inverted phase microscope	cell counting and/or checking monolayer confluency
- spectrofluorimeter (485nm, 530nm filters)	reading Na-fluorescein dye OD
-balance	weighing test materials for solubilisation
-water bath	warming cell culture materials and solubilising test materials
-fridge	storing cell culture media
-freezer, liquid nitrogen container	cell storage

4.3. Required level of training, expertise, and demonstrated proficiency needed

The availability of standardised FL assay INVITTOX Protocols No. 71, No. 82, No. 86, No.120, make the FL assay straight-forward to perform as instructions are clear and precise. However, the INVITTOX protocols are rarely cited in the literature, unless by the research groups that devised them.

In general the FL assay is relatively easy to perform for anyone with basic cell culture experience. Both strains of the MDCK cells grow well in culture and on the inserts. In comparison to other *in vitro* assays, the principal difficulty of the FL assay protocol is that special care is needed when dealing with the monolayer grown on the insert membranes to ensure that neither are damaged. Rinsing steps following test material exposure and prior to adding the sodium-fluorescein dye are incorporated into all the INVITTOX Protocols. The rinsing steps are required to ensure that the test material does not remain in the insert and produce erroneous results.

The FL assay can be used to measure recovery after an initial exposure which entails that the cells can remain cultured on the inserts for up to 96 hours. During this time, multiple FL assays are performed on the same population of cells, thus providing many opportunities for damage to the monolayer and insert to occur. After the FL assay is completed, the remaining sodium-fluorescein dye is removed from the insert and the inserts placed into new 24 well plates containing fresh medium; fresh medium is then also added to the inserts. Of the four different INVITTOX Protocols, recovery is only featured in INVITTOX Protocol No. 82, although other INVITTOX Protocols have been adapted to measure recovery.

An understanding of solubility is important to ensure that the correct solvent is used to solubilise the test materials and that concentrations are uniform; this ensures exposures are comparable for subsequent experimental repeats. Studies using coded materials should provide adequate solubility instructions. Knowledge of the properties of test materials is required to ensure that they are compatible for testing with the FL assay, as

some chemicals can interact with the sodium-fluorescein dye. Sufficient controls, i.e. with and without cells should highlight these effects.

The FL assay was designed to measure the effects of the damage caused by test materials to a cell monolayer in order to be able to predict damaging effects to the corneal epithelium *in vivo*. The assay has been predominately used to measure the effects caused by materials relevant to the cosmetics industry, such as surfactants, which are likely to enter the eye through accidental exposures. Surfactants are often supplied as stock concentrations $\geq 30\%$ that are viscous gels, creams, or formulations that can be applied neat or diluted using volume/volume aqueous solvents. Viscous materials are not always easily removed and special care is required to ensure that the monolayers and/or insert membranes are not damaged during the removal process. Some experience in efficiently applying and removing viscous test materials after a short incubation are needed to enable experiments to be replicated and to ensure exposures are not longer than defined by the protocol. This problem is a greater concern for those protocols using one minute exposures, as any variation in exposure period will be a greater proportion of one minute in comparison to the five minutes used by other protocols. It is especially important to remove all the test material in the case of the FL assay, as any remaining in the insert could cause the membrane pores to be blocked and physically prevent FL. Dyes which alter the colour of the sodium-fluorescein dye could potentially distort results if left in the well, although a no-cell control exposed to the same dye concentration can indicate this as a problem. In general, careful observation of the insert and monolayer is required at all stages of the FL assay to ensure that the monolayer is not damaged and that the test material has been fully removed.

Training is also required to ensure that the operator is able to perform all the necessary steps quickly following the short test material exposure, i.e. aspiration, washings and fixation/desorbing of the remaining dye. This should be undertaken with the positive control that can be used to confirm reproducibility of results for the operator and ensure they are comparable to published data for the chosen protocol.

A level of specific technical instruction/assistance is normally required for setting up the fixed laboratory equipment e.g. the incubator and laminar flow hood, and for operating the plate reader.

5. Between-laboratory reproducibility (Module 4)

Various FL assay protocols have featured in a number of large-scale studies which assessed the ability of the protocols for predicting the effects of chemical-induced *in vivo* ocular irritation. Normally, the FL assay was tested in two or more laboratories which also allowed the results to be used to assess between-laboratory variability. Table 5.1. shows the studies where a single FL assay protocol was tested in more than one laboratory and raw *in vitro* data were available.

5.1. Table presenting the relevant results and information for studies where raw *in vitro* data are available.

Study	No. of Chemicals	No. of Products	Code	Results (e.g., ANOVA SD, CV, Range, sim of classif.)	Lab. No.	Exp. No.	Replicate No.	Data Format (raw, summary)	Chemical Classes	Ranges of Toxicity	Physico-chemical properties
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase II) (Southee, 1998)	4	1	Not stated	A hierarchical ANOVA performed for the Southee report (1998) noted inter-well, inter-plate, inter-experiment, inter-day and inter-laboratory variation.	4	3-8	3	Raw FL20 (mg/ml) T0, T4 values per chemical per laboratory	triton X-100 (5%), CTAB, Johnson's baby shampoo, glycerol, ammonium nitrate	MMAS Range: 0-44	Viscous liquids, hydrophilic emulsifiers, liquids, MW 80.04-364.45
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	10	0	Y	ANOVA analysis found statistical difference in results from the three laboratories for four chemicals	3	Protocol specified 2 (but n=2-3 FRAME; n=2-3 ECVA M; n=2 COMP ANY # 3)	3	Raw FL20 (mg/ml) T0, T4 values per chemical per laboratory	Mild surfactants relevant to cosmetic testing with <i>in vivo</i> data readily available from BIBRA (originally from AVON and Sigma).	MMAS Range: 0-34	Viscous liquids, liquids, powders, MW: 288.37687-414.6

T=time point, MMAS= modified maximal average score, MW= molecular weight

5.2. Discussions from the literature

The following comments were taken from the ECVAM Prevalidation Report (Southee, 1998) and the accompanying independent statistical analysis report prepared by BIBRA International (Lovell, 1998).

'The Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay' allowed between-laboratory variability in Phase II and Phase III of the study to be investigated (Southee, 1998). Phase II specifically investigated protocol transferability although protocol modifications (i.e. the acceptance criteria) were possibly based upon initial results. In Phase I of the prevalidation process, Company # 7 developed a FL assay protocol that was then transferred to the FAL laboratory. The FAL was experienced in performing a slightly different FL assay protocol. In addition to Company # 7 and the FAL, ECVAM participated in Phase II and had no experience of performing any FL assay protocol. The research laboratory of Company # 3 also participated in Phase II; this laboratory had devised the TEP assay (INVITTOX Protocol No. 86) and their participation in this study allowed problems associated with setting up the assay in the USA to be highlighted. Data generated by these three laboratories allowed many aspects of protocol transferability to be tested and between-laboratory variability to be assessed.

The five materials tested in Phase II were, 5% triton X-100, CTAB, glycerol, ammonium nitrate and a Company # 3 baby shampoo. All chemicals were distributed to the other test laboratories by Company # 7. It was understood that the Company # 3 baby shampoo formulation was obtained by the laboratories independently as it was reported that Company # 3 baby shampoo only was distributed by Company # 7 to Laboratory 4 (Company # 3) as the formulation was unique to the UK.' Under statistical guidance, the five test materials were tested as follows:

- 3 chemicals on 5 separate occasions
- 2 chemicals on 2 separate occasions

The FL₂₀ (mg/ml) values and the following control values were recorded immediately following the exposure (T0) and four hours later (T4); FL% for 0.16mg/ml SLS, FL% for the untreated monolayer (0% leakage control), FL% for the insert only (100% leakage control). For FL₂₀ (mg/ml) values to be accepted, the control values had to be within pre-defined ranges.

The materials were tested at different periods by the various laboratories. However, all laboratories conducted the testing within two month periods. The results were collected by Company # 7 and sent to BIBRA International (UK) for independent statistical analyses. Variability was quantified by calculating the CVs for the 0% and 100% FL controls, and the positive control (0.16mg/ml SLS). The CVs for these controls varied considerably between laboratories with the ECVAM laboratory generally producing the most variable results. The laboratories also varied as to which control produced the greatest mean CV value. All laboratories produced only a few results that had control data within the acceptance ranges. The positive control values from FAL were consistently below 15% and outside of the specified acceptance range of 15.1-28.3%. Some of the positive control results submitted by ECVAM were also outside of the acceptance range. Both laboratories felt confident that their FL₂₀ results were valid despite failing positive control results. The Company # 3 laboratory had more control

data within the acceptance range but some runs did fail which resulted in additional runs being performed until the acceptance criteria were fulfilled. As all laboratories had some problems to fulfil the acceptance criteria of the positive control, it was undecided by the study organisers whether it was a poor choice for a positive control and subsequently no test chemical data were excluded as protocol modifications were allowed at this stage of the study. These results suggested that there were disparate levels of intra-laboratory variability in the different laboratories, and that the protocol did not transfer well. It was concluded that there was inter-laboratory variation which was either due to protocol ambiguity or the protocol not being fully adhered to.

One possible source of inter-laboratory variation was the different concentration ranges used by the various laboratories to determine the FL₂₀ value. In addition, some laboratories changed the range of concentrations used for the different experimental runs to a greater extent than others.

Some laboratories performed a greater number of experimental repeats than was originally decided. It also appeared that some laboratories included the results from range-finding experiments, although these were not adequately distinguished from the results of confirmatory assays. The inclusion of range-finding experiments from some laboratories and not others was reported to be a major cause of between-laboratory variation. Some laboratories submitted greater-than or less-than values which were considered to be redundant in the statistical analyses and furthermore were 'not allowed' by the protocol (Lovell, 1998).

Whilst some laboratories performed control experiments on a per chemical basis, other performed positive controls per plate or per day. This impacted on the likelihood of FL₂₀ values being accepted, e.g., if a control value failed it could cause one chemical run to be excluded in one laboratory, a plate of results to be excluded in another, and a number of plates to be excluded in the other.

A hierarchical ANOVA performed for the ECVAM Prevalidation study report noted inter-well, inter-plate, inter-experiment, inter-day and inter-laboratory variation for the Phase II data (Southee, 1998).

Further refinements to the protocol were made before Phase III testing began, i.e. increasing the ranges of FL% results for the untreated monolayer (0% leakage) control and the positive control.

The aim of Phase III was to determine protocol predictivity; as the protocol was tested in three different laboratories, inter-laboratory variability could also be tested. Ten mild surfactants, coded by BIBRA International, were tested twice in each laboratory. Mild surfactants were tested as results from the COLIPA study indicated that the FL assay was particularly useful for predicting lower levels of ocular irritation and showed better predictivity for surfactants (Zanvit *et al.*, 1999).

Raw and log-transformed FL₂₀ (mg/ml) T4 ECVAM Prevalidation study Phase III data were analysed for inter-laboratory variability by ANOVA analyses. Significant between-laboratory variation was reported for four of the ten surfactants when both types of data were analysed. The between-laboratory variability for each chemical was quantified by defining the value below which the difference between test results from different

laboratories would be expected to lie with 95% probability. The lowest level of between-laboratory was recorded for deoxycholic acid, and the highest level for polysorbate-60.

Kappa analysis was performed to determine the agreement of predicted classifications between the laboratories. Linear kappa values ranged from 0.44-0.82 according to which laboratory data sets were compared.

Following this study, comments from the participating laboratories were submitted to Company # 7. Those factors which could have impacted on between-laboratory variability are given below.

The FAL highlighted the problem of viscous solutions that can be difficult to pipette accurately and to wash off at the end of the exposure. Materials which are solubilised in mineral oil can precipitate out in a heterogeneous manner and lead to variable results.

ECVAM suggested that as SLS, (positive control) is hygroscopic it needed to be handled carefully. It was also stated that it was unclear whether the positive control was to be made up fresh every time it was required. Although the results for the test materials were not discarded on the basis of the control results, it was clear from the ECVAM comments that some ambiguity in the protocol existed and that the SLS results could be a source of variation in the future. It was also stated that the solubilisation instructions for the test materials were sometimes unclear. CTAB was also reported to bind to the insert membrane at concentrations above 1mg/ml.

The Company # 3 laboratory suggested that SLS may not be a suitable positive control. They reported that when assays 'failed' due to the positive control, the test samples results were in agreement with other runs that had passed. They also stated that five concentrations were too few to determine an adequate dose-response curve.

All laboratories highlighted the disadvantage of using Millicell-HA filters as the monolayer cannot be observed. Damage caused to the monolayer due to the many protocol procedures can go undetected and is therefore a potential source of inter-laboratory variability.

5.3. Compilation of results

5.3.1. Statistical approach(es) used: description & rationale for the approach used to determine between-laboratory reproducibility

For each data set, the mean, SD and CV were calculated per test material and for the entire data sets.

For each phase of the ECVAM Prevalidation study, one-way ANOVAs were performed using the mean results for each test material to determine if the variation between the results from the different laboratories was equal. Mean results were used as the number of experimental repeats varied according to test material and laboratory.

Pearson correlation coefficients were calculated for the data sets generated by the various laboratories. This allowed all the data sets produced by the different laboratories featured in each study to be compared.

For this BRD, analyses were performed to determine the agreement of predicted classifications of ocular irritation between the participating laboratories for the Prevalidation study Phase III. The PM used to transform the FL assay values into predicted classifications for ocular irritation was the same as featured in the Prevalidation report (Southee, 1998).

5.3.2. Relevant results and information for each study

Table 5.3.2.1.a. Summarised ECVAM Prevalidation Phase II FL₂₀ (mg/ml) data (All raw data for Phase II, for each laboratory provided in Annex III on CD)

i) T0

chemical	FRAME	COMPANY # 7	ECVAM	COMPANY # 3	Mean	SD	CV (%)
10% CTAB	1.59	2.12	0.73	0.57	1.25	0.73	58.28
Company # 3 baby shampoo	8.50	14.31	11.91	11.61	11.58	2.38	20.58
Glycerol	387.90	952.77	532.22	750.00	655.72	247.73	37.78
Triton X-100	19.97	15.90	15.60	22.27	18.43	3.24	17.57
Ammonium Nitrate	156.07	243.63	147.22	155.68	175.65	45.51	25.91
				Mean	172.53	59.92	32.02

ii) T4

chemical	FRAME	COMPANY # 7	ECVAM	COMPANY # 3	Mean	SD	CV (%)
10% CTAB	1.41	1.30	0.57	0.42	0.92	0.50	54.19
Company # 3 baby shampoo	9.83	32.32	12.80	8.34	15.82	11.15	70.49
Glycerol	376.57	978.87	764.61	750.00	717.51	250.22	34.87
Triton X-100	19.47	7.14	11.91	6.56	11.27	5.97	52.95
Ammonium Nitrate	327.93	353.83	242.95	330.29	313.75	48.63	15.50
				Mean	211.86	63.29	45.60

Table 5.3.2.1.b. Summarised ECVAM Prevalidation Phase III FL₂₀ (mg/ml) data (All raw data for Phase III, for each laboratory provided in Annex III on CD)

i) T0

chemical	FRAME	ECVAM	COMPANY # 3	Mean	SD	CV (%)
Benzalkonium Chloride (1%)	26.38	18.97633	17.5035	20.95	4.76	22.70
Brij-35		200.04	199.82	199.93	0.16	0.08
Cetyl stearyl alcohol (50% in corn oil)	7.12	18.35567	10.9405	12.14	5.71	47.06
Deoxycholic Acid	0.53	1.028267	0.1975	0.59	0.42	71.44
Lauryl Sulfobetaine	2.96	2.743167	0.953	2.22	1.10	49.65
N-lauroyl sarcosine sodium salt	0.485	0.773333	0.569	0.61	0.15	24.35
Polysorbate-60	436.86	57.6685	50	181.51	221.17	121.85
Sodium C14-C16 olesulfonate	0.4	0.376333	0.329	0.37	0.04	9.81
Sodium Laureth Sulfate Pres 35%	2.595	8.41	2.028	4.34	3.53	81.31
Triton X-155	1.693333	7.9386	1.2285	3.62	3.75	103.51
			Mean	42.63	24.08	53.18

ii) T4

chemical	FRAME	ECVAM	COMPANY # 3	Mean	SD	CV (%)
Benzalkonium Chloride (1%)	20.415	15.65133	26.516	20.86	5.45	26.11
Brij-35	-	121.99	200	161.00	55.16	34.26
Cetyl stearyl alcohol (50% in corn oil)	2.17	11.84217	2.629	5.55	5.46	98.37
Deoxycholic Acid	0.48	2.0211	0.468	0.99	0.89	90.25
Lauryl Sulfobetaine	1.65	2.695967	1.8465	2.06	0.56	26.93
N-lauroyl sarcosine sodium salt	1.19	1.3456	0.6285	1.05	0.38	35.76
Polysorbate-60	399.31	75.32067	37.545	170.73	198.86	116.48
Sodium C14-C16 olesulfonate	0.43	0.415317	0.58	0.48	0.09	19.18
Sodium Laureth Sulfate Pres 35%	5.1	5.116367	4.4685	4.89	0.37	7.55
Triton X-155	0.933333	7.278467	0.9665	3.06	3.65	119.43
			Mean	37.07	27.09	57.43

ECVAM Prevalidation study Phase II (Southee, 1998)

There was no statistical difference between the data sets for the five materials tested in the four different laboratories when T0 or T4 data were analysed (one-way ANOVA). The Pearson correlation coefficients for T0 data ranged 0.98-1 and for T4 data ranged 0.87-1.

ECVAM Prevalidation study Phase III (Southee, 1998)

Ten mild surfactants were coded and distributed by BIBRA. They were tested blindly at least twice in the three participating laboratories. There was no statistical difference between the data sets produced by the three participating laboratories when either T0 or T4 data were analysed (one-way ANOVA). Pearson correlation coefficients ranged 0.94-1 for T0 data, and 0.82-0.98 for T4 data.

ECVAM Prevalidation Study –Positive control data

Positive control data were available for Phase II and Phase III of the ECVAM Prevalidation study, and were used here to assess between-laboratory variation (figure 5.3.2.1. and figure 5.3.2.2.).

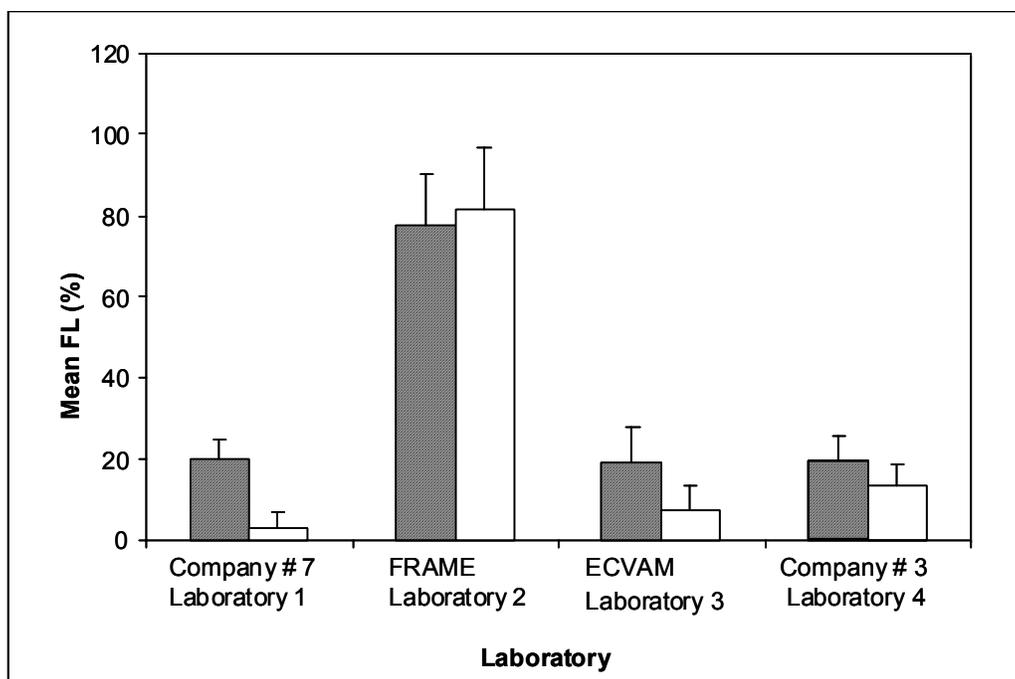


Figure 5.3.2.1. Mean FL% \pm SD values for Phase II positive control data (0.16mg/ml SLS) from each laboratory at T0 (pattern) and T4 (blank); $n \geq 5$. A version of INVITTOX Protocol No. 120 was used.

The T0 and T4 positive control data from the four test laboratories were statistically different (one-way ANOVA, $p < 0.0001$). This finding was based on different numbers of experimental runs from each laboratory. The mean values for the positive control data were similar for three of the four test laboratories at T0 and T4 (figure 5.3.2.1.). In comparison to the other laboratories, the FAL (FRAME Alternatives Laboratory) reported a greater amount of FL. At T0, the FAL had the smallest CV equal to 16%, whilst CV values for other laboratories ranged from 23% to 46%.

All laboratories had greater intra-laboratory variation for T4 data in comparison to T0 data, with CVs ranging from 19% (FAL) to 131% (Company # 7). One could expect greater variation at T4 because the relatively short chemical exposure would cause differing levels of cell damage. At T4 the monolayer begins to repair itself and the rate of the recovery would depend on the number of cells initially damaged and the extent of the damage, thus leading to high levels of variation in the amount of FL measured.

Using positive control data, the between-laboratory variation for the three laboratories that participated in Phase III was investigated; ECVAM, FRAME and Company # 3 (figure 5.3.2.2.). The main protocol differences between Phase II and Phase III were the acceptance criteria ranges for the control values which were increased slightly (0% leakage control, positive control) or the boundaries shifted (100% leakage control).

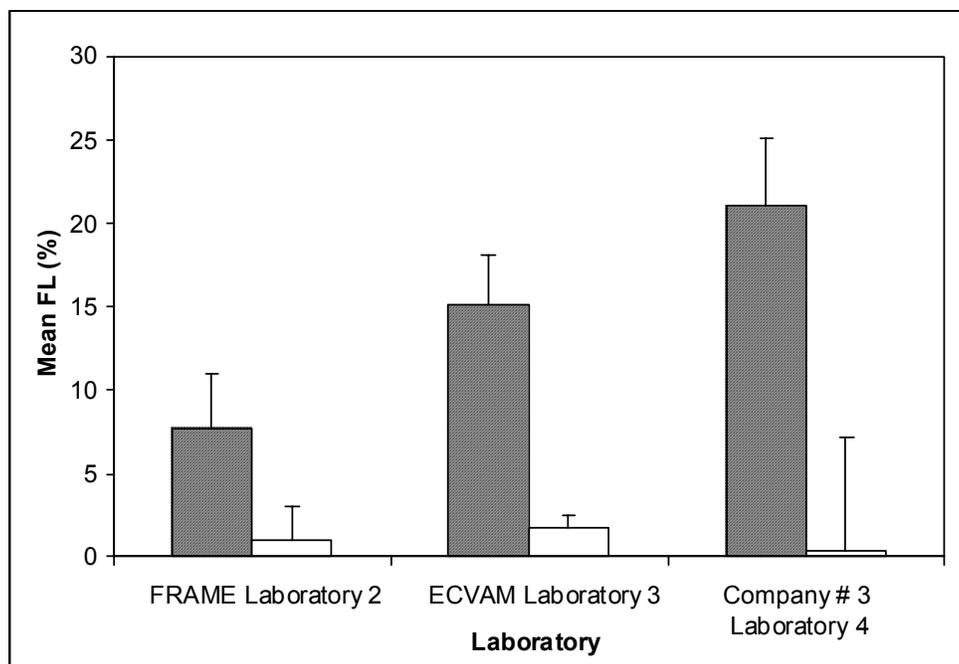


Figure 5.3.2.2. Mean FL% \pm SD values for Phase III positive control data (0.16mg/ml SLS) from each laboratory at T0 (pattern) and T4 (blank); $n \geq 5$. A version of INVITTOX Protocol No. 120 was used.

In comparison to Phase II data, all the laboratories showed relatively similar mean T0 FL assay values in Phase III. This finding was based on different amounts of data from each laboratory. There was no statistical difference between the positive control data from the different laboratories when either T0 or T4 data were compared. This indicated that protocol clarifications made after Phase II were effective as the protocol produced comparable Phase III results for the positive control between the laboratories. The SDs were similar for all T0 and T4 data with the exception of data from the Company # 3 laboratory at T4. For all laboratories, the difference between the mean values for T0 and T4 increased in each test laboratory in comparison to Phase II results.

Despite increased correlation between the Phase III T0 results from the different laboratories, only the Company # 3 laboratory had a mean value that fell well within the acceptance range of 15-30%. ECVAM had a mean value of 15% which was just within the range of acceptance, whilst the FAL had a mean value of 7%. All T4 mean values from the different laboratories failed to fall within the acceptance range. This result indicated that a different concentration of SLS may be required for it to be used as a suitable positive control, or that a different chemical would be more appropriate.

Phase III data were analysed to determine the proportion of identical predicted classifications from the three different test laboratories. The ten test surfactants covered a wide range of MMAS values. The PM applied to the data was taken from the COLIPA study (Brantom *et al.*, 1997) (table 5.3.2.2.).

Table 5.3.2.2. PM applied to ECVAM Prevalidation study Phase III data

FL20 (mg/ml) T4	Classification	Draize MMAS
>100	Non irritant/ Slight	<15
20-100	Moderate	15-30
<20	Irritant/ Severe	>30

Table 5.3.2.3. Mean FL₂₀ (mg/ml) T4 data and classification for each test material in each laboratory.

Chemical	FRAME	Classific.	ECVAM	Classific.	Company # 3	Classific.
Benzalkonium Chloride (1%)	20.42	Moderate	15.65	Irritant	26.52	Moderate
Brij-35	Not Tested	Not Tested	121.99	Non Irritant/ Slight	200	Non Irritant/ Slight
Cetyl stearyl alcohol (50% in corn oil)	2.17	Irritant	11.84	Irritant	2.63	Irritant
Deoxycholic Acid	0.48	Irritant	2.02	Irritant	0.47	Irritant
Lauryl Sulfobetaine	1.65	Irritant	2.70	Irritant	1.85	Irritant
N-lauroyl sarcosine sodium salt	1.19	Irritant	1.35	Irritant	0.63	Irritant
Polysorbate-60	399.31	Non Irritant/ Slight	75.32	Moderate	37.5	Moderate
Sodium C14-C16 olesulfonate	0.43	Irritant	0.42	Irritant	0.58	Irritant
Sodium Laureth Sulfate Pres 35%	5.1	Irritant	5.12	Irritant	4.47	Irritant
Triton X-155	0.93	Irritant	7.28	Irritant	0.97	Irritant

NB. N=3 for ECVAM laboratory and FAL, n=2 for the Company # 3 laboratory.

Table 5.3.2.4. Summary table of proportion of identical predicted classifications from each laboratory

Report	Classif. Scheme	No. of Lab.s	Materials with 100% agreement between laboratories	Materials with at least 1 divergent classif.	Materials with 3 different classif.s	All materials with classification differences compared *			Only materials which differed by 2 classifications compared**		
						FAL-ECVAM classif. comparison	FAL-Company # 3 classif. comparison	ECVAM-Company # 3 classif. comparison	FAL-ECVAM classif. comparison	FAL-Company # 3 classif. comparison	ECVAM-Company # 3 classif. comparison
ECVAM Prevalidation (Southee, 1998)	Refer to table 5.3.2.2.	3	77.8% (7/9)	22.2% (2/9)	0% (0/9)	77.8% (7/9)	88.9% (8/9)	90% (9/10)	100% (9/9)	100% (9/9)	100% (10/10)

Values in brackets are the numbers used to determine the percentage of identical predicted classifications from the total number of classifications.

* materials that differed by one or more classifications when comparing the different laboratories were counted

**only materials that produced classifications that differed by two classes were counted (i.e. non-irritant and irritant)

NB. Analyses which included data from the FAL were for nine test chemicals only as no value was produced for Brij-35

The results showed that the proportion of identical predicted classifications was relatively high and that predicted classifications for the surfactants only ever differed by one classification. One would expect the proportion of identical predicted classifications to be high as only surfactants were tested and the test materials had similar potencies, towards the mild end of the ocular irritation scale. In addition, the PM used to convert the FL assay results into the predicted classifications for ocular irritation was specifically developed for surfactants and surfactant-based formulations only. Therefore a high proportion of identical predicted classifications from the various laboratories would be expected as the protocol was not thoroughly challenged by a wide range of chemical classes and potencies.

5.4.1. Additional studies where raw data are not available: attempt to combine the data using weight-of-evidence approaches

Table 5.4.1. Relevant results and information for studies where raw data were not available

Study	No. of Chemicals	No. of Products	Coded?	Results	Lab. No.	Exp. No.	Replicate No.	Data Format (raw, summary)	Chemical Classes	Ranges	Physico-chemical properties
The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test (Balls <i>et al.</i> , 1995)	60	0	Y	Pearson correlation coefficients for inter-laboratory variability of FL20 values ranged from 0.214-0.841.	4	3	3	Mean FL20 (mg/ml) T0 values per chemical per laboratory	60 chemicals ranging in mechanisms and potency for which historical <i>in vivo</i> data were available- data primarily from ECETOC database	MMAS Range: 0-108	Liquids, crystals, powders MW: 39.99707-1228
A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom <i>et al.</i> , 1997)	11	18 (common to both FRAME and Company # 4 lab.)	Y	Kappa statistics for correlation of <i>in vivo</i> classifications and predicted classification; FAL K= 0.65± 0.3; Company # 4 K= 0.78± 0.2	2	4	3	Mean FL20 (mg/ml) T4 values	surfactants and surfactant-based materials soluble in HBSS	MMAS Range: 0-110	Solids, liquids MW 39.99707-1228

5.4.2. Discussions from the literature

Possible sources of variation acknowledged by the literature reporting these studies were discussed.

The EC/HO International validation study on alternatives to the Draize eye irritation test (Balls *et al.*, 1995)

The EC/HO International validation study on alternatives to the Draize eye irritation test (Balls *et al.*, 1995) investigated a number of *in vitro* assays for their ability to predict *in vivo* MMAS scores. The FL assay test protocol was based on the method of Tchao (1988) and was later accepted as INVITTOX Protocol No. 71. The FL assay was conducted in the following laboratories, FAL; Company # 4, State University of New York, the Philadelphia College of Pharmacy & Science. Inter-laboratory variability was assessed as part of this study.

Sixty chemicals of high purity were tested. The test chemicals were supplied by BIBRA International (UK), and when available came from the same source as those used for the previously conducted *in vivo* tests. Therefore, one can assume that where possible, the chemicals supplied to the different laboratories were from the same suppliers and that variation caused by different chemical batches was minimal. Balls *et al.*, (1995) stated that a wide range of chemicals with different mechanisms and potencies were used in order to fully challenge the various alternative methods tested. However, most alternative methods are developed and used in industry to test a specific chemical class or similar toxicity mechanisms, and thus predictive capacity would be expected to be low for a wide range of chemicals.

Pearson's correlation coefficients were calculated to compare the results from the different laboratories and ranged 0.214-0.841. (Balls *et al.*, 1995). All the low correlation co-efficients were obtained when comparisons were made with one particular laboratory. Although this finding could indicate that the protocol was not sufficiently detailed, therefore preventing all participating laboratories from following it correctly, as FAL participated in this study it was known that not all laboratories followed the FL assay test protocol equally (R Clothier, personal communication).

Various factors were stated in the Balls *et al.*, (1995) publication regarding the limitations of the FL assay method. These are discussed in detail in section 2. of this BRD. As these limitations would also impact on between-laboratory variability, they are also listed here briefly:

- viscous non-toxic materials can be difficult to remove after the one minute exposure
- binding of test chemicals to the insert membrane, (e.g. cationic surfactants) can occur
- transfer of the inserts to new wells can cause bubbles to form underneath the membranes and impede FL
- damage to the insert membrane can occur during the many handling procedures of the protocol
- some materials are incompatible with the O-ring of the membrane
- highly volatile materials need to be sealed with mineral oil to ensure continued contact with the cell monolayer
- materials need to be tested as, either an aqueous solution, mineral oil solution, suspension, or as a microemulsion formulation

The COLIPA Eye irritation international validation program (Phase I) (Brantom *et al.*, 1997; Zanvit *et al.*, 1999)

The COLIPA Eye irritation international validation program (Phase I) (Brantom *et al.*, 1997; Zanvit *et al.*, 1999) evaluated a FL assay protocol, for its predictivity of *in vivo* MMAS scores and inter-laboratory variability. The protocol was performed according to Cottin *et al.*, (1992), which was the basis for INVITTOX Protocol No. 120. The protocol was tested in only two laboratories (FAL and Company # 4) which made it difficult to draw definitive conclusions regarding inter-laboratory variability. Classifications based on the *in vivo* data were compared to the FL assay predicted classifications using kappa (κ) analysis where a value of one indicated complete agreement between the values. Although no direct analyses for inter-laboratory variability were reported in the publications, both laboratories reported a similar number of correctly predicted *in vivo* classifications; FRAME $K = 0.65 \pm 0.3$, Company # 4 $K = 0.78 \pm 0.2$ (Zanvit *et al.*, 1999). Four of the test materials tested by Company # 4 had their irritancy potential over-predicted which compared to the six over-estimated by FAL; four were common to both laboratories. None of the materials tested in Company # 4 laboratory had their irritancy potential under-estimated whilst the hand cleanser formulation was under-predicted when tested in the FAL. It was stated that this test sample contained a mineral spirit that was not very stable in the formulation; this could have caused the different results in the two laboratories.

Zanvit *et al.*, (1999) stated that there were too few test materials with mid-range irritancy potentials to enable the predictive capacity of the protocol to be fully tested. It was also stated that the protocol was not designed to measure materials with severe irritancy potentials (MMAS 50 to 110) and that another *in vitro* test should be performed before the FL assay, to detect severe materials which cannot be accurately measured using the FL assay. Forty materials were tested in this study although the results for only ~30 which were surfactants and surfactant-based formulations considered to be soluble in HBSS were accepted. The two laboratories differed as to which materials they considered to be soluble in HBSS.

It should be noted that of a range of test materials available for testing in the COLIPA study, the FL assay was only used to test surfactants and surfactant-based formulations that were soluble in HBSS. Subsequently, the range of materials tested was limited to the type which the FL assay was designed to test and therefore would be expected to perform well. If other types of materials had been tested, the results from the different laboratories may have varied to greater extents and indicated that the protocol was not very transferable. However, as the only available PM was specific for surfactants, comparisons of the predictive capacity of the protocol was limited to this chemical class.

5.4.3. Compilation of results

5.4.3.1. Statistical approach(es) used: description & rationale for the approach used to determine between-laboratory reproducibility

ANOVA analyses were not performed because there were no individual data for each experimental run per test material from each laboratory. Hence the results of an ANOVA analysis could be artifactual.

For each study, Pearson's correlation coefficients were calculated using the various data sets generated by the different laboratories. This allowed all the data sets produced by the laboratories featured in each study to be compared.

For the results from the COLIPA study, the mean CVs were calculated for the entire data set, and for 'formulations only' and for 'chemicals only.' The CVs were calculated for each laboratory.

For Prevalidation study Phase III data, analyses were performed to determine the agreement between the participating laboratories for predicted classifications of ocular irritation. The PM used to transform the FL assay values into predicted classifications of irritancy was the same as that featured in the COLIPA study.

5.4.3.2 Compilation of resultsTable 5.4.3.2.1. Between-laboratory FL₂₀ (mg/ml) data for COLIPA study test materials:

i) chemicals

Chemical	Lab 24	Lab 25	Mean	SD	CV (%)
Triton X-100 1%	58.98	51.05	55.02	5.61	10.19
SLS 3%	9.28	6.7	7.99	1.82	22.83
Triton X-100 5%	13.27	13.1	13.19	0.12	0.91
Benzalkonium chloride 1%	10.88	6.8	8.84	2.88	32.64
SLS 15%	4.36	4.3	4.33	0.04	0.98
SLS 30%	0.88	0.42	0.65	0.33	50.04
Triton X-100 10%	5.16	7.4	6.28	1.58	25.22
Benzalkonium chloride 5%	4.12	3.8	3.96	0.23	5.71
Benzalkonium chloride 10%	3.63	0.46	2.05	2.24	109.61
Cetylpyridinium bromide 6%	12.83	0.35	6.59	8.82	133.91
Cetylpyridinium bromide 10%	7.45	0.38	3.92	5	127.69
		Mean	10.26	2.61	47.25

ii) formulations

Formulation	Lab 24	Lab 25	Mean	SD	CV (%)
Perfumed skin lotion	930	>1062	n/a	n/a	n/a
Polishing scrub	649.85	311	480.43	239.6	49.87
Shampoo #1 normal	5.29	5.9	5.6	0.43	7.71
Eye make-up remover	193.12	161.5	177.31	22.36	12.61
Hand cleaner	10.3	24	17.15	9.69	56.49
Hair dye base F#1	992	>872.5	n/a	n/a	n/a
Sunscreen lotion	979	>1009	n/a	n/a	n/a
Emulsion antiperspirant	1057	>1062.5	n/a	n/a	n/a
Gel cleaner	35.38	17.75	26.57	12.47	46.93
Hand soap	4.86	-	4.86	n/a	n/a
Shampoo - baby	7.41	3.6	5.51	2.69	48.94
Sunscreen SPF 15	344.6	-	344.6	n/a	n/a
Liquid soap #1	5.82	6.15	5.99	0.23	3.9
Shampoo antidandruff	3.13	2.5	2.82	0.45	15.83
Shampoo 2-in-1	4.59	4.15	4.37	0.31	7.12
Mouthwash	244.25	44.95	144.6	140.93	97.46
Toothpaste	-	30.1	30.1	n/a	n/a
Cleansing foam III	9.67	-	9.67	n/a	n/a
Moisturiser with sunscreen	498.35	-	498.35	n/a	n/a
Shower gel	7.16	6.2	6.68	0.68	10.16
Skin cleaner	5.92	3	4.46	2.06	46.29
Hair dye base form #2	981	>1001	n/a	n/a	n/a
Hair dye base form #3	974	>750	n/a	n/a	n/a
		Mean	104.06	35.99	33.61

Table 5.4.3.2.2. Between-laboratory FL₂₀ (mg/ml) data for EC/HO study test chemicals

Chemical	Lab 18	Lab 19	Lab 20	Lab 21	Mean	SD	CV (%)
sodium hydroxide (10%)	6.3	13.3	5.8	10.0	8.9	3.5	39.6
benzalkonium chloride (10%)	19.0	>25.0	>1000.0	>25.0	n/a	n/a	n/a
trichloroacetic acid (30%)	120.0	>250.0	>1000.0	>25.0	n/a	n/a	n/a
cetylpyridinium bromide (10%)	27.0	>25.0	>960.0	>25.0	n/a	n/a	n/a
cetylpyridinium bromide (6%)	93.0	>25.0	>810.0	>100.0	n/a	n/a	n/a
benzalkonium chloride (5%)	18.0	>25.0	>1000.0	>25.0	n/a	n/a	n/a
captan 90 concentrate	>100.0	>250.0	>250.0	>100.0	n/a	n/a	n/a
chlorhexidine	>100.0	>250.0	>250.0	>100.0	n/a	n/a	n/a
cyclohexanol	473.0	>938.0	741.0	>1000.0	n/a	n/a	n/a
quiniacrine	>750.0	>250.0	>250.0	>250.0	n/a	n/a	n/a
promethazine HCl	12.0	53.0	65.1	>25.0	n/a	n/a	n/a
parafluoraniline	55.0	>1146.0	>1120.0	-	n/a	n/a	n/a
triton X-100 (10%)	90.0	99.4	339.0	665.0	298.4	270.2	90.6
acetone	523.0	709.3	678.0	839.0	687.3	129.8	18.9
hexanol	270.0	12.7	>750.0	>1000.0	n/a	n/a	n/a
1-naphthalene acetic acid, Na salt	171.0	>250.0	>500.0	245.0	n/a	n/a	n/a
sodium oxalate	>100.0	>250.0	>750.0	>250.0	n/a	n/a	n/a
isobutanol	369.0	>250.0	>770.0	>1000.0	n/a	n/a	n/a
imidazole	135.0	86.3	185.0	185.0	147.8	47.3	32.0
sodium lauryl sulphate (15%)	12.0	8.9	24.0	70.0	28.7	28.3	98.5
2-ethyl-1-hexanol	230.0	78.4	>730.0	>1000.0	n/a	n/a	n/a
4-carboxybenzaldehyde	>100.0	>250.0	>250.0	>100.0	n/a	n/a	n/a
methyl ethyl ketone	256.0	273.3	235.0	>1000.0	n/a	n/a	n/a
pyridine	176.0	315.8	371.0	989.0	463.0	360.2	77.8
1-naphthalene acetic acid	>100.0	>250.0	>250.0	>250.0	n/a	n/a	n/a
benzalkonium chloride (1%)	91.0	>100.0	>960.0	>250.0	n/a	n/a	n/a
2,2-dimethylbutanoic acid	172.0	>250.0	>810.0	>100.0	n/a	n/a	n/a
gamma-butyrolactone	133.0	160.7	482.0	152.0	231.9	167.1	72.1
thiourea	>100.0	>250.0	>250.0	>250.0	n/a	n/a	n/a
octanol	198.5	82.4	>770.0	>1000.0	n/a	n/a	n/a
methyl acetate	361.0	518.8	>870.0	>1000.0	n/a	n/a	n/a
L-aspartic acid	16.0	>250.0	>250.0	>100.0	n/a	n/a	n/a
benzoyl-L-tartaric acid	49.5	>25.0	>250.0	>100.0	n/a	n/a	n/a
triton X-100 (5%)	167.0	163.7	625.0	715.0	417.7	293.7	70.3
potassium cyanate	>750.0	>500.0	>500.0	>750.0	n/a	n/a	n/a
isopropanol	549.0	715.0	618.0	992.0	718.5	194.6	27.1
sodium perborate	50.5	>250.0	>250.0	>250.0	n/a	n/a	n/a
dibenzyl phosphate	21.0	>250.0	>250.0	>25.0	n/a	n/a	n/a
2,5-dimethylhexanediol	>750.0	>750.0	>250.0	>750.0	n/a	n/a	n/a
methyl cyanoacetate	>1100.0	>1115.0	>750.0	>100.0	n/a	n/a	n/a
sodium hydroxide (1%)	132.0	125.2	127.0	133.0	129.3	3.8	2.9
ethanol	516.0	698.6	590.0	>1000.0	n/a	n/a	n/a
2,6-dichlorobenzoyl chloride	>1500.0	>1448.0	>1400.0	>250.0	n/a	n/a	n/a
ammonium nitrate	551.0	>750.0	>750.0	>750.0	n/a	n/a	n/a

Table 5.4.3.2.2. continued

Chemical	Lab 18	Lab 19	Lab 20	Lab 21	Mean	SD	CV (%)
ethyl-2-methylacetoacetate	-	>990.0	369.0	>1000.0	n/a	n/a	n/a
sodium lauryl sulfate (3 %)	25.0	14.7	73.0	317.0	107.4	142.0	132.2
ethyl acetate	231.0	>887.0	746.0	>1000.0	n/a	n/a	n/a
maneb	>100.0	>250.0	>750.0	>250.0	n/a	n/a	n/a
fomesafen	>100.0	>250.0	>750.0	>250.0	n/a	n/a	n/a
tetraaminopyrimidine sulphate	>100.0	>250.0	>250.0	>100.0	n/a	n/a	n/a
toluene	>860.0	>800.0	>830.0	>1000.0	n/a	n/a	n/a
butyl acetate	>940.0	>874.0	>860.0	>100.0	n/a	n/a	n/a
trichloroacetic acid (3%)	803.0	>1006.0	>1000.0	>250.0	n/a	n/a	n/a
methyl isobutyl ketone	>800.0	>792.0	>770.0	>1000.0	n/a	n/a	n/a
tween 20	653.0	>1101.0	>750.0	>1000.0	n/a	n/a	n/a
ethyl trimethyl acetate	>850.0	>844.0	>800.0	>1000.0	n/a	n/a	n/a
methylcyclopentane	>750.0	>741.0	>670.0	>1000.0	n/a	n/a	n/a
cetylpyridinium bromide (0.1%)	>900.0	>984.0	>960.0	>1000.0	n/a	n/a	n/a
glycerol	>1200.0	>1254.0	>1000.0	>1000.0	n/a	n/a	n/a
polyethylene glycol 400	>1200.0	>1110.0	>1100.0	>1.0	n/a	n/a	n/a
				Mean	294.4	149.1	60.2

NB. It was known that the various laboratories differed as to the highest test concentration used per chemical. Therefore, the greater-than values from the different laboratories were not indicative of the level of toxicity of the chemical but the testing strategies adopted by the different laboratories. The overall mean, SD and CV calculated for the EC/HO study was based only on the actual data and not greater-than values. Subsequently, the overall mean, SD and CV values calculated are skewed as there were very few chemicals for which actual data were available.

The EC/HO International validation study on alternatives to the Draize eye irritation test (Balls *et al.*, 1995)

An inter-laboratory correlation of the FL₂₀ (mg/ml) data from each laboratory was reported in the EC/HO study publication (Balls *et al.*, 1995) and agreed with the Pearson's correlation coefficients performed for this BRD (table 5.4.3.2.3.).

Table 5.4.3.2.3. Inter-laboratory correlation of the FL₂₀ (mg/ml) data from the four laboratories (from Balls *et al.*, 1995)

	Laboratory			
	Lab 18	Lab 19	Lab 20	Lab 21
Lab 18	1			
Lab 19	0.841	1		
Lab 20	0.490	0.513	1	
Lab 21	0.373	0.424	0.214	1

The proportion of identical predicted classifications for the EC/HO chemicals tested in the four participating laboratories was calculated (table 5.4.3.2.5.). There was no PM featured in the EC/HO study that could be used with the FL assay data generated as the PM proposed in the publication required 72h FL assay recovery data to distinguish EU R36 and R41 classifications. The PM was modified by the authors of this BRD and

applied to the EC/HO FL₂₀ (mg/ml) data to distinguish EU irritants (R36/R41) and Not Classified classifications only (table 5.4.3.2.4.).

Table 5.4.3.2.4. Modified PM from EC/HO study used to assign irritant and non-irritant EU classifications.

FL₂₀ (mg/ml)	EU Classification
>750mg/ml	Not Classified
≤ 750mg/ml	R36/R41

The modified PM was applied to the mean results for each chemical from each laboratory. SD or SEM values were not available (table 5.4.3.2.5.).

Table 5.4.3.2.5. Predicted classifications for EC/HO study data generated using the PM featured in table 5.4.3.2.4.

Chemical Name	Chemical concentrations (mg/ml) causing FL 20%							
	Lab 18	Classif.	Lab19	Classif.	Lab 20	Classif.	Lab 21	Classif.
1-naphthalene acetic acid	>100	unknown	>250	unknown	>250	unknown	>250	unknown
1-naphthalene acetic acid, Na salt	171	R36/R41.	>250	unknown	>500	unknown	245	R36/R41.
2,2-dimethylbutanoic acid	172	R36/R41.	>250	unknown	810	NC	>100	unknown
2,5-dimethylhexanediol	>750	NC.	>750	NC	>250	unknown	>750	NC
2,6-dichlorobenzoyl chloride	>1500	NC	>1448	NC	>1400	NC	>250	unknown
2-ethyl-1-hexanol	230	R36/R41.	78.4	R36/R41.	>730	unknown	>1000	NC
4-carboxybenzaldehyde	>100	unknown	>250	unknown	>250	unknown	>100	unknown
acetone	523	R36/R41.	709	R36/R41.	678	R36/R41.	839	NC
ammonium nitrate	551	R36/R41.	>750	NC.	>750	NC	>750	NC
benzalkonium chloride (1 %)	91	R36/R41.	>100	unknown	>960	NC.	>250	unknown
benzalkonium chloride (10%)	19	R36/R41.	>25	unknown	>1000	NC.	>25	unknown
benzalkonium chloride (5%)	18	R36/R41.	>25	unknown	>1000	NC.	>25	unknown
benzoyl-L-tartaric acid	49.5	R36/R41.	>25	unknown	>250	unknown	>100	unknown
butyl acetate	>940	NC.	>874	NC.	>860	NC.	>100	unknown
captan 90 concentrate	>100	unknown	>250	unknown	>250	unknown	>100	unknown
cetylpyridinium bromide (0.1%)	>900	NC	>984	NC	>960	NC	>1000	NC
cetylpyridinium bromide (10%)	27	R36/R41	>25	unknown	>960	unknown	>25	unknown
cetylpyridinium bromide (6%)	93	R36/R41	>25	unknown	>810	NC	>100	unknown
chlorhexidine	>100	unknown	>250	unknown	>250	unknown	>100	unknown
cyclohexanol	473	R36/R41	>938	NC	741	R36/R41	>1000	NC
dibenzyl phosphate	21	R36/R41	>250	unknown	>250	unknown	>25	unknown
ethanol	516	R36/R41	699	R36/R41	590	R36/R41	>1000	NC
ethyl acetate	231	R36/R41	>887	NC	746	R36/R41	>1000	NC
ethyl trimethyl acetate	>850	NC	>844	NC	>800	NC	>1000	NC
ethyl-2-methylacetoacetate	*	N/A	>990	NC	369	R36/R41	>1000	NC

Chemical Name	Chemical concentrations (mg/ml) causing FL 20%							
	Lab 18	Classif.	Lab19	Classif.	Lab 20	Classif.	Lab 21	Classif.
Fomesafen	>100	unknown	>250	unknown	>750	NC	>250	R36/R41.
gammabutyrolactone	133	R36/R41	161	R36/R41	482	R36/R41	152	R36/R41
glycerol	>1200	NC	>1254	NC	>1000	NC	>1000	NC
hexanol	270	R36/R41	12.7	R36/R41	>750	NC	>1000	NC
imidazole	135	R36/R41	86.3	R36/R41	185	R36/R41	185	R36/R41
isobutanol	369	R36/R41	>250	unknown	>770	NC	>1000	NC
isopropanol	549	R36/R41	715	R36/R41	618	R36/R41	992	NC
L-aspartic acid	16	R36/R41	>250	unknown	>250	unknown	>100	unknown
maneb	>100	unknown	>250	unknown	>750	NC	>250	unknown
methyl acetate	361	R36/R41	519	R36/R41	>870	NC	>1000	NC
methyl cyanoacetate	>1100	NC	>1115	NC	>750	NC	>100	unknown
methyl ethyl ketone	256	R36/R41	273	R36/R41	235	R36/R41	>1000	NC
methyl isobutyl ketone	>800	NC	>792	NC	>770	NC	>1000	NC
methylcyclopentane	>750	NC	>741	unknown	>670	unknown	>1000	NC
octanol	198.5	R36/R41	82.4	R36/R41	>770	NC	>1000	NC
parafluoraniline	55	R36/R41	>1146	NC	>1120	NC	*	N/A
polyethylene glycol 400	>1200	NC	>1110	NC	>1100	NC	>1	unknown
potassium cyanate	>750	NC	>500	unknown	>500	unknown	>750	NC
promethazine HCl	12	R36/R41	53	R36/R41	65.1	R36/R41	>25	unknown
pyridine	176	R36/R41	316	R36/R41	371	R36/R41	989	NC
quiniacrine	>750	NC	>250	unknown	>250	unknown	>250	unknown
sodium hydroxide (1%)	132	R36/R41	125	R36/R41	127	R36/R41	133	R36/R41
sodium hydroxide (10%)	6.3	R36/R41	13.3	R36/R41	5.8	R36/R41	10	R36/R41
sodium lauryl sulfate (3 %)	25	R36/R41	14.7	R36/R41	73	R36/R41	317	R36/R41
sodium lauryl sulphate (15 %)	12	R36/R41	8.87	R36/R41	24	R36/R41	70	R36/R41
sodium oxalate	>100	unknown	>250	unknown	>750	NC	>250	unknown
sodium perborate	50.5	R36/R41	>250	unknown	>250	unknown	>250	unknown
tetraaminopyrimidine sulphate	>100	unknown	>250	unknown	>250	unknown	>100	unknown
thiourea	>100	unknown	>250	unknown	>250	unknown	>250	unknown
toluene	>860	NC	>800	NC	>830	NC	>1000	NC
trichloroacetic acid (3%)	803	NC	>1006	NC	>1000	NC	>250	unknown
trichloroacetic acid (30%)	120	R36/R41	>250	unknown	>1000	NC	>25	unknown
triton X-100 (10 %)	90	R36/R41	99.4	R36/R41	339	R36/R41	665	R36/R41
triton X-100 (5 %)	167	R36/R41	164	R36/R41	625	R36/R41	715	R36/R41
tween 20	653	R36/R41	>1101	NC	>750	NC	>1000	NC

* no results reported; NC =EU Not Classified, GHS No Category

The proportion of identical predicted classifications from each laboratory (table 5.4.3.2.5.) was assessed and the results summarised in table 5.4.3.2.6. Only the results for chemicals which had results leading to four classifications based on the prediction model were used (i.e. the classifications for any chemical which had one or more laboratories producing an unknown classification were not used). It was considered inappropriate to utilise 'unknown' as a classification as it was known that the various test laboratories used different criteria to determine which was the highest test concentration used. Therefore the analyses were performed using the classifications for only 25 chemicals.

Table 5.4.3.2.6. Summary table of proportion of identical predicted classifications from each laboratory

Report	Classif. Scheme	No. of Lab.'s	Materials with 100% agreement between laboratories	Materials with at least 1 divergent classif. from any laboratory	Materials with 3 different classif.s	All materials with classification differences compared					
						Lab 18-Lab 19 classif comparison	Lab 18-Lab 20 classif comparison	Lab 18-Lab 21 classif comparison	Lab 19-Lab 20 classif comparison	Lab 19-Lab 21 classif comparison	Lab 20-Lab 21 classif comparison
EC/HO Study (Balls <i>et al.</i> , 1995)	Refer table 5.4.3.2.4.	4	52% (13/25)	48% (12/25)	N/A	84% (21/25)	80% (20/25)	52% (13/25)	80% (20/25)	68% (17/25)	72% (18/25)

Similar analyses performed for other datasets (i.e., Prevalidation study Phase III, COLIPA study data) calculated the proportion of chemicals which produced more than one class difference in the predicted classifications, i.e the number of chemicals that were predicted to be Not Classified and R41 irritants. As the PM for the EC/HO dataset was unable to distinguish R36 and R41 irritants, this analysis could not be performed for this dataset.

Only approximately 50% of the test materials produced the same classification in all four participating laboratories. The proportion of identical classifications from the laboratories was lowest when the classifications from all other laboratories were compared with laboratory 21; excluding analyses for laboratory 21, the proportion of identical classifications for the other laboratories was consistently above 80%.

The COLIPA Eye irritation international validation program (Phase I) (Brantom *et al.*, 1997; Zanvit *et al.*, 1999)

A one-way ANOVA compared the FL₂₀ values produced by Company # 4 laboratory and the FAL for materials found to be compatible for testing with this assay (33 test materials for Company # 4, 30 test materials for the FAL). Interestingly, the two participating laboratories did not agree totally about which materials they found to be soluble. Twenty-nine identical materials were tested in both laboratories. There was no statistical difference between the data sets, subsequently the Pearson correlation coefficient was also found to be high; 0.98. The mean CV for the 'chemicals only' analysis was calculated to be 43% whilst the mean CV for the 'formulations only' analysis was 25%. These results indicate relatively good inter-laboratory variability although it is acknowledged that the test materials were those specifically suited for testing with the FL assay.

The number of experimental repeats performed and the acceptance criteria were not stated.

Predicted classifications for the COLIPA data were generated using the PM as featured in Brantom *et al.*, (1997) (table 5.3.2.2.). Table 5.4.3.2.7. shows the predicted classifications for both laboratories. Data were provided by COLIPA.

Table 5.4.3.2.7. Predicted classifications for COLIPA Study data generated using the PM featured in table 5.3.2.2.

Test Chemical	Chemical concentrations causing FL20% (mg/ml) T4			
	Lab 24 Company # 4	Lab 24 Classif.	Lab 25 FRAME	Lab 25 Classif.
Benzalkonium chloride 1%	10.9	Irritant/ Severe	6.8	Irritant/ Severe
Benzalkonium chloride 10%	3.6	Irritant/ Severe	0.5	Irritant/ Severe
Benzalkonium chloride 5%	4.1	Irritant/ Severe	3.8	Irritant/ Severe
Cetylpyridinium bromide 10%	7.5	Irritant/ Severe	0.4	Irritant/ Severe
Cetylpyridinium bromide 6%	12.8	Irritant/ Severe	0.4	Irritant/ Severe
Cleansing foam III	9.7	Irritant/ Severe	*	*
Emulsion antiperspirant	1057.0	NI/Slight	>1062.50	NI/Slight
Eye make-up remover	193.1	NI/Slight	161.5	NI/Slight
Gel cleaner	35.4	Moderate	17.8	Irritant/ Severe
Hair dye base F#1	992.0	NI/Slight	>872.5	NI/Slight
Hair dye base F#2	981.0	NI/Slight	>1001.00	NI/Slight

	Chemical concentrations causing FL20% (mg/ml) T4			
Test Chemical	Lab 24 Company # 4	Test Chemical	Lab 24 Company # 4	Test Chemical
Hair dye base F#3	974.0	NI/Slight	>750.00	NI/Slight
Hand cleaner	10.3	Irritant/ Severe	24.0	Moderate
Hand soap	4.9	Irritant/ Severe	*	*
Liquid soap #1	5.8	Irritant/ Severe	6.2	Irritant/ Severe
Moisturiser with sunscreen	498.4	NI/Slight	*	*
Mouthwash	244.3	NI/Slight	45.0	Moderate
Perfumed skin lotion	930.0	NI/Slight	>1062	NI/Slight
Polishing scrub	649.9	NI/Slight	311.0	NI/Slight
Shampoo - baby	7.4	Irritant/ Severe	3.6	Irritant/ Severe
Shampoo #1 normal	5.3	Irritant/ Severe	5.9	Irritant/ Severe
Shampoo 2-in-1	4.6	Irritant/ Severe	4.2	Irritant/ Severe
Shampoo antidandruff	3.1	Irritant/ Severe	2.5	Irritant/ Severe
Shower gel	7.2	Irritant/ Severe	6.2	Irritant/ Severe
Skin cleanser	5.9	Irritant/ Severe	3.0	Irritant/ Severe
SLS 15%	4.4	Irritant/ Severe	4.3	Irritant/ Severe
SLS 3%	9.3	Irritant/ Severe	6.7	Irritant/ Severe
SLS 30%	0.9	Irritant/ Severe	0.4	Irritant/ Severe
Sunscreen lotion	979.0	NI/Slight	>1009.00	NI/Slight
Sunscreen SPF 15	344.6	NI/Slight	*	*
Toothpaste	*	*	30.1	Moderate
Triton X-100 1%	59.0	Moderate	51.1	Moderate
Triton X-100 10%	5.2	Irritant/ Severe	7.4	Irritant/ Severe
Triton X-100 5%	13.3	Irritant/ Severe	13.1	Irritant/ Severe

* Material not tested by the laboratory; NI =non-irritant

Analyses were performed to determine the proportion of identical predicted classifications from each laboratory (table 5.4.3.2.8.).

Table 5.4.3.2.8. Summary table of proportion of identical predicted classifications from each laboratory

Report	Classif. Scheme	No. of Lab.'s	Materials with 100% agreement between laboratories	Materials with at least 1 divergent classification*	Only materials which differed by 2 classifications compared**
COLIPA (Brantom <i>et al.</i> , 1997)	Refer to table 5.3.2.2.	2	89.7% (26/29)	10.3% (3/29)	0% (0/29)

Values in brackets are the numbers used to determine the percentage of proportion of identical classifications from the different laboratories.

* materials that differed by one or more classifications when comparing the different laboratories were counted

**only materials that produced classifications that differed by two classes were counted (i.e. non-irritant and irritant)

NB. The analyses were limited due to only two participating laboratories.

The proportion of identical predicted classification from the two laboratories was high. There were no test materials that caused the classifications from the two laboratories to differ by two classes (i.e. non-irritant and severe irritant). This was probably due to the laboratories only testing surfactants and surfactant-based formulations that they considered to be soluble in HBSS. Unknown solubility of test materials is likely to account for some between-laboratory variability as laboratories are likely to use different solvents or obtain different solubilised concentrations.

There were nine test chemicals that featured both in the EC/HO study (Balls *et al.*, 1995) and the COLIPA study (Brantom *et al.*, 1997). Different protocols were used in these studies. The FAL and Company # 4 participated in both studies. The results for these nine test chemicals from both studies and for both laboratories are shown (table 5.4.3.2.9.) (data from Zanvit *et al.*, 1999).

Table 5.4.3.2.9. FL₂₀ (mg/ml) data from the EC/HO study and the COLIPA study for the nine common test chemicals (from Zanvit *et al.*, 1999). A version of INVITTOX Protocol No. 120 was used.

Chemical	FL20 COLIPA *		FL20 EC/HO **	
	Company # 4	FRAME	Company # 4	FRAME
SLS 3%	9.28	6.7	14.7	25
Triton X-100 5%	13.27	13.10	163.72	167
Benzalkonium chloride 1%	10.88	6.8	>100	91
SLS 15%	3.6	4.3	8.87	12
Triton X-100 10%	5.16	7.4	99.44	90
Benzalkonium chloride 5%	4.12	3.8	>25	18
Cetylpyridinium bromide 6%	9.49	0.35	>25	93
Cetylpyridinium bromide 10%	7.45	0.38	>25	27
Benzalkonium chloride 10%	3.63	0.46	>25	19

* The protocol tested in the COLIPA study used a 15 minute chemical exposure duration and measured FL 4h following the chemical exposure

** The protocol tested in the EC/HO study used a 1 minute chemical exposure duration and measured FL immediately following the chemical exposure

As there was no PM for the EC/HO FL₂₀ (mg/ml) study data which converted the values into classifications of ocular irritation it is difficult to compare the predictive capacity of the two protocols for ocular irritation. As the protocols differed greatly, it was not considered appropriate to apply the PM from the COLIPA study to the EC/HO study data. Zanvit *et al.*, (1999) reported Pearson's and Spearman's correlation coefficients for these data respective to *in vivo* MMAS. There was higher correlation with the *in vivo* MMAS for the COLIPA data for both laboratories in comparison to the EC/HO study data produced by the same laboratories (Zanvit *et al.*, 1999).

5.4.3.3. Attempt to combine the data using weight-of-evidence approaches

If COLIPA study and ECVAM Prevalidation study data are considered representative of INVITTOX Protocol No. 120, there were more data available to enable the inter-laboratory variability of INVITTOX Protocol No. 120 to be evaluated in comparison to the other INVITTOX Protocols. There were no inter-laboratory data available for INVITTOX Protocols No. 82 and No. 86. A protocol similar to INVITTOX protocol No 71 was tested in the EC/HO study (Balls *et al.*, 1995). This was a multi-laboratory study where 60 pure chemicals were tested in four laboratories. In comparison, INVITTOX Protocol No. 120 was tested in the COLIPA study where 29 surfactants and surfactant-based formulations were tested in two laboratories, and also in the ECVAM Prevalidation study; four chemicals and one formulation were tested in four laboratories for Phase II and ten surfactants were tested in three laboratories for Phase III. In addition to the test material data available for this protocol there was also a considerable amount of positive control (0.16mg/ml SDS) data. Subsequently, in terms of the quantity of data, INVITTOX Protocol No. 120 had greater weighting than INVITTOX Protocol No. 71.

INVITTOX Protocol No. 71 was used to test a wide range of chemicals (EC/HO study) and solubility was uncertain for some. In terms of the ranges of chemical classes, INVITTOX Protocol No. 71 would receive greater weighting than INVITTOX Protocol No. 120. Data for both INVITTOX Protocol No. 120 and INVITTOX Protocol No. 71 covered the same range of *in vivo* irritancy scores and therefore both protocols would receive equal weighting for this criterion.

It was known that not all of the laboratories that participated in the EC/HO study fully adhered to the FL assay protocol, therefore there was some doubt as to whether all the results were representative of the same protocol (R Clothier, personal communication). This fact was supported by the Pearson's correlation coefficient calculations in the publication and also the statistical analyses performed for this BRD; both found that there were significant differences between the data sets from each laboratory. Unfortunately, there was no other study reported where INVITTOX Protocol No. 71 had been tested in more than one laboratory, which would have enabled one to determine if the EC/HO study FL assay results for inter-laboratory variability were typical for this protocol. There was no statistical difference between the data sets submitted from the various laboratories that participated in the COLIPA study and also the ECVAM Prevalidation Study Phase II and Phase III. Therefore, there was greater evidence in support of the reproducibility INVITTOX Protocol No. 120 rather than irreproducibility of INVITTOX Protocol No. 71.

- Annex III Between-laboratory FL assay data for a. COLIPA test materials:
 i) chemicals ii) formulations b. EC/HO study chemicals; c. ECVAM
 Prevalidation Phase II; d. ECVAM Prevalidation Phase III –provided on CD
- Annex A CTFA study Phase III formulation ingredients (from draft HET-CAM
 BRD: Appendix C2 (ICCVAM/NICEATM, 2004)
- Annex B COLIPA Study test chemicals and formulations compositions (from
 COLIPA)
- Annex C Formulation compositions from Company # 3

6. Predictive Capacity (Module 5)

The majority of the literature that featured a FL assay protocol assessed the assay's ability to predict chemical-induced *in vivo* eye irritation, or less commonly skin irritation. Most of the chemicals/materials tested using the FL assay have been cosmetic ingredients and formulations, i.e. those that humans are likely to be routinely exposed to and are likely to gain entry to the eye, rather than exposure to industrial chemicals in the workplace. In the literature, the *in vivo* data used to assess the predictive ability of the FL assay were often Draize MAS or MMAS. In some cases, the *in vivo* ocular irritation data were not provided in the publications but a statement made regarding the correlation of the FL assay ranking of materials with the *in vivo* ranking for toxicity, based on the chemical structures and properties, i.e. Hubbard *et al.*, (1994). EU risk phrases featured in only one study (Clothier *et al.*, 1994) and were assigned based on the historical *in vivo* data available. No study featured classifications according to the GHS or the EPA classification systems for ocular irritation.

It is to be noted that summarised *in vivo* data were often presented in the literature and the variability of this data was rarely discussed. There is no consensus for how the variability of *in vivo* data should be addressed. The impact of *in vivo* data variability differs according to the EU, GHS and EPA classification schemes which use the raw *in vivo* data in different ways, to determine the various classifications. ECVAM is currently evaluating the effect of *in vivo* variability in these different classification systems.

Table 6.1.1. Provides information regarding the CTFA study Phase III (Gettings *et al.*, 1996) which tested the TEP assay/ INVITTOX Protocol No. 86. This was the only data set for which raw *in vitro* and *in vivo* data were available.

6.1. Studies with available raw data

Table 6.1.1. Table presenting the relevant information for each study where raw *in vitro* and *in vivo* data were available

INVITTOX Protocol No. 86

Study	No. of Laboratories	No. of Chemicals	No. of Products	Coded (Y/N)	Chemical Classes	Ranges of Toxicity Tested	Physico-chemical properties
CTFA Evaluation of Alternatives Program: An Evaluation of <i>In Vitro</i> Alternatives to the Draize Primary Eye Irritation Test: Phase III (Gettings <i>et al.</i> , 1996) IRAG LAB B (Botham <i>et al.</i> , 1997)	1	0	23 (data for 2 formulations were not reported)	independent coding	representative surfactant-based personal care formulations	MAS Range: 2.3-43	liquid formulations -compositions known

6.1.2. *In vivo* reference data used to assess the performance of the FL assay for studies where raw *in vitro* and *in vivo* data were available

Table 6.1.2. *In vivo* reference data used to assess the performance of the FL assay for studies where raw *in vitro* and *in vivo* data were available

INVITTOX Protocol No. 86

Study	Species and Protocols used as ref. Data	Sources of Information	No. of Labs	No. of experiments	Quality of Data (GLP)	Data Format
CTFA Evaluation of Alternatives Program: An Evaluation of <i>In Vitro</i> Alternatives to the Draize Primary Eye Irritation Test: Phase III (Gettings <i>et al.</i> , 1996) IRAG LAB B (Botham <i>et al.</i> , 1997)	Draize test -albino rabbits	<i>In vivo</i> and <i>in vitro</i> experiments conducted in parallel. Random block design resulted in materials tested x6 on different days in (3M, 3F) albino rabbits, diluted formulations tested on M rabbits only	1	6	GLP	Raw animal data available. In publication, summarised MAS values provided for various ocular tissues and total MAS scores

6.1.3. Brief description of the studies with available raw *in vitro* and *in vivo* data

a. Statistical approaches detailed in the CTFA publications (Gettings *et al.*, 1996)

The 'response profile' of the TEP assay (INVITTOX Protocol No. 86) data from a single test laboratory was compared to MAS values for each formulation, and the agreement between the data sets assessed. A post-hoc threshold value was assigned to the TEP assay data, which produced the least number of false positive and false negative classifications of irritancy according to the FHSA classifications based on the *in vivo* data. The ability of the TEP assay to separate pairs of test materials with significantly different MAS values and to determine which was the least irritant according to the MAS values was evaluated. This analysis was termed 'concordance analysis' and used because it did not require subjective interpretation of the *in vitro* data. A disadvantage of this type of analysis was that when TEP assay results separated test materials that were not separated by the *in vivo* data, it was considered to be a discordance. However, the seemingly greater sensitivity of the *in vitro* test was questionable in the absence of supporting human data.

A Monte Carlo computer re-sampling procedure was used to give estimates of the variability of the discordance rates for test materials that had different Draize MAS. This analysis was performed to determine which of the numerous test assays had similar or dissimilar discordance rates. The re-sampling was performed 100 times; the SDs of the simulated values were used as estimates of the SDs for the discordance rates based on observed MAS. A separation index was produced which indicated the extent of agreement between the TEP assay data and the Draize data in their abilities to separate the irritancy potential of pairs of test materials. The assays were ranked in order to select the assay(s) that had discordance rate(s) similar to the discordance rate for the *in vivo* data. The ranking was also performed using separation index values as this reflected discordance rates when Draize data *did* and *did not* separate the pairs of test materials.

Regression analyses were performed to describe the predictive capacity of the TEP assay for MAS. Base-10 logarithm TEP assay data were plotted against *in vivo* MAS, and regression models fitted using weighted least squares; weights were inversely proportional to the sum of the variability of the *in vivo* data and TEP assay data. The regression models were fitted in order to smooth and interpolate the *in vitro* response within the range of the observed *in vivo* data. 95% prediction bounds were calculated for the range of logarithmic *in vitro* scores.

b. Principal results reported in the literature

Twenty-three of the 25 surfactant-based formulations were tested/reported for the TEP assay from one laboratory only. A post-hoc threshold value enabled 14 of the FHSA irritants and all of the non-irritants to be correctly identified. The specificity of the TEP assay and PM was greater than the sensitivity, 100% and 78% respectively. Base-10 logarithm transformed data were used for the statistical comparisons with the quantitative *in vivo* data. Analyses of the discordance rates when the Draize test data did and did not separate test materials and the separation index led to the TEP assay being selected amongst 13 other assays for having greater concordance with the *in vivo* data than the other assays evaluated.

Table 6.1.3. The findings reported in the literature for those studies which used PMs to assess the predictive capacity of the TEP assay.

Study	Protocol	Prediction/ Classification Model	Time- point(s) used for PM	No of test materials	Type of Test materials	No. of materials under- predicted*	No. of materials over- predicted*
CTFA Evaluation of Alternatives Program: An Evaluation of <i>In Vitro</i> Alternatives to the Draize Primary Eye Irritation Test: Phase III (Gettings <i>et al.</i> , 1996)	TEP assay (INVITTOX Protocol No. 86)	A post-hoc threshold value (≤ 2.6) was assigned to achieve the least number of false positives and false negatives according to the FHSA classification system	15min	23	Surfactant-based formulations	4	0

* based on the threshold value in featured in Gettings *et al.*, (1996)

NB. The assay featured in the CTFA publication (Gettings *et al.*, 1996) was later accepted as INVITTOX Protocol No. 86.

6.1.4. Compilation of data on predictive capacity of the test method from studies with raw data available

Refer to Annex Vai -Compilation of raw TEP assay data and EU, GHS and EPA classifications generated by entering raw *in vivo* data into the ECVAM template v6.

6.1.4.1. Description & rationale for the PM applied and statistical approaches used

a. Classification systems

The aim of these analyses was to determine the predictive capacity of the TEP assay for the EU, GHS and EPA classification systems. The CTFA study did not report *in vivo* eye irritation classifications for the EU, GHS nor the EPA classification systems. Therefore, raw *in vivo* data (supplied by Company # 1) were entered into the ECVAM template v6 which allowed raw Draize data to be converted into EU, GHS and EPA classifications. These *in vivo* based classifications were used for comparisons with the predicted classifications based on the TEP assay results.

b. Prediction models (PMs)

INVITTOX Protocol No. 86 includes a PM (table 6.1.4.1.1.). The EC₅₀ (%) values produced by the test materials are used to determine if the irritancy levels 'pass' or 'fail.' The PM is only capable of distinguishing potential irritants from non-irritants and not capable of distinguishing different levels of irritancy (i.e. R36 and R41 irritants). The PM featured in INVITTOX Protocol No. 86 does not correlate the predicted classifications with any classification schemes used for regulatory proposes. The PM contains a range of TEP assay values where it was not possible to determine conclusively if the materials were irritants or non-irritants, these materials were classified as 'borderline'.

Table 6.1.4.1.1. PM from INVITTOX Protocol No. 86

EC50 (%)	TEP Rating
≤1.8	fail
>1.8 < 2.2	borderline/undetermined
≥ 2.2	pass

The PM from INVITTOX Protocol No. 86 had been applied to TEP assay data for surfactant-based formulations tested in-house at Company # 3. Therefore it was considered suitable to apply this PM to the data generated from the CTFA study Phase III (Gettings *et al.*, 1996). A modified version of the PM featured in INVITTOX Protocol No. 86 was submitted to ECVAM; the same ranges of values were used, but correlated to EU, GHS and EPA classifications (table 6.1.4.1.2.).

Table 6.1.4.1.2. Modified TEP assay PM for predicting EU, GHS and EPA classifications.

EC50 (%)	EU Classification	GHS Classification	EPA Classification
≤1.8	R36 or R41	Cat 1, Cat 2A or Cat 2B	Category I/ II
>1.8 < 2.2	Borderline/Undetermined	Borderline/Undetermined	Borderline/Undetermined
≥ 2.2	Not Classified	No Category	Category III/ IV

The predicted EU, GHS and EPA classifications based on the TEP assay data were compared to the actual classifications obtained by entering raw *in vivo* data into the ECVAM template v6. The same ranges of TEP assay values were used to distinguish the three principal classes used in the EU and GHS classification systems. The EPA classification system has four principal classes, but the PM was not able to distinguish Category IV and Category III test materials. The PM was not capable of distinguishing sub-categories, i.e. Category 2A and Category 2B GHS classifications. All Category IV and Category III materials were treated as non-irritants for analyses in this BRD. The conclusions formed regarding the predictive capacity of the TEP assay would be the same for all classification systems unless the *in vivo* classifications did not correlate, i.e. a test material with an R36 EU classification would be expected to have a Category 2 GHS classification and a Category II EPA classification.

Company # 3 also submitted to this BRD, TEP assay/INVITTOX Protocol No. 86 data for formulations, and corresponding raw *in vivo* data. The data set supplied by Company # 3 could not be analysed to assess the protocol's predictive capacity for ocular irritation, due to the format of the *in vivo* data; greater-than values were predominately reported which prevented EU, GHS and EPA classifications from being generated by the ECVAM template v6. The raw *in vitro* and *in vivo* data from Company # 3 were provided in Annex Vc of this BRD.

6.1.4.2. Description of the performance compared to the reference and eventually the human situation for each study

Concordance analyses were performed according to ECVAM guidance. Contingency tables (tables 6.1.4.2.1.a-c.) show the concordance of predicted and actual classifications for the EU, GHS and EPA classification systems.

TEP assay data from the CTFA study Phase III were analysed (Gettings *et al.*, 1996). Of 25 test formulations, TEP assay data for only 23 formulations were reported; no data were reported for the polishing scrub (HZZ) and the facial cleanser (HZZ). It was not stated in the publication (Gettings *et al.*, 1996) whether these two formulations were not tested or whether the results were not accepted or submitted.

Table 6.1.4.2.1.a. Contingency tables for CTFA Phase III data generated using INVITTOX Protocol No. 86 and the PM featured in table 6.1.4.1.2. -EU classification system

Test Prediction	EU Classification			
	NC	R36	R41	
NC	8	0	2	
Borderline	2	0	0	
R36/R41	3	0	8	
Total	13	0	10	23

The results showed that the TEP assay and PM had a tendency to over-predict the irritancy potential of the surfactant-based formulations. Nearly 40% of the EU Not Classified formulations were classified as irritants or borderline materials. All but two of the R41 formulations were correctly classified; the facial cleansing foam and gel cleanser were misclassified as Not Classifieds. This would seem to indicate that either the TEP assay was not capable of detecting the irritancy potential of all types of surfactant-based formulations or that the ranges of values of the PM needed some adjustment. As a version of this assay is routinely used at Company # 3 for testing surfactants and surfactant-based formulations, it is hypothesised that the PM required some refinement rather than the assay being incompatible with the test formulations.

Table 6.1.4.2.1.b. Contingency tables for CTFA Phase III data generated using INVITTOX Protocol No. 86 and the PM featured in table 6.1.4.1.2. -GHS classification system

Test Prediction	GHS Classification			
	NC	Cat 2	Cat 1	
NC	10	0	0	
Borderline	1	1	0	
Cat 1/ Cat 2	3	0	8	
Total	14	1	8	23

The results for the GHS classifications show that all the Category 1 irritants were correctly predicted. This contrasts to the results for the EU and EPA classification system where two severe irritants were predicted to be non-irritants. One Category 2 irritant and one No Category formulation were considered 'borderline' and did not receive any classification.

Table 6.1.4.2.1.c. Contingency tables for CTFA Phase III data generated using INVITTOX Protocol No. 86 and the PM featured in table 6.1.4.1.2. -EPA classification system

Test Prediction	EPA Classification				
	Cat IV	Cat III	Cat II	Cat I	
Cat III/IV	2	6	0	2	
Borderline	0	2	0	0	
Cat I/ Cat II	0	3	0	8	
Total	2	11	0	10	23

The predictive capacity of INVITTOX Protocol No. 86 was similar for the EU, GHS and EPA classification systems due to the general concordance for the classifications based on the *in vivo* data, i.e. a R36 (EU) irritants were generally also labelled Category 2 (GHS) and Category II (EPA) irritants.

The predictive capacity for the three different classification systems indicated that further work was required to increase the predictive capacity of the protocol and/or PM. If INVITTOX Protocol No. 86 was to be used for regulatory purposes, work should also focus on determining a range of values that could be used to distinguish the mild irritants from the severe irritants.

Summary tables were compiled in order to compare the predictive capacity of the TEP assay for the different classification schemes (tables 6.1.4.2.2.a-c). Under ECVAM guidance the predictive capacity of the assay was ascertained by determining the predictive capacity for the 'non-irritants versus the rest' and the 'severe irritants versus the rest.' It was determined by ECVAM that if there were fewer than 10% mild irritants in the data set, only the 'non-irritants versus the rest' analysis should be performed, as the 'severe irritants versus the rest' analysis would only produce similar results in the absence of data for mild irritants. For the data sets featured in tables 6.1.4.2.2.a-c, there were less than 10% of the formulations classed as mild irritants in all the classification systems, therefore only the 'non-irritants versus the rest' analysis was performed for these data.

For interpretation of the 'non-irritant versus the rest' analyses below, the following should be noted:

-Concordance referred to the number of correctly predicted non-irritants and irritants by the FL assay; irritants that were identified but predicted to have the wrong classification were still counted as correctly identified.

-Sensitivity referred to the number of correctly predicted irritants by the FL assay as a proportion of the number of actual irritants. This analysis only determined that irritants were predicted to be irritants by the FL assay but did not distinguish between the different irritancy classifications predicted (e.g., R36 and R41). 'Borderline' formulations were not counted as an irritancy classification.

-Specificity referred to the number of correctly predicted non-irritants by the FL assay as a proportion of the total number of actual non-irritants.

-Positive Predictivity referred to the number of correctly predicted irritants as a proportion of the total number of predicted irritants i.e. this analysis only determined that irritants were predicted to be irritants by the FL assay but did not distinguish between the different irritancy classifications predicted (e.g., R36 and R41).

-Negative Predictivity referred to the number of correctly predicted non-irritants as a proportion of the total number of predicted non-irritants

-False Positive Rate referred to the number of non-irritants predicted to be irritants as a proportion of the total number of non-irritants

-False Negative Rate referred to the number of irritants predicted to be non-irritants as a proportion of the total number of irritants

These descriptions are relevant for all 'non-irritant versus the rest' analyses throughout Section 6 of this BRD.

Table 6.1.4.2.2.a. Evaluation of the performance of INVITTOX Protocol No. 86 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
CTFA Phase III (Gettings <i>et al.</i> , 1996)	Refer to Table 6.1.4.1 .2.	23	69.6	16/23	80	8/10	61.5	8/13	72.7	8/11	80	8/10	27.3	3/11	20	2/10

Table 6.1.4.2.2.b. Evaluation of the performance of INVITTOX Protocol No. 86 for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
CTFA Phase III (Gettings <i>et al.</i> , 1996)	Refer to Table 6.1.4.1 .2.	23	78.3	18/23	88.9	8/9	71.4	10/14	72.7	8/11	100	10/10	23.1	3/13	0	0/8

Table 6.1.4.2.2.c. Evaluation of the performance of INVITTOX Protocol No. 86 for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
CTFA Phase III (Gettings <i>et al.</i> , 1996)	Refer to Table 6.1.4.1 .2.	23	69.6	16/23	80	8/10	61.5	8/13	72.7	8/11	80	8/10	27.3	3/11	20	2/10

NB. As the PM was not able to distinguish EPA Category III and Category IV materials, materials with these classifications were considered in the analyses concerning non-irritants. Refer to Annex Vai for the origins of the *in vitro* data and *in vivo* classifications.

6.1.4.3. Discussions

a Descriptions of limitations of the test method (i.e. applicability domain based on the results of the compiled data)

Analyses to determine the applicability domain of the TEP assay/INVITTOX Protocol No. 86 were limited to analyses of the data for the surfactant-based formulations tested in the CTFA study Phase III (Gettings *et al.*, 1996). As only formulations were tested, the predictive capacity of this protocol and PM for chemicals could not be established.

The PM was not capable of distinguishing different levels of irritancy, and there was also a range of values within which the protocol and PM could not distinguish materials as either irritants or non irritants, but which were labelled as 'borderline.' Two of the 23 formulations could not have their EU, GHS, EPA classifications predicted because the results fell within the range of 'borderline'. If it was known that the assay had not been able to test the two materials for which data was not submitted, in addition to the two formulations classed as 'borderline' materials, the PM was not able to predict the irritancy for 16% (4/25) of the test materials. As there were only data for 21 test formulations, the various results summarised in tables 6.1.4.2.2.a-c. were greatly affected by each individual result.

The concordance rates for the 'non-irritants versus the rest' for the three classification systems ranged 70-78%. The protocol and PM had a greater predictive capacity for the GHS classifications rather than the EU and EPA classifications. Concordance rates and sensitivity were higher, and false positive and false negative rates were lower for the GHS classifications in comparison to the EU and EPA classifications. The negative predictivity was the same for the EU and EPA classifications but higher for the GHS classifications.

There were more formulations classified as non-irritants than irritants in each of the three classification systems, and the data indicated that the protocol and PM had a higher sensitivity rather than specificity for all three classification systems. This finding suggested that further work was required to modify the protocol and/or PM to increase the specificity of the assay and PM.

The physical properties of the test formulations were evaluated to determine if certain features affected the predictive capacity of INVITTOX Protocol No. 86. As the predictive capacity of INVITTOX Protocol No. 86 for the EU, GHS and the EPA classifications were similar, comparisons were made with EU classifications only. All of the test formulations were liquids so it could not be determined if the assay had a good predictive capacity for solids which are known to be difficult to measure in many *in vitro* assays. Two of the three EU Not Classified formulations that were misclassified as being EU irritants and both the EU R41 irritants misclassified as Not Classified formulations were coloured materials. As 17 of 23 formulations were coloured, the data did not conclusively indicate that colour did or did not affect the predictive capacity of the TEP assay. There were 13 viscous formulations (12 viscous plus one gel) in the data set, ten of these viscous formulations were also coloured. Only one of the five materials with their irritancy potential over-predicted was a viscous liquid. One of the two coloured formulations that had their irritancy potential under-predicted was also viscous. One Not Classified formulation that was neither coloured nor viscous was also misclassified as a R36/R41 irritant. As 13 of the 23 test materials were viscous solutions, the data did not indicate

that the assay was unable to measure materials with viscosity and that the predictive capacity of the assay was adversely affected. As many test formulations were both coloured and viscous it was difficult to distinguish the effects of these two physical properties on the predictive capacity of the protocol. Overall, there was greater evidence in support of colour rather than viscosity reducing the predictive capacity of the protocol and PM.

In consideration that the TEP assay had only been used to measure the effects of materials which it was designed to test, the overall predictive capacity of the assay based on a limited data set was not promising. In addition, the assay was only evaluated for its ability to distinguish irritants from non-irritants as the PM was not able to distinguish different levels of irritancy. As there were few formulations labelled as 'mid-range' irritants according to the *in vivo* data, it was difficult to surmise from the range of data whether the assay would be capable of distinguishing mid-range irritants if the PM was modified. Further work could investigate if there are different ranges of values that could be used to distinguish mild and severe irritants, and this would probably involve removing the range of ('borderline') values within which the PM is unable to distinguish materials as either irritants or non-irritants.

b. Possible rationale for the differences observed

As only one type of material was tested using the TEP assay/INVITTOX Protocol No. 86 in a single laboratory, possible rationale for differences observed was limited to the variation between the predictive capacity of the protocol and PM for the various classification systems (EU, GHS, EPA). As there were few test materials in the data set, the concordance values etc., calculated were greatly affected by the prediction of every single test formulation. For example, sensitivity was calculated as 80% for the EU classification system and the EPA classification based on the correct prediction of 8/10 irritants, whilst sensitivity was calculated as 88.9% for the GHS classification system based on the correct prediction of 8/9 irritants. Therefore, although the data indicate that the predictive capacity of the protocol varies considerably according to the different classification systems, further testing is required to give significance to the results presented.

Table 6.2.1. Table presenting the relevant information for each study where raw data were not availableINVITTOX Protocol No. 71 (and similar)

Study	No. of Laboratories	No. of Chemicals	No. of Products	Coded (Y/N)	Chemical Classes	Ranges of Toxicity Tested	Physico-chemical properties
Loss of Trans-epithelial Impermeability of a Confluent Layer of Madin-Darby Canine Kidney (MDCK) Cells as a Determinant of Ocular Irritancy Potential (Shaw <i>et al.</i> , 1990)	1	22	0	N	metals, surfactants, alcohols, solvents	non-irritants, R41 irritants; one R36 irritant	compounds only, solids and liquids. MW: 32.042-1228
The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test (Balls <i>et al.</i> , 1995)	4	60	0	Y	pure chemicals ranging in mechanisms and potency	MMAS Range: 0-108	liquids, crystals, powders. MW: 39.99707-1228
The Evaluation of Pesticide Ingredients and Formulations <i>In Vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995)	1	2	4	Y	pesticides and vehicles	permethrin, severe irritant (K-C) and cypermethrin, minimal irritant (K-C)	Cypermethrin -pale brown, viscous liquid/semi-solid, MW 416.31; permethrin - Seige II- gel, lanasol-liquid, MW 391.28
Evaluation of Tissue Culture Insert Membrane Compatibility in the Fluorescein Leakage Assay. (Ward <i>et al.</i> , 1997a)	1	6	0	N	surfactants (non-ionics and ionics)	MMAS Range: 0-108	MW: 288.37687-1228, liquids

FRAME In-house (non-published) -INVITTOX Protocol No. 71

Study	No. of Laboratories	No. of Chemicals	No. of Products	Coded (Y/N)	Chemical Classes	Ranges of Toxicity Tested	Physico-chemical properties
Report on Comparison of 40 Cosmetic and Domestic Formulations Supplied by Company # 8 and Evaluated by 3 <i>In Vitro</i> Cytotoxicity Tests (FRAME, 1992)	1	0	40	Y	variety of cosmetics and industrial cleaners representing different levels of toxicity, mainly surfactant-based formulations	24h LVET MMAS Range: 0-44.7	formulations, liquids, powders, creams. MW: 149.1894-1228 (2 compounds).

INVITTOX Protocol No. 82

Study	No. of Laboratories	No. of Chemicals	No. of Products	Coded (Y/N)	Chemical Classes	Ranges of Toxicity Tested	Physico-chemical properties
Development of a Fixed Dose Approach for The Fluorescein Leakage Test. (Clothier <i>et al.</i> , 1994)	1	21	0	Y	mixture of industrial chemicals representing those found in industry	non-irritants and R41 chemicals; only 2 R36 chemicals	liquids, solids acetyldehyde volatile. MW: 39.997 – 376.275
Comparative Evaluation of Five <i>In Vitro</i> Tests for Assessing the Eye Irritation Potential of Hair-care Products (Jones <i>et al.</i> , 2001)	1	0	17	Y	Surfactant-based formulations -shampoos and conditioners	mild, moderate and substantial irritants relative to the toxicities of these product types	formulation gels

INVITTOX Protocol No. 86

Refer to Section 6.1.

INVITTOX Protocol No. 120 (and similar)

Study	No. of Laboratories	No. of Chemicals	No. of Products	Coded (Y/N)	Chemical Classes	Ranges of Toxicity Tested	Physico-chemical properties
Fluorescein Leakage Test: a Useful Tool in Ocular Safety Assessment (Cottin and Zanvit, 1997)	1	20	23	Surfactant-based formulations only	non-ionic, anionic, cationic surfactants and surfactant-based products	MAS Range: 0.3-50.3	limited information, MW 228.4172-1310. lotions, gels for formulations
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase II) (Southee, 1998)	4	4	1	Not stated	glycerol, triton X-100, CTAB, ammonium nitrate, Company # 3 Baby shampoo	MMAS Range: 0-33.8	MW: 80.04-364.45, liquids
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	3	10	0	Y	mild surfactants	MMAS Range: 0-37	MW: 288.37687-414.6, liquids, powders, flakes
Ocular Irritancy Assessment of Cosmetics Formulations and Ingredients: Fluorescein Leakage Test. (Zanvit <i>et al.</i> , 1999) DATA FROM COLIPA	2	11	23	Y	surfactants and surfactant-based formulations soluble in HBSS	MMAS Range: 0-108	For chemicals only: liquids, viscous liquids, crystals, flakes, MW: 250.3802-384.4419

Others

Study	No. of Laboratories	No. of Chemicals	No. of Products	Coded (Y/N)	Chemical Classes	Ranges of Toxicity Tested	Physico-chemical properties
Investigations of the MDCK Permeability Assay as an <i>In Vitro</i> Test of Ocular Irritancy (Gautheron <i>et al.</i> , 1994) IRAG LAB A (Botham <i>et al.</i> , 1997)	1	32	0	N	surfactants, alcohols, miscellaneous	MAS Range: 0-98	liquids, solids, MW: 41.0524-448.0873
Evaluation of a Human Corneal Epithelial Cell Line as an <i>In Vitro</i> Model for Predicting Ocular Irritation (Kruszewski <i>et al.</i> , 1997)	1	20	25	N	Chemicals, many surfactants	MMAS Range: 1.33-64.75	MW 39.99707-384.4419, liquids, crystals, solids, 25 surfactants from CTFA Phase III study
Assessment of Initial Damage and Recovery Following Exposure of MDCK Cells to an Irritant (Clothier <i>et al.</i> , 1999)	1	3	0	N	known irritants (surfactants)	MMAS Range: 4- 56	Tween 20, isopropanol, benzalkonium chloride

Blue font indicates that raw *in vivo* data were available.

6.2.2. *In vivo* reference data used to assess the performance of the FL assay

Table 6.2.2. *In vivo* reference data used to assess the performance of the FL assay

INVITTOX Protocol No. 71 (and similar)

Study	Species and Protocols used as ref. Data	Sources of Information	No. of Labs	No. of experiments	Quality of Data (GLP)	Data Format
Loss of Trans-epithelial Impermeability of a Confluent Layer of Madin-Darby Canine Kidney (MDCK) Cells as a Determinant of Ocular Irritancy Potential (Shaw <i>et al.</i> , 1990)	Draize test	Botham <i>et al.</i> , (1989) Acute and topical toxicity profiles for substances involved in the <i>in vitro</i> validation of the <i>in vivo</i> ocular irritation model. Nottingham, FRAME	1	1-6 (according to Botham <i>et al.</i> , (1989)).	N	Data summarised into NC, R36, R41 categories and ocular irritation categories (OIC)
The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test (Balls <i>et al.</i> , 1995)	Draize test -ECETOC (1992) normally NZ White rabbits	Historical data from Bagley <i>et al.</i> , (1992), ECETOC, (1992)	unknown	3-6.	GLP	Summarised MMAS for individual chemicals
The Evaluation of Pesticide Ingredients and Formulations <i>In Vitro</i> and Correlations with <i>In Vivo</i> Data (Clothier <i>et al.</i> , 1995)	OECD Guidelines for Testing of Chemicals (1987) No. 45 "Acute Eye Irritation/ Corrosion" using NZ White Rabbit-stated in confidential report only	Rat oral LD ₅₀ (mg/ml) from either Purchase <i>et al.</i> (1993) and/or RTECS (1985). Summarised Draize scores for individual tissues	1	OECD Guideline 405	OECD Guideline 405	<i>In vivo</i> Draize scores provided by manufacturer-not stated if MAS or MMAS
Evaluation of Tissue Culture Insert Membrane Compatibility in the Fluorescein Leakage Assay. (Ward <i>et al.</i> , 1997a)	Draize test -normally NZ White rabbits	Draize MMAS data from Balls <i>et al.</i> , (1995) and ECETOC (1992)	unknown	3-6.	GLP	Mean MAS ±SD

FRAME In-house (non-published) -INVITTOX Protocol No. 71

Study	Species and Protocols used as ref. Data	Sources of Information	No. of Labs	No. of experiments	Quality of Data (GLP)	Data Format
Report on Comparison of 40 Cosmetic and Domestic Formulations Supplied by Company # 8 and Evaluated by 3 <i>In Vitro</i> Cytotoxicity Tests (FRAME, 1992)	24h LVET	24h LVET MAS from Company # 8	1	unknown	not stated	Mean LVET 24h MAS

INVITTOX Protocol No. 82

Study	Species and Protocols used as ref. Data	Sources of Information	No. of Labs	No. of experiments	Quality of Data (GLP)	Data Format
Development of a Fixed Dose Approach for The Fluorescein Leakage Test. (Clothier <i>et al.</i> , 1994)	Draize test	Historical Draize test data interpreted and reclassified from Botham <i>et al.</i> , (1989)	many	1-6 (according to Botham <i>et al.</i> , (1989)).	N	Materials categorised as R41, R36 or NC
Comparative Evaluation of Five <i>In Vitro</i> Tests for Assessing the Eye Irritation Potential of Hair-care Products (Jones <i>et al.</i> , 2001)	Draize test -species not stated	Historical MAS scores and market research	1	N/A	not stated	No <i>in vivo</i> scores given. Test materials ranked according to benchmark values.

INVITTOX Protocol No. 120 (and similar)

Study	Species and Protocols used as ref. Data	Sources of Information	No. of Labs	No. of experiments	Quality of Data (GLP)	Data Format
Fluorescein Leakage Test: a Useful Tool in Ocular Safety Assessment (Cottin and Zanvit, 1997)	Draize test	MAS data	unknown	3	Method of the Officiel de la Republique Français, (24/10/1984).	MAS and mean Draize scores for every chemical at 1h, 24h, 48h, 72h, 96h, 7 days
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase II) (Southee, 1998)	Draize test	MMAS data -assume historical but source not stated	unknown	unknown	unknown	MMAS scores given for individual chemicals
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	Draize test	Historical MMAS data supplied by Company # 1 and Gautheron <i>et al.</i> , (1994b) which met requirements set by BIBRA	unknown	unknown	unknown	Mean MMAS \pm SD scores given for individual chemicals
Ocular Irritancy Assessment of Cosmetics Formulations and Ingredients: Fluorescein Leakage Test. Zanvit <i>et al.</i> , (1999)	Draize test (as for COLIPA)	Draize MMAS data provided by BIBRA International from the COLIPA study (Botham <i>et al.</i> , 1997)	unknown	3	OECD Guideline 405	Raw animal data available (from IIVS); MMAS in publication

Others

Study	Species and Protocols used as ref. Data	Sources of Information	No. of Labs	No. of experiments	Quality of Data (GLP)	Data Format
Investigations of the MDCK Permeability Assay as an <i>In Vitro</i> Test of Ocular Irritancy (Gautheron <i>et al.</i> , 1994a) IRAG LAB A (Botham <i>et al.</i> , 1997)	Draize test	24h total Draize scores from EEC study (Gautheron <i>et al.</i> , 1994b) and Kennah <i>et al.</i> , (1989)	1	3	data fulfils IRAG criteria, but GLP status not stated	17 chemicals with 24h total Draize scores, cornea, conjunctival scores, and days to clear
Evaluation of a Human Corneal Epithelial Cell Line as an <i>In Vitro</i> Model for Predicting Ocular Irritation (Kruszewski <i>et al.</i> , 1997)	Draize test -normally NZ White rabbits	Data from ECETOC Eye Irritation Reference Chemicals Data Bank (ECETOC 1992) and from CTFA Phase III (Gettings <i>et al.</i> , 1996)	unknown	3-5 depending on material and assay	GLP	Mean Draize MAS w/o SD or SEM
Assessment of Initial Damage and Recovery Following Exposure of MDCK Cells to an Irritant (Clothier <i>et al.</i> , 1999)	Draize test (ECETOC-normally NZ White rabbits)	Historical MMAS data from ECETOC (1992)	1	n≥3	unknown	Summarised MMAS

6.2.3. Brief description of the studies without raw data available

a. Statistical Approaches

The statistical approaches used for analysing the predictive capacity of the FL assay protocols featured in table 6.2.1. were outlined in brief below. Studies featured in table 6.2.1. and not discussed below did not statistically analyse the data. In these cases, the predicted classifications or rankings of the *in vitro* and the *in vivo* data were usually reported.

INVITTOX Protocol No. 71 (and similar)

The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test (Balls *et al.*, 1995).

The statistics featured in the EC/HO study were employed to evaluate inter-laboratory variability and the relationship between the *in vitro* data and the *in vivo* data (MMAS).

The *in vitro* data were plotted against MMAS data from the ECETOC Technical Report on Eye Irritation (ECETOC, 1992) and additional *in vivo* data for 14 other test chemicals. Greater-than and less-than values were also plotted and were included as the maximal or minimal concentration tested in calculations for; linear regression, 95% confidence intervals and correlation coefficients.

Evaluation of Tissue Culture Insert Membrane Compatibility in the Fluorescein Leakage Assay (Ward *et al.*, 1997a).

A MMAS value of less than 15 was used to indicate surfactants with mild or no irritancy. A cut-off value of 50mg/ml was used to discriminate irritants from non-irritants according to FL₂₀ (mg/ml) values measured immediately following the chemical exposure.

INVITTOX Protocol No. 82 (and similar)

Comparative Evaluation of Five *In Vitro* Tests for Assessing the Eye Irritation Potential of Hair-care Products (Jones *et al.*, 2001)

The amount of FL (%) measured immediately after and 72 hours following the chemical exposure were compared to the respective FL (%) results from a benchmark shampoo and conditioner. If the FL (%) value was lower than the benchmark, the test material was considered to have an 'acceptable' level of irritancy; if the FL (%) value was higher than the benchmark, the test material was considered to have an 'unacceptable' level of irritancy. FL (%) values that were similar to the benchmark, i.e. $\pm 20\%$ were considered to require 'further investigation' before concluding if the test material had acceptable or unacceptable levels of ocular irritancy.

INVITTOX Protocol No. 120 (and similar)

A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom *et al.*, 1997)

Data submitted to the COLIPA study underwent an independent quality check by BIBRA International (UK). The publication stated that the protocol used was according to Cottin

et al., (1992) but a very similar version of this protocol was later accepted as INVITTOX Protocol No. 120.

The FL assay PM was based on data from 43 surfactants and surfactant-based formulations that were dissolved or suspended in HBSS. The PM used log₁₀ FL assay results taken 4h after the 15 minute chemical exposure. Cut-off values for the *in vitro* data indicated that three classes of observed *in vivo* ocular irritation could be distinguished; 'non-irritant/slight', 'moderate', and 'irritant/severe'.

Log₁₀ FL assay 4h results were plotted against observed MMAS values. Cut-off values were applied to the *in vitro* and *in vivo* data. The classifications based on the *in vitro* results showed good correlation to the *in vivo* classifications.

The lead laboratories provided a 95% prediction interval for the PMs. The 95% prediction interval refers to the range in which 95% of the predicted values were likely to fall within. The 95% prediction/confidence intervals (CIs) were produced for Altman and Bland comparisons which were used as they allow false positives and false negatives to be considered. The difference between the average observed MMAS values and the average predicted MMAS values was plotted as a continuous line. The $\pm 2SD$ from the mean was plotted on the graph which represented the 95% CI. Wide intervals indicated a reduced predictive capacity of the *in vitro* assay for *in vivo* ocular irritation. A mean difference greater than zero indicated that the assay had the tendency to under-predict *in vivo* eye irritation; a mean difference less than zero indicated that the assay had the tendency to over-predict *in vivo* eye irritation. The mean difference of the FL assay was plotted as zero. Values that were plotted above the CI represented materials with their *in vivo* ocular irritancy potential under-predicted. Values that were plotted below the CI indicated materials that had their *in vivo* ocular irritancy potential over-predicted.

A classification PM was also produced and analysed by kappa statistics using equal weightings, linear weightings and quadratic weightings. Linear weightings give a greater weight to predicted values that differ by two (rather than one) classification groups in relation to the *in vivo* classification. Quadratic weightings give an even greater weighting than the linear analyses, to results that differ by two classifications in respect to the *in vivo* classifications. The 95% CIs were included in all of the analyses. Altman Bland comparisons were carried out on the classifications.

Fluorescein Leakage Test: A Useful Tool in Ocular Safety Assessment (Cottin and Zanvit, 1997).

Pearson (linear) and Spearman (rank) correlations were computed for the FL assay data and corresponding historical MAS data at 0h, 1h, 24h and day 7 for the 43 test samples. NB. Greater-than 1000mg/ml data were plotted as '1000mg/ml' data. Weighted kappa statistics were used to measure the agreement between classifications based on *in vivo* and *in vitro* data.

Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998).

The data were independently analysed by BIBRA International. FL₂₀ (mg/ml) results measured 4h after the 15 minute chemical exposure were used for comparisons with *in vivo* MMAS data. The results were classified into various categories of ocular irritation.

Kappa statistics were used to compare inter-laboratory reproducibility of predicted *in vivo* classifications, and to assess the capacity of the FL assay protocol for predicting *in vivo* classes of ocular irritation. Linear and quadratic kappa statistics were used. Quadratic weightings give an even greater weighting, than the linear analyses, to results that differ by two classifications in respect to the *in vivo* classifications.

Others

Investigations of the MDCK Permeability Assay as an *In Vitro* Test of Ocular Irritancy (Gautheron *et al.*, 1994).

The FL₅₀ (mg/ml) results were compared to the *in vivo* data (Gautheron *et al.*, (1994b) (Kennah *et al.*, 1989)) using the Pearson correlation test. It was indicated that due to the low amount of variation in the *in vivo* data the Pearson's correlation test could only be performed for comparative purposes, thus the correlation was given little weighting.

Evaluation of a Human Corneal Epithelial Cell Line as an *In Vitro* Model for Predicting Ocular Irritation (Kruszewski *et al.*, 1997).

FR₈₅ (%) values were used which measure the concentration of the test material causing fluorescein retention to decrease to 85% relative to the negative control. FR₈₅ (%) values were compared to MMAS values, corneal scores, corneal opacity scores, corneal area scores, iris scores, conjunctival redness scores, conjunctival discharge scores, and 'days to clear' scores for the 25 CTFA study Phase III surfactant-based formulations.

b. Main results

Table 6.2.3. The findings reported in the literature for those studies which used PMs to assess the predictive capacity of the various FL assay protocols.

INVITTOX Protocol No. 71 (and similar)

Study	Protocol	Prediction/ Classification Model	Time- point(s) used for PM	No of test materials	Type of Test materials	No. of materials under- predicted [§]	No. of materials over- predicted [§]
The EC/HO International Validation Study on Alternatives to the Draize Eye Irritation Test (Balls <i>et al.</i> , 1995)	According to Tchao (1988) (i.e., INVITTOX Protocol No. 71)	FL20 (mg/ml) values: <100mg/ml =R36 or R41; >750mg/ml =NI; 100-750mg/ml =R41 if no recovery after 72h, R36 if recovery after 72h	1min	60	Chemicals with various mechanisms	N/A**	N/A**
Evaluation of Tissue Culture Insert Membrane Compatibility In the Fluorescein Leakage Assay (Ward <i>et al.</i> , 1997a)	According to Shaw <i>et al.</i> , (1990) –Anopore insert used	FL20 (mg/ml) values: <50mg/ml =irritants; ≥50mg/ml =non-irritants	1min	12	surfactants	0	0

INVITTOX Protocol No. 82 (and similar)

Study	Protocol	Prediction/ Classification Model	Time- point(s) used for PM	No of test materials	Type of Test materials	No. of materials under- predicted [§]	No. of materials over- predicted [§]
Development of a Fixed Dose Approach for The Fluorescein Leakage Test. (Clothier <i>et al.</i> , 1994)	INVITTOX Protocol No. 82 (Fixed Dose)	50mg/ml hypothesised to be the cut-off point to distinguish R36/R41 chemicals from NC if \geq FL20 (%) was taken to indicate significant toxicity.	1min and consideration of 72h recovery rates	22	Chemicals with various mechanisms	1	1
Comparative Evaluation of Five <i>In Vitro</i> Tests for Assessing the Eye Irritation Potential of Hair-care Products (Jones <i>et al.</i> , 2001)	Essentially a Fixed Dose method using INVITTOX Protocol No. 71*	Results compared to the FL assay values for benchmark shampoo and conditioner. Material acceptable if lower than benchmark; 'further investigation' defined by \pm 20% from benchmark value	1min (and assume with consideration of 72h values)	17	Shampoo and conditioner formulations	0 (2 non-classified)	10 (2 non-classified)

INVITTOX Protocol No. 86 (and similar)

See Section 6.1.

INVITTOX Protocol No. 120 (and similar)

Study	Protocol	Prediction/ Classification Model	Time- point(s) used for PM	No of test materials	Type of Test materials	No. of materials under- predicted [§]	No. of materials over- predicted [§]
A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom <i>et al.</i> , 1997) –FRAME data	Cottin <i>et al.</i> , (1992) (i.e., INVITTOX Protocol No. 120)	For FL20 (mg/ml) value at T4: >100mg/ml =non irritant/slight, MMAS <15; 20-100mg/ml =moderate, MMAS 15-30; <20mg/ml =irritant/severe, MMAS> 30	4h	30	Surfactants and surfactant-based formulations soluble in HBSS	1	6
A Summary Report of the COLIPA International Validation Study on Alternatives to the Draize Rabbit Eye Irritation Test (Brantom <i>et al.</i> , 1997) –Company # 4 data	Cottin <i>et al.</i> , (1992) (i.e., INVITTOX Protocol No. 120)	For FL20 (mg/ml) value at T4: >100mg/ml =non irritant/slight, MMAS < 15; 20-100mg/ml =moderate, MMAS 15-30; <20mg/ml =irritant/severe, MMAS> 30	4h	33	Surfactants and surfactant-based formulations soluble in HBSS	0	4
Fluorescein Leakage Test: A Useful Tool in Ocular Safety Assessment (Cottin and Zanvit, 1997)	Not stated (similar to INVITTOX Protocol No. 120)	For FL20 (mg/ml) value at T0: ≥100mg/ml =slightly irritant, ≥20,100mg/ml =moderate irritant <20mg/ml =irritant	0h	43	Surfactants and surfactant-based formulations soluble in HBSS	2***	1***
Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay (Phase III) (Southee, 1998)	INVITTOX Protocol No. 120	Based upon COLIPA PM using FL20 (mg/ml) at T4: results: >100mg/ml =non-irritant/slight; 20-100mg/ml =moderate; <20 mg/ml =irritating/severe	4h	10	Mild surfactants	0 (based on mean classif. from three labs)	6 (based on mean classif. from three labs)

* The protocol used different incubation periods; 10 seconds for the shampoos and 30 seconds for the conditioners. ** PM reported in the publication was not applicable to the data reported. *** results were taken directly from graphical representation of concordance of classifications (Cottin and Zanvit *et al.*, 1997).[§] the results were based on the classification systems featured in the publications.

Additional discussions and comparisons of the conclusions for the studies featured in table 6.2.3. are detailed below. There are many different elements to the studies featured in table 6.2.3. that cause the conclusions regarding the predictive capacity of the various FL assays to differ from study to study. Examples include; use of different FL assay protocols, types of test materials, format and quality of *in vivo* data (i.e., MAS, MMAS, qualitative classification groups), PM used, level of statistical analyses. The principal reasons for the various findings regarding the FL assays predictive capacities in the different studies are outlined.

INVITTOX Protocol No. 71 (and similar)

The predictive capacity of INVITTOX Protocol No. 71 for ocular irritation has been evaluated in a number of studies. The largest study was the EC/HO study where 60 pure chemicals were tested in four laboratories (Balls *et al.*, 1995). The FL assay was initially designed to detect immediate damage caused by test materials to the impermeability of an epithelial monolayer. The chemicals tested in the EC/HO study covered a wide range of chemical classes and encompassed many different physical properties. The FL₂₀ (mg/ml) data were compared to Draize MMAS scores using Pearson's correlation coefficients and Spearman's rank correlation coefficients. The Pearson's correlation coefficients from the four laboratories testing the FL assay ranged from -0.151 to -0.565. This range of values was comparable to other correlation coefficients produced by a number of *in vitro* tests evaluated as part of this study.

INVITTOX Protocol No. 71 also featured in the study of Shaw *et al.*, (1991). Twenty-two chemicals were tested which covered a range of potencies and mechanisms. The FL assay results were ranked and the order of the chemicals showed good correlation to *in vivo* classifications which ranged from non-irritants, mild irritants and severe irritants. The conclusions from study of Shaw *et al.*, (1991) differed to those from the EC/HO study predominately because Shaw *et al.*, (1991) compared the FL values to a qualitative classification scheme rather than quantitative MMAS scores. Additionally, fewer chemicals were also tested which created less opportunity for the assay to be challenged by different chemical mechanisms and potencies.

This protocol was also used within a test battery, comprised of the NRR assay and the kenacid blue assay, to test two pesticides and their formulations (Clothier *et al.*, 1995). The formulation vehicles without the pesticide ingredients were found to be more toxic than the pure pesticide ingredients in all three assays. Permethrin was considered to be a severe eye irritant and cypermethrin a minimal eye irritant according to the Kay and Calandra scoring system.

The protocol of Shaw *et al.*, (1990) was used to test the effects of EC/Home Office validation trial surfactants (Balls *et al.*, 1995) on four types of inserts which can be used for the FL assay; 0.2 and 0.02 pore sized Anopore inserts with aluminium oxide membranes, 0.4 pore sized Millicell-CM inserts with PTFE biopore membranes and 0.45 pore sized Millicell-HA inserts with a membrane comprised of mixed cellulose esters (Ward *et al.*, 1997a). It was concluded that the Anopore membrane produced results that correlated better with MMAS than those obtained using a Millicell-HA membrane (Ward *et al.*, 1997a).

INVITTOX Protocol No. 82 (and similar)

INVITTOX Protocol No. 82 was developed as a consequence of the findings of the EC/HO study (Balls *et al.*, 1995). A set concentration was used to test all chemicals and the amount of induced FL was used to distinguish irritants and non-irritants according to the EU classification system. The protocol was very similar to INVITTOX Protocol No. 71 with the following exceptions; only one chemical concentration was tested, the length of time for FL to occur and the endpoint measured (FL%). The authors reported that a chemical concentration of 50mg/ml was effective at distinguishing irritants from non-irritants. This was the only reported study where this particular Fixed Dose FL assay has been used. The test chemicals were identical to those tested by Shaw *et al.*, (1991) who also reported that the protocol was able to distinguish different levels of irritancy.

Jones *et al.*, (2001) stated that a method similar to that of Tchao (1989) was used to test hair care products although an evaluation of the protocol described showed greater concordance with INVITTOX Protocol No 82. However, the exposure durations and the time allowed for FL were shorter than those featured in INVITTOX Protocol No. 82. A total of 17 shampoos and conditioners were tested neat and the FL assay performed at 0h and 72h. The results were compared to benchmark shampoo and conditioner formulations for which *in vivo* data were available. The amount of FL produced by the test formulations had to be lower than that caused by the respective benchmarks in order for the test formulations to be considered to have 'acceptable' levels of potential ocular irritation. No formulation had its irritancy potential under-predicted but ten materials had their irritancy potentials over-predicted. It is difficult to compare the results of this study with others that also featured INVITTOX Protocol No. 82 as here formulations were tested rather than pure chemicals. In addition, the endpoint used by this protocol differed to the FL₂₀ (mg/ml) endpoint generally used with INVITTOX Protocol No. 82. The type of *in vivo* data used for this study was not as clearly defined as in the other publications.

INVITTOX Protocol No. 120 (and similar)

This protocol was accepted as an INVITTOX Protocol following the promising results from the COLIPA study (Brantom *et al.*, 1997) and the ECVAM Prevalidation study (Southee, 1998). In the COLIPA study, only surfactants and surfactant-based formulations that were soluble in HBSS were tested. This contrasted to the studies for INVITTOX Protocol No. 71 where numerous types of chemicals with various levels of solubility were tested. A classification PM was defined prior to the start of the COLIPA study. Four test materials were over-predicted by Company # 4 and six materials were over-predicted by the FAL; only one formulation was under-predicted by the FAL. INVITTOX Protocol No. 120 was concluded to be a promising *in vitro* test based upon the COLIPA results (Zanvit *et al.*, 1999). This finding contrasted to the results presented for INVITTOX Protocol No. 71 tested in the EC/HO study. However, the types of materials tested in the COLIPA study were limited to surfactants soluble in HBSS. There was also approximately half the number of materials tested in the COLIPA study in comparison to the EC/HO study. However, the COLIPA study tested both chemicals and formulations that enabled the predictive capacity of INVITTOX Protocol No. 120 for surfactants and surfactant-based formulations to be ascertained.

Results for the nine materials common to the COLIPA study and the EC/HO study were compared to the *in vivo* MMAS data using Pearson's correlation and Spearman's rank

coefficients. INVITTOX Protocol No. 120 (COLIPA study data) was found to show a greater correlation to the *in vivo* data in comparison to INVITTOX Protocol No. 71 (EC/HO study data) (Zanvit *et al.*, 1999).

Cottin and Zanvit (1997) used the assay to measure the effects of surfactants and surfactant-based products for 43 samples. The authors reported that different recovery rates were observed for anionic (greater recovery) and cationic surfactants (less recovery) despite similar amounts of initial damage. The FL assay data were concluded to correlate well with the historical Draize *in vivo* data according to Spearman's correlation test results.

A protocol similar to INVITTOX Protocol No. 120 was also tested as part of the ECVAM Prevalidation study (Southee, 1998). The predictive capacity of the assay was tested in Phase III. Ten mild surfactants were tested in three independent laboratories. The mean results (compiled from the three laboratories) were entered into the same PM that featured in the COLIPA study. Six materials had their irritancy over-predicted and none had their irritancy under-predicted. These results supported those from the COLIPA study which indicated that the FL assay and PM had a tendency to over-predict rather than under-predict ocular irritation potential. As the same PM was used in both studies it was difficult to determine whether the protocol and/or the PM needed refinement to reduce the incidence of over-predictions.

Others

Gautheron *et al.*, (1994a) tested a number of materials including surfactants and alcohols. The authors concluded that the FL assay data correlated well with 24h historical *in vivo* data for surfactants rather than other types of test materials. Some surfactants had their irritancy potential over-estimated. *In vivo* data were taken from two different studies (Gautheron *et al.*, 1994b; Kennah *et al.*, 1989) that were judged to be 'not comparable' according to the authors of the publication.

INVITTOX Protocol No. 71 was also used combined with the AB assay in the study of Clothier *et al.*, (1999; Assessment of initial damage and recovery following exposure of MDCK cells to an irritant). The FL₁₅ (%) was measured to coincide with data from another type of FL assay performed with HCE-T cells. FL₁₅ (%) data for three test materials (tween 20, isopropanol, benzalkonium chloride (1%)) showed some correlation to Draize MMAS.

A modified TEP assay, using HCE-T cells, was used to test 20 chemicals selected from the ECETOC database (1992) and the 25 formulations tested in CTFA study Phase III (Gettings *et al.*, 1996) (Kruszewski *et al.*, 1997). FR₈₅ (%) data were correlated to Draize MAS scores using Pearson Correlations which was $r=0.71$ for the chemicals and $r=0.86$ for the CTFA study Phase III formulations. For both data sets, outlier chemicals were as equally prevalent throughout the entire range of Draize MMAS scores. It was reported that the most variable FR₈₅ (%) results were for the CTFA formulations that had low Draize scores leading the authors to conclude that the assay was not able to distinguish mildly irritating surfactant-based formulations (Kruszewski *et al.*, 1997). The data for the chemicals indicated that the assay was most effective for predicting the ocular irritation potential of surfactants, alcohols and least effective for acids, bases and acetates (Kruszewski *et al.*, 1997).

6.2.4. Compilation of data on predictive capacity of the test method from studies without raw data

Refer to Annex V*ai*

6.2.4.1. Description & rationale for the Prediction Model(s) applied and statistical approach(es) used

The aim of these analyses was to determine the abilities of the different FL assay protocols to predict *in vivo* ocular irritation according to the EU, GHS and EPA classification systems. However, only the publication reporting results for the Fixed Dose FL assay assessed the protocol's predictive capacity for EU classifications. No studies featuring the FL assay assessed the assays predictive capacity for GHS or EPA classifications. Subsequently there was limited relevant *in vivo* data available from the literature to compare the predictive capacities of the various FL assay protocols.

ECVAM provided a template that allowed raw data from either three or six test animals to be converted into EU, GHS and EPA classifications. ECVAM used this template to produce EU, GHS and EPA classifications for the chemicals tested in the EC/HO study (Balls *et al.*, 1995). The classifications were provided directly from ECVAM for concordance analyses in this BRD. The chemicals tested in the COLIPA study were based on those from the EC/HO study which enabled the same classifications based on the *in vivo* data, provided by ECVAM for the EC/HO study, to be used.

Prediction models (PMs)

In order to evaluate the predictive capacity of the various FL assays without bias, PMs were submitted to ECVAM before all the FL assay data and the *in vivo* based EU, GHS, and EPA classifications were compiled. The various PMs used to evaluate the data collected for this BRD are detailed.

INVITTOX Protocol No. 71 with recovery data

The EC/HO study publication (Balls *et al.*, 1995) proposed/devised a PM that could be used for FL₂₀ (mg/ml) data produced by INVITTOX Protocol No. 71, that also included 72h recovery data (refer to table 2.2.1.2.). This PM correlated FL₂₀ (mg/ml) values with the EU classification system. This PM was not used to evaluate the FL assay data featured in the EC/HO study (Balls *et al.*, 1995) as 72h data were not produced; the PM was only stated as an 'application' for the results. Subsequently, for this BRD the PM was only applied to the relatively few chemicals that also had 72h data which were taken from the publications of Shaw *et al.*, (1991) and Ward *et al.*, (1997a). Similar threshold values as published in the PM for the EU classifications were also applied to the respective GHS and EPA classifications. The PM was modified to include clarifications made by the authors of this BRD to specify recovery or deterioration; 30% increases or decreases at the 72h FL measurement relative to the initial measurement were used (table 6.2.4.1.1.). In addition, the ranges of values used to determine R41 and R36 irritants were increased from >100 ≤ 750mg/ml to 0 ≤ 750mg/ml. The PM was not able to distinguish Category III and Category IV EPA classifications.

Table 6.2.4.1.1. PM submitted to ECVAM for FL₂₀ (mg/ml) values with recovery data, based upon the PM featured in the EC/HO study (Balls *et al.*, 1995).

FL ₂₀ (mg/ml)	EU Classification	GHS Classification	EPA Classification
>750mg/ml	Not Classified	No Category	Category III/IV
≤750mg/ml and ≥30% recovery at 72h	R36	Cat 2A or Cat 2B	Category II
≤750mg/ml and ≥30% deterioration at 72h	R41	Cat 1	Category I

In order to evaluate the predictive capacity of this protocol and PM, the *in vivo* data were taken from other sources. An explanation of the *in vivo* data sources and how EU, GHS and EPA classes were assigned are given below.

ECVAM provided some raw *in vivo* data (originally from ICCVAM/NICEATM), which were to be entered into the ECVAM template v6 to generate the EU, GHS and EPA classifications. The raw *in vivo* data for six out of 12 chemicals requiring classifications could not be found in the TNO file (Prinsen and Koeter, 1993) due to chemical coding. Other documents containing raw *in vivo* data also contained coded chemicals. As the pairing of *in vitro* and *in vivo* data was based purely on chemical name and concentration due to a lack of chemical information, there was some doubt as to the suitability of the pairings of *in vitro* and *in vivo* data. It was hypothesised that the variability generated by possible differences between the ICCVAM/NICEATM and ECVAM classifications would be minor in comparison to the variability generated by pairing the *in vitro* and *in vivo* data together based only on chemical name. Therefore it was decided to take EU, GHS and EPA classifications directly from the ICCVAM/NICEATM 'Current Status of *In Vitro* Test Methods for Detecting Ocular Corrosives and Severe Irritants Draft Background Review Documents (BRDs) (2004); *in vivo* classifications from the Hen's Egg Test -Chorioallantoic Membrane (HET-CAM) draft BRD (2004) were predominately used. Therefore, the *in vivo* data consisted of those classifications generated by entering data into the ECVAM template v6 and those classifications taken directly from the ICCVAM/NICEATM publications. The authors of this BRD acknowledge that the quality of the analyses performed with paired *in vitro* and *in vivo* data was considerably lower than those with raw *in vivo* data available. Therefore the results and conclusions drawn from these analyses are considered to be speculative and only potentially indicative of the predictive capacities of the different protocols.

The criteria used by ICCVAM/NICEATM for assigning the EU, GHS and EPA classifications to the various types of Draize data are outlined in brief. Data from one to six animals were used. The criteria were:

- the test material had to be tested in at least three rabbits unless a severe effect (e.g. corrosion of cornea, lesion persistence) was observed in a single animal upon which a classification could then be assigned
- a volume of 0.1ml or 0.1g had to be tested. An exception was made for materials that were tested in lower quantities but which produced a severe effect.
- the minimum observations of the eye had to be made at 24h, 48h and 72h after material instillation if the lesions were not severe.

For the GHS classification system (UN, 2003), classifications were assigned to each individual animal based on the average scores for corneal opacity, iritis, conjunctival redness and/or conjunctival chemosis for the 24h, 48h and 72h time points. The combinations of different GHS classifications given to the various animal results were used to assign the irritancy classification to the test material. In order to use all the data available, an additional rule was introduced into the classification scheme by ICCVAM/NICEATM (table 6.2.4.1.2.)

Table 6.2.4.1.2. ICCVAM/NICEATM criteria for assigning GHS classifications to Draize data (from draft HET-CAM BRD (2004)).

GHS Category	Criteria Necessary for Substance Classification
Category 1	1. At least 1 of 3 animals or 2 of 6 animals classified as Category 1, Group A 2. <i>One of 6 animals classified as Category 1, Group A and at least 1 of 6 animals classified as Category 1, Group B</i> 3. At least 2 of 3 animals or 4 of 6 animals classified as Category 1, Group B
Category 2A	At least 2 of 3 animals or 4 of 6 animals classified as Category 2A or Category 2B
Category 2B	At least 2 of 3 animals or 4 of 6 animals classified as Category 2B
Nonirritant	At least 2 of 3 animals or 4 of 6 animals classified as nonirritant

Text in italics indicates the additional criteria added by ICCVAM/NICEATM.

The ICCVAM/NICEATM rules for applying the EU classifications (EU, 2001) to the available Draize data are described. Calculation of the EU classifications varied according to the number of test animals. In the case of data from three animals, average corneal opacity, iritis and conjunctival chemosis were scored for each animal across the 24h, 48h and 72h time-points. When data were available for more than three animals, the overall corneal opacity, iritis, conjunctival chemosis and redness were scored for all of the test animals across the 24h, 48h and 72h time-points. Using these values, classifications were assigned based on the minimal positive average when data were available for only three animals, and on the overall average when data were available for more than three animals (table 6.2.4.1.3.). For test materials that produced many different classifications for the group of test animals, no classification was assigned

Table 6.2.4.1.3. ICCVAM/NICEATM criteria for assigning EU classifications to Draize data (from draft HET-CAM BRD (2004)).

EU Category	Three Animals Tested	Greater than Three Animals Tested
R41	Two or more animals where the average animal Draize scores over Days 1, 2, and 3 were: Opacity ≥ 3 Iritis = 2 Or At least one animal (at end of observation period) where the effect has not reversed ¹	Overall mean animal Draize scores over Days 1, 2, and 3 were: Opacity ≥ 3 or Iritis > 1.5 Or At least one animal (at end of observation period) where the effect has not reversed
R36	Two or more animals where the average animal Draize scores over Days 1, 2, and 3 were: $2 \leq \text{Opacity} < 3$ $1 \leq \text{Iritis} < 2$ Redness ≥ 2.5 Chemosis ≥ 2	Overall mean animal Draize scores over Days 1, 2, and 3 were: $2 \leq \text{Opacity} < 3$ $1 \leq \text{Iritis} < 2$ Redness ≥ 2.5 Chemosis ≥ 2

The majority of data used by ICCVAM/NICEATM for assigning these classifications were based on average score data, average animal data or irritancy classifications.

For EPA classifications to be assigned, normally at least three animals are tested (EPA, 1996). A positive response is classified as an opacity or iritis score of equal to or greater-than one, or a redness or chemosis score of equal to or greater-than 2 (table 6.2.4.1.4.). The observed score can occur at any time up to 21 days following the material exposure. The most severe classification recorded for any of the test animals is assigned to the test material.

Table 6.2.4.1.4. EPA classification system for ocular irritation (from draft HET-CAM BRD (2004)).

EPA Category	Criteria for Animal Classification
Category I	- Corrosive, corneal involvement or irritation (iris or cornea score ≥ 1 or redness or chemosis ≥ 2) persisting more than 21 days or - Corneal effects that are not expected to reverse by 21 days
Category II	- Corneal involvement of irritation clearing ¹ in 8-21 days
Category III	- Corneal involvement of irritation clearing in 7 days or less
Category IV	- Minimal or no effects clearing in less than 24 hours

It is to be noted that the explanations for the *in vivo* data provided above were relevant only for the *in vitro* data set which included recovery data. The remaining FL assay protocols and PMs analysed were assessed according to classifications generated by entering *in vivo* data into the ECVAM template v6. Original individual animal data were available for the CTFA study Phase III (Gettings *et al.*, 1996), the EC/HO study (Balls *et al.*, 1995) and the COLIPA study (Brantom *et al.*, 1997).

INVITTOX Protocol No. 71 (and similar)

The majority of data collected for this BRD were generated following a one minute exposure and FL measured immediately afterwards. As no PM existed for this type of data, a new PM was devised using the data from the EC/HO study (table 6.2.4.1.5.). In devising the range of values for the three principal classes of EU, GHS and EPA classification systems, greater efforts were made to reduce the number of false negative predictions rather than false positive predictions. The PM was unable to distinguish EPA Category III and Category IV irritancy classifications.

Table 6.2.4.1.5. PM for FL₂₀ (mg/ml) values

FL ₂₀ (mg/ml)	EU Classification	GHS Classification	EPA Classification
>750mg/ml	Not Classified	No Category	Category III/IV
>100 ≤ 750mg/ml	R36	Cat 2A or Cat 2B	Category II
≤100mg/ml	R41	Cat 1	Category I

INVITTOX Protocol No. 120 (and similar)

A substantial amount of data was collected for materials that had been tested for a 15 minute exposure period. The majority of data were from the COLIPA study (Brantom *et al.*, 1997), where the FL was measured 4h following the original chemical exposure. A PM for this data type featured in the COLIPA study and distinguished three classes of irritancy; non-irritant/slight, moderate, irritant/severe. The same ranges of FL₂₀ (mg/ml) values were used to distinguish the three principal classes featured in the EU, GHS and EPA classification systems (table 6.2.4.1.6.). The PM was not able to distinguish EPA Category III and Category IV classifications.

Table 6.2.4.1.6. PM for FL₂₀ (mg/ml) values produced 4h following a 15 minute chemical exposure period (modified from Brantom *et al.*, 1997).

FL ₂₀ (mg/ml)	EU Classification	GHS Classification	EPA Classification
>100 mg/ml	Not Classified	No Category	Category III/IV
20 -100 mg/ml	R36	Cat 2A or Cat 2B	Category II
<20 mg/ml	R41	Cat 1	Category I

INVITTOX Protocol No. 82

The PM as featured in the publication of Clothier *et al.*, (1994) was submitted for the Fixed Dose FL assay. Clothier *et al.*, (1994) proposed that 50mg/ml could be used to distinguish non-irritants from R36 and/or R41 chemicals if FL₂₀ (%) was taken to indicate significant toxicity (table 6.2.4.1.7.). In the publication, consideration of recovery or deterioration of effects at 72h following the chemical exposure, was used to distinguish R36 and R41 irritants. The amount of recovery or deterioration required to be considered significant of an effect was not stated in the publication of Clothier *et al.*, (1994). Therefore, for analyses of this BRD, a 30% reduction in FL at 72h in comparison to the initial FL assay immediately following the exposure was taken to indicate recovery. Less than 30% recovery or any further deterioration at 72h in comparison to the initial FL assay was taken to indicate R41 chemicals.

Although the authors (Clothier *et al.*, 1994) stated that further work was required to modify this threshold value according to the type of materials being tested, the same threshold value was used here as no additional data for this protocol were collected. The EU classifications as stated in the publication were used to determine the predictive capacity of the protocol and PM.

Table 6.2.4.1.7. Modified PM for INVITTOX Protocol No. 82 data (from Clothier *et al.*, (1994)).

Chemical tested at 50mg/ml	EU Classification
causes <FL20 (%)	Not Classified
causes ≥FL20 (%) with more than 30% recovery at 72h	R36
causes ≥FL20 (%) with less than 30% recovery or further deterioration at 72h	R41

INVITTOX Protocol No. 86

The only *in vitro* and *in vivo* data available for analyses of the TEP assay/ INVITTOX Protocol No. 86 were analysed in section 6.1. of this BRD.

FL Assay with HCE-T cells

A modified FL assay was devised by Kruszewski *et al.*, (1997). This protocol uses human transfected corneal epithelium cells (HCE-T) in order to try and increase the predictive capacity of the assay for human ocular irritation by using relevant cell types of human origin. Data for this protocol were available for the CTFA study Phase III surfactant-based formulations. The percentage of the test formulation which enabled 85% of the sodium-fluorescein dye to be retained was measured (FR₈₅(%)). No PM was available for this protocol so the CTFA study Phase III data were plotted against the EU classifications generated by entering *in vivo* data into the ECVAM template v6 (figure 6.2.4.).

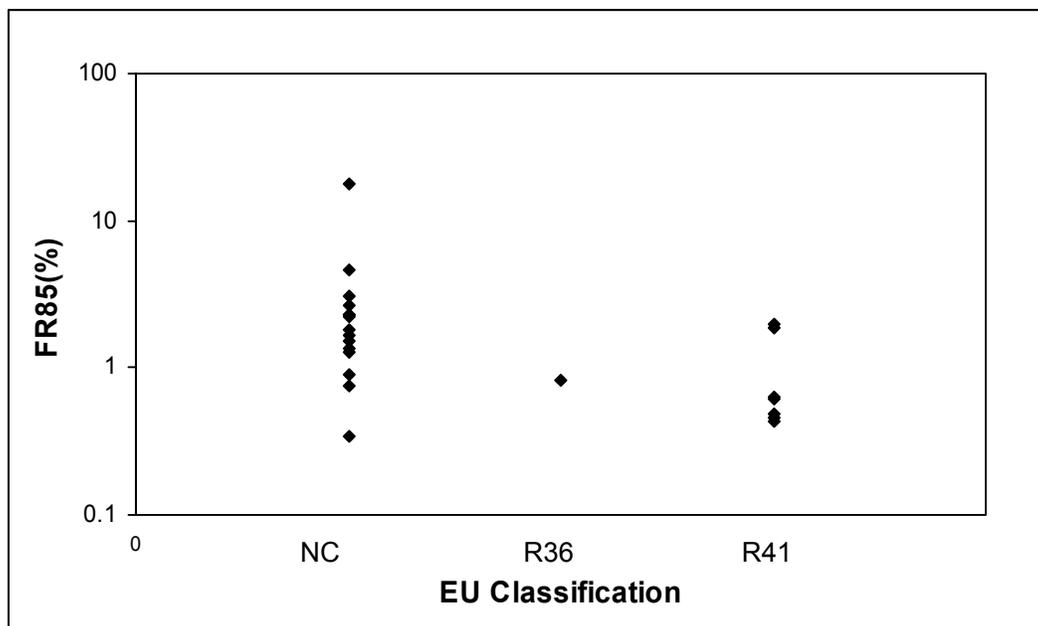


Figure 6.2.4. Correlation of FR₈₅ (%) data with EU classifications.

Both Not Classified formulations and R41 formulations were plotted within a similar range of values (figure 6.2.4.). Based upon this limited data set, this protocol did not appear able to distinguish surfactant-based formulations with different potencies as defined by EU classifications.

Comparisons of PMs

The range of values used in the PMs for the FL assay protocols, to distinguish the various categories featured in the EU, GHS and EPA classification systems, showed some correlation when protocol differences were taken into consideration. The PM for INVITTOX Protocol No. 71, uses the widest range of values to distinguish different irritancy classifications in comparison to the other PMs. As only a one minute chemical exposure is used, the concentration of the test material required to induce FL₂₀ and label a test material as an irritant is higher in comparison to the PM for INVITTOX Protocol No. 120. In comparison to INVITTOX Protocol No. 71, INVITTOX Protocol No. 120 uses, a different cell strain, a 15 minute chemical exposure and measures FL 4h after the chemical exposure. The FL assay at 4h entails that some recovery or further damage could occur before the assay is carried out. Due to these protocol differences, INVITTOX Protocol No. 120 generally produces lower values than INVITTOX Protocol No. 71 which is shown in the different ranges of values used for the PMs of these protocols.

INVITTOX Protocol No. 82 uses a fixed concentration of the test material. The amount of FL produced immediately following the exposure and 72h later determines the classification. The test concentration featured in the publication of Clothier *et al.*, (1994) was 50mg/ml although it was stated that this could be modified according to the type of test material. As the Fixed Dose FL assay is very similar to INVITTOX Protocol No. 71, the fixed concentration is within the range of values used to classify R41 irritants in the PM for INVITTOX Protocol No. 71.

The PM reported in INVITTOX Protocol No. 86 uses relatively low values to distinguish irritants and non-irritants. This protocol shows many similarities to INVITTOX Protocol No. 120 but the lower values reported for the PM of INVITTOX Protocol No. 86 was primarily due to the use of EC₅₀ values rather than FL₂₀ values.

6.2.4.2. Description of performance compared to reference and eventually, to the human situation for each study

The predictive capacities of the various FL assay protocols and their respective PMs for the EU, GHS and EPA classifications were evaluated. The predicted EU, GHS and EPA classifications were compared to the *in vivo* based EU, GHS and EPA classifications obtained using either the ECVAM template v6 or from the ICCVAM/NICEATM BRDs for organotypic models. As the predicted classifications used the same boundaries for all three classification systems (EU, GHS, EPA), the conclusions formed regarding the predictive capacity of the FL assay protocols were the same for all classification systems unless the *in vivo* classifications did not correlate (i.e. a test material with an R36 EU classification would be expected to have a Category 2 GHS classification and a Category II EPA classification). None of the PMs were capable of distinguishing sub-categories nor Category III or Category IV EPA classifications.

Contingency tables were created to display the concordance of the predicted classifications based on FL assay data with EU, GHS and EPA classifications based on *in vivo* data.

For all the different analyses performed in this section, tables of the *in vitro* and *in vivo* data, including the sources for both types, were provided (Annex Viii).

INVITTOX Protocol No. 71

Data for this protocol were generated in the EC/HO study (Balls *et al.*, 1995). The protocol was accepted as INVITTOX Protocol No. 71 following the results of this study. The mean data reported for each of the four laboratories that participated in the EC/HO study (Balls *et al.*, 1995) were analysed individually as it was known that some laboratories did not sufficiently adhere to the test protocol (personal communication, R Clothier); to some extent this prevented the predictive capacity of the protocol from being significantly affected by the results of one or more laboratories that failed to adequately follow the protocol. The different testing practices of the various laboratories was partially indicated by the different greater-than values that were reported by the different laboratories; e.g. some laboratories tested the chemicals up to their maximum level of solubility whilst others ceased testing at lower concentrations. Greater-than values that did not allow an irritancy classification to be assigned or indicated two possible classifications, were not reported in the contingency tables (6.2.4.2.1.a-c.) but referred to in the text following the appropriate table.

INVITTOX Protocol No. 71 results from the four laboratories were produced for all of the 59 test chemicals except for parafluoraniline for which laboratory 21 failed to report a result and ethyl-2-methylacetoacetate for which laboratory 18 failed to report a result. No formulations were tested and therefore, the conclusions regarding the predictive capacity of this protocol and PM were limited to chemicals only.

The EU, GHS and EPA classifications based on the *in vivo* data were provided by ECVAM directly. The concordance of the predicted classifications generated by entering INVITTOX Protocol No. 71 data into the PM featured in table 6.2.4.1.5., with the EU, GHS and EPA classifications based on the *in vivo* data, was assessed (tables 6.2.4.2.1.a-c.).

Table 6.2.4.2.1.a. Contingency table for EC/HO study data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. -EU classification system

Test Prediction	EU classification			
	NC	R36	R41	
NC	41	19	12	
R36	12	20	13	
R41	4	3	21	
Total	57	42	46	145

Based on the *in vivo* data, the 'study criteria were not met' to be able to determine the ocular irritation classifications for quiniacrine, potassium cyanate, L-aspartic acid, maneb, sodium oxalate, parafluoranile, triton X-100 (5%), formesafen, chlorhexidine and 2,2-dimethylbutanoic acid. Thiourea was reported to have killed all the test animals. Predictions for thiourea were not included in table 6.2.4.2.1.a. although all four

laboratories produced results which according to the PM rendered the chemical as either a NC or a R36 irritant.

There were nine incidences of NC and nine R36 classified chemicals according to the *in vivo* data that could not be distinguished as either NC or R36 chemicals based on the INVITTOX Protocol No. 71 data available. Dibenzyl phosphate was a R36 classified chemical according to the *in vivo* data that could not be given any classification according to the *in vitro* result from laboratory 21. There were ten incidences of chemicals labelled as R41 according to the *in vivo* data that could not be assigned a predicted classification because of the greater-than value being so low. There were 20 incidences of R41 labelled chemicals according to the *in vivo* data that produced *in vitro* results which did not distinguish if they were NC or R36 irritants.

Based on the results presented in table 6.2.4.2.1.a., for which definitive prediction were made, INVITTOX Protocol No. 71 had a greater predictive capacity for the NC chemicals (41/57 correctly predicted) rather than the R36 chemicals (20/42 correctly predicted) and the R41 chemicals (21/46 correctly predicted).

There were slightly more GHS classifications based on the *in vivo* data, in comparison to the number of available EU classifications. The concordance of predicted and actual GHS classifications was shown (table 6.2.4.2.1.b.). Based on the *in vivo* data, the study criteria were not met to be able to determine the GHS ocular irritation classifications for 2,2-dimethylbutanoic acid, fomesafen, L-aspartic acid, maneb, parafluoraniline, potassium cyanate, and thiourea. Data for thiourea which killed the test animals were not included in the table 6.2.4.2.1.b. This results for this chemical led to all chemicals stating that either thiourea was a No Category or a Category 2 irritant; none of the laboratories could distinguish between these two classifications.

Table 6.2.4.2.1.b. Contingency table for EC/HO study data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. -GHS classification system

Test Prediction	GHS classification			
	NC	Cat 2	Cat 1	
NC	36	24	14	
Cat 2	4	32	13	
Cat 1	3	4	21	
Total	43	60	48	151

In addition to the data presented in table 6.2.4.2.1.b. there were the following incidences of chemicals that could not be given a definitive predicted classification; eight incidences of No Category chemicals that were predicted to be No Category/Category 2, ten incidences of Category 2 chemicals that were predicted to be No Category/Category 2 chemicals, 30 incidences of Category 1 chemicals that were predicted to be No Category/ Category 2 chemicals. Also, there was one incidence of a No Category, one incidence of a Category 2A, and ten incidences of Category 1 chemicals that could not be given a predicted classification due to low greater-than values.

The predictive capacity of INVITTOX Protocol No. 71 for the GHS classifications was comparable to the protocol's predictive capacity for the EU classifications. There was a

greater proportion of Category 2 irritants, based on the *in vivo* data, featured in the data set for GHS classifications (60/151) in comparison to the number of R36 irritants (42/145). A similar proportion of GHS Category 2 irritants were correctly identified by INVITTOX Protocol No. 71 in comparison to the number of R36 irritants correctly identified; 32/60 and 20/42 respectively. More than half of the Category 1 irritants were under-predicted to be Category 2 irritants or No Category chemicals.

The predictive capacity of INVITTOX Protocol No. 71 for the EPA classification system was also investigated. The PM was not capable of distinguishing materials classified as Category III and Category IV. Only glycerol and PEG 400 were classified as Category IV irritants according to the *in vivo* data. There were no EPA classifications for the following materials: cetylpyridinium bromide 6%, 2,2-dimethylbutanoic acid, pyridine, benzoyl-L-tartaric acid, L-aspartic acid, parafluoraniline, potassium cyanate, promethazine HCl, quiniacrine, chlorhexidine, imidazole and sodium oxalate as the 'study criteria were not met'.

Table 6.2.4.2.1.c. Contingency table for EC/HO study data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. -EPA classification system

Test Prediction	EPA classification				
	Cat IV	Cat III	Cat II	Cat I	
Cat III/ IV	7	29	16	10	
Cat II	0	23	15	5	
Cat I	0	3	6	13	
Total	7	65	37	28	137

Data for thiourea which killed the test animals were not included in the table 6.2.4.2.1.c. All laboratories produced data which could not distinguish the chemical as being either Category II, Category III or Category IV.

In addition to the data presented in table 6.2.4.2.1.c. there were the following incidences of chemicals that could not be given definitive predicted classification; fourteen incidences of a Category III chemicals that were predicted to be Category II/III/IV, ten incidences of Category II chemicals that were predicted to be Category II/III/IV chemicals, 17 incidences of Category I chemicals that were predicted to be Category II/III/IV chemicals. Also, there was one incidence of a Category IV, one incidence of a Category II and seven incidences of Category I chemicals that could not be given a predicted classification due to low greater-than values.

Based on the *in vivo* data, there were fewer *in vivo* EPA classifications in comparison to the number of EU and GHS classifications. The predictive capacity of INVITTOX Protocol No. 71 and PM for the EPA classifications for the non-irritants (Category IV) and mild irritants (Category III) was lower than predictive capacity for the EU and GHS non-irritants and mild irritants. A similar predictive capacity was calculated for the severe irritants for all three classification systems.

The data featured in tables 6.2.4.2.1.a-c. were summarised in tables 6.2.4.2.2.a-c and 6.2.4.2.3.a-c for analyses of 'non-irritants versus the rest' for all three classification systems.

Table 6.2.4.2.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5.	49	67.6	98/145	64.8	57/88	71.9	41/57	78.1	57/73	56.9	41/72	28.1	16/57	35.2	31/88

Table 6.2.4.2.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5.	49	70.2	106/151	64.8	70/108	83.7	36/43	90.9	70/77	48.6	36/74	16.3	7/43	35.2	38/108

Table 6.2.4.2.c. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5.	49	62.0	85/137	60.0	39/65	63.9	46/72	60.0	39/65	63.9	46/72	36.1	26/72	40.0	26/65

NB. As the PM was not able to distinguish EPA Category III and Category IV materials, materials with these classifications were considered in the analyses concerning non-irritants. Refer to Annex Viii for the origins of the *in vitro* data and *in vivo* classifications.

As more than 10% of the test materials were classed as mild irritants for the EU, GHS and EPA data sets, 'severe irritants versus the rest' analyses was performed for these classification systems only (tables 6.2.4.2.3.a-c.).

For interpretation of the 'severe irritants versus the rest' analyses, the following should be noted:

-Concordance referred to the number of correctly predicted severe irritants by the FL assay relative to the number of non-irritants and mild irritants; non-irritants and mild irritants that were identified but predicted to have the wrong classification were still counted as correctly identified.

-Sensitivity referred to the number of correctly predicted severe irritants by the FL assay as a proportion of the total number of actual severe irritants.

-Specificity referred to the number of correctly predicted non-irritants and mild irritants by the FL assay as a proportion of the total number of actual non-irritants and mild irritants.

-Positive Predictivity referred to the number of correctly predicted severe irritants as a proportion of the total number of predicted severe irritants.

-Negative Predictivity referred to the number of correctly predicted non-irritants and mild irritants as a proportion of the total number of predicted non-irritants and mild irritants

-False Positive Rate referred to the number of non-irritants and mild irritants predicted to be severe irritants as a proportion of the total number of non-irritants and mild irritants

-False Negative Rate referred to the number of severe irritants predicted to be non-irritants or and mild irritants as a proportion of the total number of severe irritants

These descriptions are relevant for all 'severe irritant versus the rest' analyses throughout Section 6 of this BRD.

Table 6.2.4.2.3.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Severe irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5.	49	77.9	113/145	45.7	21/46	92.9	92/99	75	21/28	78.6	92/117	7.1	7/99	54.3	25/46

Table 6.2.4.2.3.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Severe irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5.	49	77.5	117/151	43.8	21/48	93.2	96/103	75.0	21/28	78.0	96/123	6.8	7/103	56.3	27/48

Table 6.2.4.2.3.c. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EPA classification system –Severe irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5.	49	82.5	113/137	46.4	13/28	91.7	100/109	59.1	13/22	87.0	100/115	8.3	9/109	53.6	15/28

NB. As the PM was not able to distinguish EPA Category III and Category IV materials, materials with these classifications were considered in the analyses concerning non-irritants. Refer to Annex Viii for the origins of the *in vitro* data and *in vivo* classification.

INVITTOX Protocol No. 71 with recovery data

The predictive capacity of INVITTOX Protocol No. 71 which also produced recovery data was analysed. Recovery data were only available for chemicals, therefore the predictive capacity of the protocol and PM for formulations could not be ascertained. The *in vitro* data were taken from the study of Shaw *et al.*, (1991) and Ward *et al.*, (1997a); Annex Viii provides the source of the *in vitro* data for each chemical. Chemicals that were only tested once by Shaw *et al.*, (1991) were not included in the analyses. It was also decided not to include Shaw *et al.*, (1991) data for SDS as the test concentration was not stated. FL assay data for SDS were available from the publication of Ward *et al.*, (1997a) where two different test concentrations were used.

There were no matching *in vivo* data for these studies so the *in vitro* and the *in vivo* data were paired from different sources (i.e. classifications produced using the ECVAM template v6 or taken directly from the ICCVAM/NICEATM BRDs) as described in section 6.2.4.1. Annex Viii provides the source of the *in vivo* data for each chemical. It is acknowledged that the *in vivo* and *in vitro* data which may have been generated by testing slightly different forms of the test chemical could largely account for predicted misclassifications. The results of this analysis are only to be used as an *indication* of the protocol's and PM's predictive capacity

The PM used to analyse the data was taken from the EC/HO study (Balls *et al.*, 1995) (table 6.2.4.1.1.). It was proposed that R36 and R41 irritants should be distinguished according to whether there was recovery or deterioration at the 72h time-point in comparison to the amount of FL measured immediately after the chemical exposure (Balls *et al.*, 1995). The degree of recovery or deterioration required in order to be classed as significant was not stated in the publication and for the analyses of this BRD was taken to be a 30% difference in relation to the first FL measurement.

Table 6.2.4.2.4.a. Contingency table for data generated using INVITTOX Protocol No. 71 with recovery time-point and the PM featured in table 6.2.4.1.1. -EU classification system

Test Prediction	EU classification			
	NC	R36	R41	
NC	1	0	0	
R36	1	0	3	
R41	1	0	7	
Total	3	0	10	13

In addition to the data presented in table 6.2.4.2.4.a., there were four chemicals that were Not Classified according to the *in vivo* data that were predicted to be either R36 or R41 irritants based on the FL assay results, these were; 1,2-propanediol, DMSO, methanol and brij 35. Allyl alcohol was a R36 irritant and 1-butanol was a R41 irritant that were predicted to be either a R36 or a R41 irritant. As the amount of recovery or deterioration at the 72h time-point was not 30% different to the amount of FL at the original measurement, the PM could not distinguish if the chemicals were R41 or R36 irritants. These results were not included in the calculations reported in table 6.2.4.2.5.a. For the definitive classifications featured in table 6.2.4.2.4.a., the protocol and PM predicted all of the R41 irritants as irritants but only one of three Not Classified chemicals were correctly predicted.

There were fewer chemicals with GHS *in vivo* classifications in comparison to the number of chemicals with EU classifications; there were no GHS classifications for allyl alcohol, benzethonium chloride, brij 35, DMSO, mercuric chloride and methanol. On the contrary, triton X-100 (5%) which did not have an EU irritancy classification was included in this data set.

Table 6.2.4.2.4.b. Contingency table for data generated using INVITTOX Protocol No. 71 with recovery time-point and the PM featured in table 6.2.4.1.1. -GHS classification system.

Test Prediction	GHS classification			
	NC	Cat 2	Cat 1	
NC	1	0	0	
Cat 2	1	0	2	
Cat 1	1	2	5	
Total	3	2	7	12

In addition to the data presented in table 6.2.4.2.4.b., there was also one No Category chemicals according to the *in vivo* data, which was predicted to be either Category 1 or Category 2 irritant; 1,2-propanediol. 1-Butanol is a Category 1 chemical that was predicted to be either a Category 1 or Category 2 irritant. The exact classifications could not be determined as the recovery rate did not differ by 30% respective to the amount of FL first measured immediately after the chemical exposure. These results were not included in table 6.2.4.2.4.b. and the calculations reported in tables 6.2.4.2.5.b and 6.2.4.2.6. The predictive capacity of the FL assay and PM for the GHS classification system was comparable to the predictive capacity for the EU classification system.

There were fewer EPA classifications in comparison to the number of EU and GHS classifications. Based on the *in vivo* data, the following chemicals did not have EPA classifications; 1-butanol, allyl alcohol, benzethonium chloride, brij 35, cetylpyridinium bromide 6%, DMSO, mercuric chloride and methanol.

Table 6.2.4.2.4.c. Contingency table for data generated using INVITTOX Protocol No. 71 with recovery time-point and the PM featured in table 6.2.4.1.1. -EPA classification system.

Test Prediction	EPA classification				
	Cat IV	Cat III	Cat II	Cat I	
Cat III/ IV	0	1	0	0	
Cat II	0	1	0	2	
Cat I	0	2	1	4	
Total	0	4	1	6	11

In addition to the data presented in table 6.2.4.2.4.c., there was also one Category IV chemicals (1,2-propanediol), according to the *in vivo* data, that was predicted to be either a Category I or Category II irritant. This result was not included in the calculations reported in table 6.2.4.2.5.c for analyses of the predictive capacity for 'non-irritants versus the rest.'

The predictive capacity of INVITTOX Protocol No. 71 with recovery data, for the EPA classifications was very similar to the predictive capacity for the EU and GHS classification systems.

Table 6.2.4.2.5.a Evaluation of the performance of INVITTOX Protocol No. 71 with recovery data for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest

Data Source	Anal.	No.*	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Shaw et al., (1991), Ward et al., (1997a)	Refer to table 6.2.4.1.1.	20	84.6	11/13	100	10/10	33.3	1/3	83.3	10/12	100	1/1	66.7	2/3	0	0/10

Table 6.2.4.2.5.b. Evaluation of the performance of INVITTOX Protocol No. 71 with recovery data for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest

Data Source	Anal.	No.*	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Shaw et al., (1991), Ward et al., (1997a)	Refer to table 6.2.4.1.1.	20	83.3	10/12	100	9/9	33.3	1/3	81.8	9/11	100	1/4	66.7	2/3	0	0/9

Table 6.2.4.2.5.c. Evaluation of the performance of INVITTOX Protocol No. 71 with recovery data for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest

Data Source	Anal.	No.*	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Shaw et al., (1991), Ward et al., (1997a)	Refer to table 6.2.4.1.1.	20	72.7	8/11	100	7/7	25	1/4	70	7/10	100	1/1	75	3/4	0	0/7

* refers to the total number of chemicals included in all the analyses (refer to Annex Viii) NB. As the PM was not able to distinguish EPA Category III and Category IV materials, materials with these classifications were considered in the analyses concerning non-irritants.

As more than 10% of the test materials were classified as mild irritants according to the GHS classifications based on the *in vivo* data, an analysis of ‘severe irritants versus the rest’ was performed for this classification system (table 6.2.4.2.6.).

Table 6.2.4.2.6. Evaluation of the performance of INVITTOX Protocol No. 71 with recovery data for predicting ocular irritation according to the GHS classification system –Severe irritants versus the rest

Data Source	Anal.	No.*	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.		
Shaw et al., (1991), Ward et al., (1997a)	Refer to table 6.2.4.1.4	20	58.3	7/12	71.4	5/7	40	2/5	62.5	5/8	40	2/5	60	3/5	28.6	2/7

* refers to the total number of chemicals included in all the analyses (refer to Annex Viii).

INVITTOX Protocol No. 120

The Company # 4 assay protocol was tested in the COLIPA study (Brantom *et al.*, 1997). This protocol was slightly modified and later accepted as INVITTOX Protocol No. 120. In the COLIPA study the protocol was used to test only surfactants and surfactant-based formulations that were soluble in HBSS. The PM used results from the FL assay performed 4h following the material exposure. The assay was tested in two laboratories (FAL and Company # 4) and the mean results from each of the laboratories were used to determine the predictive capacity of this protocol. As the laboratories sometimes differed in their judgement as to whether a certain material was soluble in HBSS, not all materials were tested in each laboratory; twenty-nine identical materials were tested in both laboratories.

Raw *in vivo* data were entered into the ECVAM template v6 to generate the EU, GHS and EPA classifications

Table 6.2.4.2.7.a. Contingency table for data generated using INVITTOX Protocol No. 120 and the PM featured in table 6.2.4.1.6. –EU classification system

Test Prediction	EU classification			
	NC	R36	R41	
NC	19	0	0	
R36	5	1	0	
R41	6	3	27	
Total	30	4	27	61

Based on the *in vivo* data, there was no EU classification, for triton X-100 tested at 5%

INVITTOX Protocol No. 120 had a tendency to over-predict the irritancy potential of the test materials. Nearly 40% of the Not Classified materials were misclassified as either R36 or R41 irritants. More chemicals were over-predicted than formulations, as three of 11 different chemicals (including different chemical concentrations) had their irritancy potential over-predicted in comparison to four of 22 formulations that had their irritancy potential over-predicted.

INVITTOX Protocol No. 120 and the PM only classified one R36 material correctly whilst three were misclassified as R41 irritants. All R36 test materials were chemicals with the exception of the hand cleanser which was classified correctly as an R36 irritant by the FAL but misclassified by Company # 4 as a R41 irritant. This was the only test material which produced different classifications in the two laboratories.

The results in table 6.2.4.2.7.b show the predictive capacity of the protocol for GHS classifications.

Table 6.2.4.2.7.b. Contingency table for data generated using INVITTOX Protocol No. 120 and the PM featured in table 6.2.4.1.6. –GHS classification system

Test Prediction	GHS classification			
	NC	Cat 2	Cat 1	
NC	17	2	0	
Cat 2	5	1	0	
Cat 1	3	8	27	
Total	25	11	27	63

The predictive capacity of the protocol for the GHS classifications was similar to that for the EU classifications. The number of No Category materials misclassified as irritants, and the number of Category 2 irritants misclassified as Category 1 irritants indicated that the assay and PM had a tendency to over-predict irritancy. There were no Category 1 test materials under-predicted which supported the finding that the assay over-predicted rather than under-predicted potential irritancy. The Category 2 irritants were all misclassified with the exception of the hand cleanser formulation tested at the FAL.

Table 6.2.4.2.7.c. Contingency table for data generated using INVITTOX Protocol No. 120 and the PM featured in table 6.2.4.1.6. –EPA classification system

Test Prediction	EPA classification				
	Cat IV	Cat III	Cat II	Cat I	
Cat III/ IV	15	4	0	0	
Cat II	1	4	1	0	
Cat I	0	5	8	23	
Total	16	13	9	23	61

The concordance of the predicted and actual *in vivo* classifications were similar for all three classification systems.

The data featured in tables 6.2.4.2.7.a-c. were summarised to determine the predictive capacity of the protocol and PM for ‘non-irritants versus the rest’ (table 6.2.4.2.8.a-c.) and ‘severe irritants versus the rest’ (table 6.2.4.2.9.a-b.) for the EU, GHS and EPA classification systems

Table 6.2.4.2.8.a Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the EU classification system –Non-irritant versus the rest

Data Source	Anal.	No.*	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Brantom <i>et al.</i> , (1997)	Refer to table 6.2.4.1 .6.	33	82.0	50/61	100	31/31	63.3	19/30	73.8	31/42	100	19/19	36.7	11/30	0	0/31

Table 6.2.4.2.8.b. Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the GHS classification system –Non-irritant versus the rest

Data Source	Anal.	No.*	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Brantom <i>et al.</i> , (1997)	Refer to table 6.2.4.1 .6.	33	84.1	53/63	94.7	36/38	68	17/25	81.8	36/44	89.5	17/19	32	8/25	5.3	2/38

Table 6.2.4.2.8.c. Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the EPA classification system –Non-irritant versus the rest

Data Source	Anal.	No.*	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Brantom <i>et al.</i> , (1997)	Refer to table 6.2.4.1 .6.	33	83.6	51/61	100	32/32	65.5	19/29	76.2	32/42	100	19/19	34.5	10/29	0	0/32

* 33 different materials were tested (33 by Company # 4 and 30 by FAL) NB. As the PM was not able to distinguish EPA Category III and Category IV materials, materials with these classifications were considered in the analyses concerning non-irritants.

As more than 10% of the test materials were classified as mild GHS and EPA irritants according to the *in vivo* data, analyses of 'severe irritants versus the rest' were performed for these classification systems (tables 6.2.4.2.9.a-b.).

Table 6.2.4.2.9.a. Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the GHS classification system –Severe irritants versus the rest

Data Source	Anal.	No.*	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Brantom <i>et al.</i> , (1997)	Refer to table 6.2.4.1 .6.	33	82.5	52/63	100	27/27	69.4	25/36	71.1	27/38	69.4	25/36	30.6	11/36	0	0/27

Table 6.2.4.2.9.b. Evaluation of the performance of INVITTOX Protocol No. 120 for predicting ocular irritation according to the EPA classification system –Severe irritants versus the rest

Data Source	Anal.	No.*	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Brantom <i>et al.</i> , (1997)	Refer to table 6.2.4.1 .6.	33	78.7	48/61	100	23/23	65.8	25/38	63.9	23/36	65.8	25/38	34.2	13/38	0	0/23

* 33 different materials were tested (33 by Company # 4 and 30 by FAL) NB. As the PM was not able to distinguish EPA Category III and Category IV materials, materials with these classifications were considered in the analyses concerning non-irritants.

INVITTOX Protocol No. 82: Fixed Dose FL assay

For this data set, the percentage of FL (FL%) induced by 50mg/ml of the test chemical was recorded. The PM (table 6.2.4.1.7.) as featured in the publication of Clothier *et al.*, (1994) was applied to distinguish irritants and non-irritants. It was stated that recovery and deterioration should be taken into account when distinguishing R36 and R41 irritants. As the amount of recovery or deterioration that should take place in order to be considered as an effect was not stated in the publication, it was decided for the purpose of this BRD that a 30% difference in relation to the original FL assay data would be used to indicate significant differences. In the Clothier *et al.*, (1994) publication, the classifications for the test chemicals were based on a literature search of historical Draize data carried-out in the early 1990's,. These classifications were used for the analyses below as it was assumed that the correct *in vivo* data were matched to the chemical tested *in vitro*, i.e. the same chemical purity, concentration, CAS etc (Annex Vⁱⁱ). Further analyses to determine the predictive capacity of the assay for the GHS and EPA classifications were not performed as these classifications were not readily available and the suitability of the paired *in vitro* and *in vivo* data would be questionable.

Table 6.2.4.2.10. Contingency table for data generated using INVITTOX Protocol No. 82 and the PM featured in table 6.2.4.1.7. –EU classification system

Test Prediction	EU classification			
	NC	R36	R41	
NC	9	0	1	
R36	0	0	0	
R41	1	2	9	
Total	10	2	10	22

The protocol and PM performed well and allowed the majority of Not Classified chemicals and irritant chemicals to be classified correctly. There were only two chemicals classified as R36 irritants which entailed that the predictive capacity of this protocol and PM for mild irritants was not sufficiently tested; 2-methoxyethanol and chloroform were misclassified as R41 irritants.

The data in table 6.2.4.2.10. were analysed for the predictive capacity of 'non-irritants versus the rest,' and the results presented in the summary table 6.2.4.2.11.

Table 6.2.4.2.11. Evaluation of the performance of INVITTOX Protocol No. 82 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Clothier <i>et al.</i> , (1994)	Refer to table 6.2.4.1.7	22	90.9	20/22	91.7	11/12	90	9/10	91.7	11/12	90	9/10	10	1/10	8.3	1/12

As fewer than 10% of the test materials were classified as mild irritants according to the *in vivo* data, the analysis of ‘severe irritants versus the rest’ was not carried-out for this data set.

INVITTOX Protocol No. 71–Surfactants only

Data from the EC/HO study generated using INVITTOX Protocol No. 71 were analysed to determine the predictive capacity of the assay and PM for testing surfactants only. There were no INVITTOX Protocol No. 71 data available for surfactant-based formulations. The chemicals that were tested as part of the EC/HO study were tested in four independent laboratories and therefore the results per chemical presented in tables 6.2.4.2.12.a-c. (excluding greater-than values and where results for chemicals were not reported) were in quadruplicate.

Table 6.2.4.2.12.a. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –EU classification system

Test Prediction	EU classification			
	NC	R36	R41	
NC	10	0	5	
R36	2	0	6	
R41	3	0	11	
Total	15	0	22	37

In addition to the results presented in table 6.2.4.2.12.a. there were three incidences of R41 chemicals which received mixed Not Classified/R36 classifications and seven incidences where no classification could be assigned due to the low greater-than value reported.

The predictive capacity for ‘surfactants only’ did not differ greatly to the predictive capacity for when all the EC/HO test chemicals were analysed. For each surfactant, the classifications produced by each laboratory tended to agree and when there was discordance they differed by only one classification. All the incidences of Not Classified surfactants misclassified as R41 surfactants were for SDS tested at 3%; one laboratory also misclassified SDS tested at 3% as an R36 irritant. There were no surfactants classified as R36 irritants which prevented the predictive capacity of this protocol and PM for mildly irritating surfactants from being evaluated. All five incidences of R41 chemicals that were misclassified as Not Classified chemicals were reported by Laboratory 20. However, a total of three incidences of chemicals that were reported to be Not Classified/R36 irritants and seven incidences of chemicals that could not be given a predicted classification came from Laboratories 19 and 21. The FAL participated in the EC/HO study and stated that the format of the data submitted indicated that not all participating laboratories followed the protocol exactly.

Table 6.2.4.2.12.b. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –GHS classification system

Test Prediction	GHS classification			
	NC	Cat 2	Cat 1	
NC	10	0	5	
Cat 2	2	4	2	
Cat 1	3	0	11	
Total	15	4	18	37

In addition to the results presented in table 6.2.4.2.12.b. there was one No Category chemical which could not be given a classification based on the result reported by one laboratory, three incidences of R41 chemicals that were classified as No Category/Category 2, and seven incidences of R41 chemicals where no classification could be assigned due to the low greater-than value reported.

The predictive capacity for the GHS classification system was very similar to the EU classification system. The four incidences of Category 2 surfactants that were correctly classified by the protocol were all for triton X-100 tested at 5%. For the EU classifications, the 'study criteria were not met' for triton X-100 tested at 5%.

Table 6.2.4.2.12.c. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –EPA classification system

Test Prediction	EPA classification				
	Cat IV	Cat III	Cat II	Cat I	
Cat III/ IV	3	7	0	4	
Cat II	0	6	2	0	
Cat I	0	3	2	8	
Total	3	16	4	12	35

For the EPA classification system, the 'study criteria were not met' for cetylpyridinium bromide tested at 6%. Additionally, there was one Category IV chemical for which a classification could not be predicted by Laboratory 21, six incidences of Category I chemicals that could not be given a predicted classification and two incidences of Category I chemicals that were predicted to be either Category II, Category III or Category IV chemicals (i.e. the classifications could not be distinguished)

The predictive capacity of the protocol and PM for the EPA classification system was similar to that for the GHS classification system. The protocol and PM showed a similar rate of over-predicting the non-irritants and under-predicting the severe irritants for all three classification systems. The only material classified as a Category II irritant, according to the *in vivo* data was triton X-100 tested at 10%. Triton X-100 (10%) was classified as a R41 and a Category 1 irritant in the EU and GHS classification systems respectively.

The data in tables 6.2.4.2.12.a-c. were analysed and the results presented in the summary tables 6.2.4.2.13.a-c. for analyses of 'non-irritants versus the rest'. As fewer than 10% of the test materials were classified as mild irritants according to the *in vivo* data, the analysis of 'severe irritants versus the rest' was not carried-out for this data set.

Table 6.2.4.2.13.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.		
Balls <i>et al.</i> , 1995	Refer to table 6.2.4.1.5	37	73.0	27/37	77.3	17/22	66.7	10/15	77.3	17/22	66.7	10/15	33.3	5/15	22.7	5/22

Table 6.2.4.2.13.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.		
Balls <i>et al.</i> , 1995	Refer to table 6.2.4.1.5	37	73.0	27/37	77.3	17/22	66.7	10/15	77.3	17/22	66.7	10/15	33.3	5/15	22.7	5/22

Table 6.2.4.2.13.c. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.		
Balls <i>et al.</i> , 1995	Refer to table 6.2.4.1.5	37	62.9	22/35	75.0	12/16	52.6	10/19	57.1	12/21	71.4	10/14	47.4	9/19	25.0	4/16

NB. As the PM was not able to distinguish EPA Category III and Category IV materials, materials with these classifications were considered in the analyses concerning non-irritants. Refer to Annex Vaai for the origins of the *in vitro* data and *in vivo* classifications.

INVITTOX Protocol No. 71–Alcohols only

All of the data included in this analysis were generated from the EC/HO study and therefore results from four laboratories for each chemical were presented in table 6.2.4.2.14.a-c (excluding greater-than values and where results for chemicals were not reported). Data for the exact same chemicals were analysed for the predictive capacity of the protocol and PM for all three classification systems. Of all the data sets analysed in this section, this data set for alcohols contained the highest proportion of mild irritants. Materials with this level of potency were known to be difficult to classify correctly. Therefore one would expect the concordance of predicted and actual classifications to be lower for this data set in comparison to those containing few chemicals classified as mild irritants.

Table 6.2.4.2.14.a. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –EU classification system

Test Prediction	EU classification			
	NC	R36	R41	
NC	7	6	5	
R36	7	3	2	
R41	1	2	0	
Total	15	11	7	33

In addition to the data presented in table 6.2.4.2.14.a. there was one incidence of a Not Classified chemical which was classified as a No Category/R36, one incidence of a R36 chemical which was classified as a No Category/R36 and one incidence of a R41 chemical which was classified as a No Category/R36.

The results showed that this FL assay protocol had a poor predictive capacity for alcohols classified according to the EU classification system. More than 50% of the Not Classified, R36 and R41 chemicals were classified incorrectly. In comparison to the analyses for surfactants, there was a higher incidence of discordance between the predicted classifications from each laboratory.

Table 6.2.4.2.14.b. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –GHS classification system

Test Prediction	GHS classification			
	NC	Cat 2	Cat 1	
NC	4	9	5	
Cat 2	0	10	2	
Cat 1	0	3	0	
Total	4	22	7	33

In addition to the data presented in table 6.2.4.2.14.b. there were two incidences of Category 2 chemicals which were classified as a No Category/Category 2 chemicals and one incidence of a Category 1 chemical which was classified as a No Category/Category 2 chemicals.

The predictive capacity of INVITTOX Protocol No. 71 for the GHS classification system was slightly better than for the EU classification system. Eight of the 16 Not Classified alcohols in the EU system were classified as Category 2 irritants based the *in vivo* data and subsequently were not misclassified by the assay and PM for GHS classifications. The predictive capacity for the Category 1 irritants was comparable to the predictive capacity for EU R41 irritants.

Table 6.2.4.2.14.c. Contingency table for data generated using INVITTOX Protocol No. 71 and the PM featured in table 6.2.4.1.5. –EPA classification system

Test Prediction	EPA classification				
	Cat IV	Cat III	Cat II	Cat I	
Cat III/ IV	4	2	6	5	
Cat II	0	6	4	2	
Cat I	0	0	3	0	
Total	4	8	13	7	32

In addition to the data presented in table 6.2.4.2.14.c. there were two incidences of Category II chemicals which were classified as a Category IV/Category III/Category II chemicals and one incidence of a Category I chemical which was classified as a Category IV/Category III/Category II.

The predictive capacity of the protocol and PM for the EPA classifications showed greater similarity to the EU rather than the GHS classifications, i.e. a high number of irritants (EPA Cat I) were misclassified as Category II or Category III/IV chemicals.

The data in tables 6.2.4.2.14.a-c. were analysed and the results presented in the summary tables for 'non-irritants versus the rest' (6.2.4.2.15.a-c.) and 'severe irritants versus the rest' (table 6.2.4.2.16.a-c.).

Table 6.2.4.2.15.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5	33	42.4	14/33	38.9	7/18	46.7	7/15	46.7	7/15	38.9	7/18	53.3	8/15	61.1	11/18

Table 6.2.4.2.15.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5	33	57.6	19/33	51.7	15/29	100	4/4	100	15/15	22.2	4/18	0	0/4	48.3	14/29

Table 6.2.4.2.15.c. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EPA classification system –Non-irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5	33	46.9	15/32	45.0	9/20	50.0	6/12	60	9/15	35.3	6/17	50.0	6/12	55.0	11/20

NB. As the PM was not able to distinguish EPA Category III and Category IV materials, materials with these classifications were considered in the analyses concerning non-irritants. Refer to Annex Vaai for the origins of the *in vitro* data and *in vivo* classifications.

Table 6.2.4.2.16.a Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EU classification system –Severe irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5	33	69.7	23/33	0	0/7	88.5	23/26	0	0/3	76.7	23/30	100	3/3	23.3	7/30

Table 6.2.4.2.16.b. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the GHS classification system –Severe irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5	33	69.7	23/33	0	0/7	88.5	23/26	0	0/3	76.7	23/30	100	3/3	23.3	7/30

Table 6.2.4.2.16.c. Evaluation of the performance of INVITTOX Protocol No. 71 for predicting ocular irritation according to the EPA classification system –Severe irritants versus the rest

Data Source	Anal.	No.	Concordance		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Balls et al., 1995	Refer to table 6.2.4.1.5	33	68.8	22/32	0	0/7	88.0	22/25	0	0/3	75.9	22/29	100	3/3	24.1	7/29

NB. As the PM was not able to distinguish EPA Category III and Category IV materials, materials with these classifications were considered in the analyses concerning non-irritants. Refer to Annex Viii for the origins of the *in vitro* data and *in vivo* classifications.

6.2.4.3. Discussions

a. Description of the limitations of the test method (e.g., applicability domain based on results from compilation of data)

The results of the analyses conducted in section 6.2.4.2. were evaluated in an attempt to determine and compare the predictive capacities of the various FL assay protocols and PMs for which summarised *in vitro* and *in vivo* data were available.

The predictive capacities of the various FL assay protocols were evaluated in relation to *in vivo* data from regulated testing using rabbits' eyes. Currently, no database containing reliable data of chemical-induced *human* ocular irritation exists. Although rabbits' eyes are known to have differences in comparison to the human eye that potentially affect the results to a significant level, available human data tends to lack detailed exposure information, i.e. test material concentration, formulation composition, exposure duration. This information is required in order to accurately assess the predictive capacity of any *in vitro* assay. In the case of products that have been tested in humans' eyes, the irritancy potentials of these materials were generally known to be mild before testing. This tends to produce data sets which are biased towards the mild end on a scale of irritation. As the rabbits' eye is generally acknowledged to be more sensitive to ocular irritation than humans' eyes, one can state that the *in vivo* test using the rabbit provides an error safety margin, i.e. a material that is predicted to be an irritant by the Draize test is likely to be less irritant in the human eye.

Efforts were made to compare the predictive capacities of the various FL assay protocols whilst acknowledging the different amounts and quality of the *in vitro* and *in vivo* data. The largest data set analysed in section 6.2.4.2. was for INVITTOX Protocol No. 71 which featured in the EC/HO study (Balls *et al.*, 1995). There was no PM available for the type of data generated using this protocol, therefore a PM was created by the authors of this BRD, based on the EC/HO FL assay data and the *in vivo* based EU classifications provided by ECVAM. The PM was applicable for EU, GHS and EPA classifications. The EU, GHS and EPA classifications were generated by entering the raw *in vivo* data into the ECVAM template v6. As the PM was developed post-hoc, one could expect this protocol and PM to have a good predictive capacity, especially for the EU classifications. The PM was only applied to the definitive predicted classifications (i.e. not the mixed classifications generated by the greater-than values reported by the test laboratories). The data set contained information for a large number of chemicals with different potencies, but no data for formulations. Therefore the conclusions for the predictive capacity of INVITTOX Protocol No. 71 and the PM were limited to chemicals only. When all the chemicals were analysed, the predictive capacity was relatively low and concordance values ranged 62-70% for 'non-irritants versus the rest' analyses and 78-83% for 'severe irritants versus the rest' analyses according to the different classification systems. Specificity was consistently higher than sensitivity for each classification system for analyses of 'non-irritants versus the rest' and for analyses of 'severe irritants versus the rest.' Overall, these results indicated that the protocol had a greater predictive capacity for non-irritants for the EU, GHS and EPA classification systems and mild irritants for the EU and GHS classification systems, rather than severe irritants.

For the 'alcohols only' analyses, for analyses of 'non-irritants versus the rest,' concordance values were 58% and 47% for the GHS and EPA classifications

respectively and lower for the EU classification system (42%). For the alcohols, specificity was greater than sensitivity for all three classification systems for analyses of 'non-irritants versus the rest.' For analyses of 'severe irritants versus the rest' the concordance values were lower than those calculated for when all chemicals were analysed. For 'alcohols only,' sensitivity was 0% and specificity $\geq 88\%$ for all three classification systems. These findings indicated that for alcohols, INVITTOX Protocol No. 71 had a better predictive capacity for non-irritants and mild irritants rather than severe irritants, as was the finding for the analysis for the entire data set.

For all three classification systems, the concordance values for 'alcohols only' were lower than those for 'surfactants only' and when the entire set of EC/HO chemicals was evaluated. Approximately 25% of the EC/HO test chemicals were surfactants. In relation to analyses for all test chemicals, when the predictive capacity of INVITTOX Protocol No. 71 for 'surfactants only' was analysed, the concordance values showed slight increases for the EU and GHS classification systems and a very slight increase for the EPA classification system for analyses of 'non-irritants versus the rest.' Additional calculations for 'surfactants only' showed increased sensitivity and decreased specificity in comparison to the analysis for the entire data set. As the protocol uses a short incubation period, one would expect that the predictive capacity for surfactants would be greater than for the larger set of chemicals which would have also included chemicals that required a longer exposure period to induce an effect. As there were more surfactants tested in comparison to the number of alcohols, there was greater evidence for the predictive capacity of the protocol and PM for surfactants rather than for 'alcohols only.'

The concordance of predicted and actual EU classifications was used to determine the predictive capacity of INVITTOX Protocol No. 71 for solids. Of the sixty test chemicals, 20 were tested in a solid form. Thiorea was tested as a solid and *in vivo* ocular irritation data for this chemical were not reported as the test animals died (Balls *et al.*, 1995). Analyses for the effect of solids on the predictive capacity of the protocol were carried out using EU Risk Phrase classifications. The study criteria were not for *in vivo*-based classifications to be assigned to the following chemicals; chlorhexidine, fomesafen, L-aspartic acid, maneb, quiniacrine, sodium oxalate, potassium cyanate. Of the remaining 12 solid test materials the only chemical classified as a Not Classified, tetraaminopyrimidine sulphate had its potential irritancy predicted to be either a Not Classified/R36 irritant. Only three R36 chemicals were tested; 4-carboxybenzaldehyde was predicted to be either a Not Classified/R36 irritant by all four test laboratories, ammonium nitrate was assigned the correct classification by one laboratory but under-predicted by the remaining three laboratories and dibenzyl phosphate could not be classified by one lab, was classified as either a Not Classified/R36 irritant and as a R41 by another laboratory. There were eight R41 test chemicals of which one had the correct irritancy classification assigned by all four test laboratories. The following chemicals had their correct irritancy classification assigned by one or more test laboratories; benzoyl-L-tartaric acid (one laboratory), imidazole (one laboratory), promethazine HCl (three laboratories), sodium perborate (one laboratory). The remaining test chemicals were under-predicted to be either Not Classified, Not Classified/R36, R36 or a classification could not be assigned due to the low greater-than value. Overall, the data indicated that INVITTOX Protocol No. 71 and the PM had a greater predictive capacity for the Not Classified and R36 chemicals rather than the R41 chemicals tested as solids. This finding is in accordance with the predictive capacity of the protocol and PM calculated for chemicals that were tested as liquids. Subsequently

the data did not support the hypothesis that solid test chemicals adversely affected the predictive capacity of INVITTOX Protocol No. 71 and the respective PM. Due to the relatively few solid test chemicals for which definitive predictive classifications based on INVITTOX Protocol No. 71 data and *in vivo*-based classifications were available, it is acknowledged that findings reported here can only be considered potentially indicative of the effect of solid test chemicals on the predictive capacity of this protocol and respective PM.

The concordance of predicted and actual EU classifications was used to determine the predictive capacity of the FL assay for coloured chemicals. There were nine materials tested as part of the EC/HO study that were known to be coloured. As there were too few test chemicals that had definitive classifications, observations are made for chemicals that had mixed classifications, i.e. Not Classified/R36. Two of these chemicals (quiniacrine, maneb) did not have EU classifications as the study criteria were not met. The two Not Classified chemicals (tetraaminopyrimidine sulphate, trichloroacetic acid 3%) were classified correctly or as Not Classified/R36 chemicals by the various laboratories. There were no R36 coloured chemicals according to the *in vivo* data. Of the five R41 chemicals, all had their irritancy under-predicted by all four laboratories with the following exceptions; imidazole (one laboratory) and promethazine HCl (three laboratories). There were too few coloured chemicals tested to definitively determine the predictive capacity of the assay for coloured chemicals although these findings did seem to indicate that the predictive capacity of INVITTOX Protocol No. 71 was not good for coloured chemicals. Many chemicals that were not coloured also had their irritancy over-predicted or under-predicted. In general, the results indicated that the ability of the protocol to test and predict the irritancy potential of coloured materials needs to be investigated further.

Overall, the EC/HO study data indicated that the predictive capacity of INVITTOX Protocol No. 71 varied according to the type of test material, i.e. the predictive capacity was better for surfactants rather than alcohols. For all classification systems, the protocol had a greater predictive capacity for the non-irritants rather than the irritants for analyses of the entire data set and 'alcohols only' and the contrary for 'surfactants only'. Irritant materials that are not detected are of greater concern than non-irritant materials that are misclassified as irritants. Further work is needed to increase both specificity and sensitivity of the protocol and/or PM. In addition, formulations also need to be tested with this protocol as it is known that different PMs or modifications to the testing protocol, from those used for chemicals, can sometimes be required when predicting the irritancy of formulations.

From the data collected for this section, both INVITTOX Protocol No. 71 and No. 120 produced results that spanned five orders of magnitude. An advantage of such a large range of values is that wide ranges can be used to distinguish different classes of irritation. The PM devised for INVITTOX Protocol No. 71 for this BRD used relatively wide ranges of values to distinguish three levels of irritancy. In comparison, the PM for INVITTOX Protocol No. 120 used smaller ranges of values to distinguish three classifications of irritancy; this entails that INVITTOX Protocol No. 120 reproducibility needs to be high.

INVITTOX Protocol No. 120 and the PM were only used to test surfactants and surfactant-based formulations that were soluble in HBSS. Therefore, only the predictive capacity of this protocol for this limited but relevant chemical class could be determined.

There was a higher incidence of chemicals that were misclassified in comparison to the number of misclassified formulations. This finding indicated that the predictive capacity of INVITTOX Protocol No. 71 for formulations could also differ to the predictive capacity calculated in this BRD for chemicals only. The predictive capacity of INVITTOX Protocol No. 120 for surfactants and surfactant-based formulations was greater than that for INVITTOX Protocol No. 71 for 'all chemicals' and 'surfactants only' when the concordance values for the EU, GHS and EPA classification systems were compared for analyses of 'non-irritant versus the rest.' In the absence of 'severe irritants versus the rest' analyses for INVITTOX Protocol No. 71 for surfactants only, it was considered useful to compare the available results, i.e. severe irritants versus the rest for INVITTOX Protocol No. 71 for all chemicals and INVITTOX Protocol No. 120 for surfactants only for the GHS and EPA classification systems. Comparisons of the GHS and EPA results for 'severe irritants versus the rest' analyses showed that the concordance values were very similar but higher for INVITTOX Protocol No. 120 rather than INVITTOX Protocol No. 71 for the GHS classifications and lower for the EPA classifications. Sensitivity was 100% for INVITTOX Protocol No. 120 and only approximately 45% for INVITTOX Protocol No. 71 for the GHS and EPA classification systems. Overall, these findings suggested that the predictive capacity of INVITTOX Protocol No. 120 was greater than INVITTOX Protocol No. 71 for surfactants. There are significant differences between the protocols including the test materials exposure duration and the time-point FL is measured. Zanvit *et al.*, (1999) stated that the COLIPA FL assay protocol (INVITTOX Protocol No. 120) is purposely performed four hours after the chemical exposure as the results for surfactants are more predictive of *in vivo* effects at this time-point rather than immediately following the exposure. Further testing is required to determine the predictive capacity of INVITTOX Protocol No. 120 for other types of materials.

It is to be noted that the findings reported above were based on different data sets, which varied in the amount and quality of *in vitro* and *in vivo* data. For example, there were approximately 100% more data analysed for INVITTOX Protocol No. 71 in comparison to INVITTOX Protocol No. 120. Both chemicals and formulations were included in the data set for INVITTOX Protocol No. 120 whereas only chemicals were tested using INVITTOX Protocol No. 71. Additionally, the PMs used with the various protocols were developed differently. The PM established for INVITTOX Protocol No. 71 was devised based on the EC/HO FL study data and the EU classifications provided by ECVAM, therefore one would expect the protocol and PM to have a good predictive capacity as the PM was challenged by the same data set used to define its thresholds. In comparison, a PM was devised for INVITTOX Protocol No. 120 before the testing of the COLIPA test materials. The PM was based on historical data supplied by Company # 4 for surfactants and surfactant-based formulations. This PM was adapted by the authors of this BRD so that the threshold values correlated to the three principal classes featured in the EU, GHS and EPA classification systems; the PM could not distinguish Category III and Category IV EPA classified test materials. The predictive capacity of INVITTOX Protocol No. 120 and the PM was good as it was used to measure the effects of the same class of materials it was devised to test. Taking into account these different factors, makes it difficult to judge which protocol and PM were challenged to a greater extent by the various data sets.

In comparison to the data analysed for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120, the quality and amount of data was less for INVITTOX Protocol No. 82. The *in vivo* based EU classifications used in the publication concerning this protocol were devised by considering historical *in vivo* data from a number of different sources.

These data were not available in a form compatible with the ECVAM template v6. Therefore the EU classifications featured in the publication (Clothier *et al.*, 1994) were used to determine the predictive capacity of this protocol and PM for this classification system. There were no analyses for the predictive capacity of the assay for the GHS and EPA classification systems. Twenty-one chemicals were tested at a concentration previously noted as being able to distinguish irritants and non-irritants. The authors stated that consideration of 72h recovery data enabled some R36 irritants to be distinguished from R41 irritants. Although the amount of recovery or deterioration at the 72h time-point was not stated in the publication of Clothier *et al.*, (1994), a 30% difference in 72h values respective to the initial FL assay values was used by the authors of this BRD to indicate a substantial difference for analyses for this BRD. Using this criterion, the concordance value was the highest recorded for any data set analysed in this BRD for analyses of 'non-irritants versus the rest.' Sensitivity was not as high as recorded for INVITTOX Protocol No. 120 but specificity was higher than that recorded for any other data set for analyses of 'non-irritants versus the rest.' Further work would be required to determine if the predictive capacity was as good for a larger set of test materials that also included formulations. As the data indicated that the recovery data increased the predictive capacity of this assay, it would be interesting to investigate if recovery data could increase the predictive capacity of other FL assay protocols.

A preliminary analysis of the effect of recovery data on the predictive capacity of INVITTOX Protocol No. 71 was carried out. Based on the results of INVITTOX Protocol No. 71 tested in the EC/HO study, a PM was proposed in the publication of Balls *et al.*, (1995) which distinguishes R36 and R41 irritants on the basis of recovery or deterioration 72h following the initial exposure. There was no data set which contained recovery FL assay data and accompanying *in vivo* data for ocular irritation classifications. Subsequently, *in vitro* data from the publications of Shaw *et al.*, (1991) and Ward *et al.*, (1997a) were matched to *in vivo* data taken from a number of different sources. Due to a lack of chemical information the suitability of the paired *in vitro* and *in vivo* data was unknown. Therefore, the findings of this analysis can only be interpreted as indicators of the protocol's and PMs predictive capacity. A wide range of chemical classes were tested although there were only a maximum of 20 data points for the EU classification system and fewer for the GHS and EPA classification systems (results for some chemicals were reported by both Shaw *et al.*, (1991) and Ward *et al.*, (1997a)). INVITTOX Protocol No. 71 results which led to mixed classifications (i.e. R36/R41) were not included in the data analyses. The concordance values ranged from 72.7 to 84.6% for the different classification systems for analyses of 'non-irritants versus the rest.' For analyses of 'non-irritants versus the rest' the concordance rates for all three classification systems were substantially better for the recovery data set in comparison to INVITTOX Protocol No. 71 without recovery data, although the recovery data set was smaller. To definitively determine the predictive capacity of this protocol and PM, both chemicals and formulations need to be tested for which there are existing high quality *in vivo* data.

Overall, there were certain types of data (i.e., chemical classes, chemicals, formulations) missing for all protocols which prevented the predictive capacity of each protocol from being fully determined. There was greater evidence for the predictive capacity of INVITTOX Protocol No. 71 as it had been used in more laboratories, to test a greater number of test materials belonging to a wider range of different chemical classes. In comparison, INVITTOX Protocol No. 120 had only been used to test surfactants and surfactant-based formulations known to be soluble in HBSS in two laboratories. The

predictive capacity for surfactants and surfactant-based formulations tested with INVITTOX Protocol No. 120 was greater than INVITTOX Protocol No. 71. The predictive capacity of INVITTOX Protocol No. 71 for formulations requires investigation. Modifications to the PM could potentially increase the predictive capacity of INVITTOX Protocol No. 71 for the chemicals. Insufficient high quality *in vitro* and *in vivo* data has been obtained for INVITTOX Protocol No. 82 to be able to determine its predictive capacity. INVITTOX Protocol No. 86 was analysed in section 6.1.4.

b. Possible rationale for differences observed

Protocol differences

There are some fundamental protocol differences which would have impacted on the predictive capacity calculated for the various FL assay protocols. In comparison to INVITTOX Protocol No. 71, INVITTOX Protocol No. 120 has a different MDCK cell strain, a longer chemical exposure, and the amount of FL is measured 4h after the chemical exposure. The cell strain used for INVITTOX Protocol No. 120 forms a tighter barrier than the cell strain used in INVITTOX Protocol No. 71, therefore a longer chemical exposure period is required for INVITTOX Protocol No. 120 for damage to the monolayer to be measured. It was possible that the 4h FL measurements used by INVITTOX Protocol No. 120 could have increased the predictive capacity of this protocol relative to INVITTOX Protocol No. 71. In addition the different PMs used in conjunction with the various protocols will have affected the predictive capacity. It should be noted that the PM for INVITTOX Protocol No. 120 was devised and used to test surfactants and surfactant-based formulations only.

Quality of *in vivo* data

The quality of the *in vitro* and *in vivo* data sets differed per protocol. The EC/HO study and the COLIPA study tested the same chemicals and formulations in the *in vitro* test and in the *in vivo* tests. In comparison, the analysis of INVITTOX Protocol No. 82 was carried out using EU classifications based on historical *in vivo* data from a number of different sources. It is assumed that those involved in analysing this protocol knew the suitability of the paired *in vivo* based classifications with the *in vitro* data although this is not documented in the publication of Clothier *et al.*, (1994). For the analysis of INVITTOX Protocol No. 71 with recovery data, the *in vitro* data were paired with *in vivo* classifications by the authors of this BRD and the suitability of such pairings were unknown as information regarding the CAS number and purity were not available for the *in vitro* data.

For complete data sets, INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120 differed as to which classification systems the protocols and PMs had the highest and lowest concordance rates. For all three classification systems, INVITTOX Protocol No. 71 and the respective PM had a higher predictive capacity for the non-irritants and mild irritants whilst INVITTOX Protocol No. 120 and the PM had a greater predictive capacity for the mild and severe irritants.

Number of test materials

For all analyses, the mean *in vitro* results/classifications produced by each laboratory were analysed. INVITTOX Protocol No. 71 was evaluated by testing 59 chemicals in four laboratories although only the definitive classifications were considered in the analyses. INVITTOX Protocol No. 120 was evaluated by testing 33 chemicals in two laboratories (not all the same chemicals were tested in each laboratory). For the other protocols tested; there were fewer *in vitro* data available, the *in vitro* data were produced in a single laboratory and the *in vivo* data were of a lower quality. INVITTOX Protocol No. 120 had a higher predictive capacity than INVITTOX Protocol No. 71 which could have been due to the fewer test materials. The amount of data often entails that the protocol has been more extensively tested by a greater range of different types of materials. However, the data set for INVITTOX Protocol No. 120 was purposely limited to only surfactants and surfactant-based formulations, therefore in this case the number of test materials was not related to the number of different types of materials tested.

Types of test materials

With the exception of INVITTOX Protocol No. 120, a wide range of chemicals was tested using the different protocols analysed in this section. INVITTOX Protocol No. 120 was used to test surfactants and surfactant-based formulations only. Of all the data sets, INVITTOX Protocol No. 71 was used to test the widest range of different chemicals, although no formulations were tested.

Potency range of test materials

The GHS *in vivo* based classifications of the data sets for INVITTOX Protocol No. 71 (EC/HO study) and INVITTOX Protocol No. 120 were compared for potency ranges as there were a greater number of GHS classifications in comparison to the number of EU and EPA classifications. For INVITTOX Protocol No. 120 there were a similar number of No Category and Category 1 test materials and fewer Category 2 irritants. For the classifications used in the analyses for INVITTOX Protocol No. 71, there were similar numbers of Category 1 irritants and No Category chemicals and more Category 2. This data does not represent the proportions of the different classifications for the entire data sets (not analysed due to mixed classifications). The number of No Category and Category 1 chemicals was similar for each protocol and INVITTOX Protocol No. 71 had a higher proportion of Category 2 chemicals in comparison to INVITTOX Protocol No. 120. Category 2 (mild irritants) are known to be difficult to detect using both *in vitro* and *in vivo* methods and consequently one could expect the predictive capacity calculated for INVITTOX Protocol No. 71 would be lower than that calculated for INVITTOX Protocol No. 120 which was assessed using a data set containing relatively fewer Category 2 chemicals.

The data sets for INVITTOX Protocol No. 82 and INVITTOX Protocol No. 71 with recovery data, consisted of significantly fewer data in comparison to those for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120. The ranges of potency varied between these smaller data sets. Although one could expect these different potency ranges to affect the predictive capacity calculated for these protocols, differences in the amount and the quality of the *in vivo* data were hypothesised to have affected the predictive capacity to a greater extent.

6.3.1. Attempt to combine the data using weight-of-evidence approaches

A weight-of-evidence analysis was predominately carried-out for the FL assay INVITTOX Protocols which had the greatest amounts of *in vitro* and *in vivo* data available, to enable the predictive capacity to be determined. The weight-of-evidence analysis applied was devised by the authors of this BRD and acknowledged to be a preliminary study.

INVITTOX Protocol No. 86

INVITTOX Protocol No. 86 was the only protocol that had raw *in vitro* and *in vivo* data available. In a weight-of-evidence approach, the results from this analysis would receive greater weighting due to the quality of the data, i.e. the same highly relevant formulations were tested in both the *in vitro* and *in vivo* tests and compositions were known. However, the data set was only comprised of the surfactant-based formulations tested in the CTFA study Phase III (Gettings *et al.*, 1996). Data were available for 23 out of 25 formulations that were tested in a single laboratory. This was considered a small data set and would receive little weighting based on this criterion. As no chemicals were tested, the predictive capacity for chemicals could not be ascertained. Therefore, the limited range of test materials adds little weight in evidence of the various applicability domains of this protocol.

The PM for INVITTOX Protocol No. 86 was only capable of distinguishing irritants from non-irritants. In addition, there was also a range of TEP assay values that could not be used to classify formulations as either irritants or non-irritants. For the formulations, the predictive capacity for the EU, GHS and EPA classifications differed but all showed a greater predictive capacity for the irritants rather than the non-irritants in all three classification systems. These findings contrasted to those reported in the publication of Gettings *et al.*, (1996) where the assay was compared to FHSA classifications. A post-hoc threshold value enabled 14/18 of the FHSA irritants and all of the non-irritants to be correctly identified. Although the *in vivo* classification systems have different criteria for labeling materials as irritants or non-irritants, as these two findings are contradictory, the literature does not give any further weighting to the results calculated for this BRD.

Preliminary analyses of the effects of certain physical properties on the predictive capacity of INVITTOX Protocol No. 86 were inconclusive regarding the affect of coloured materials, and indicated that viscosity did not adversely affect the protocol's predictive capacity.

In comparison to the analyses for other INVITTOX Protocols which allowed three irritancy classifications to be predicted, there was no evidence of the protocol's and PM's ability to distinguish mild and severe ocular irritants. Comparisons of the predictive capacity of INVITTOX Protocol No. 86 with other INVITTOX protocols analysed in this section showed that the concordance values for INVITTOX Protocol No. 86 for 'non-irritants versus the rest' was comparable to the respective values for INVITTOX Protocol No. 71 for 'all chemicals' and 'surfactants only' and lower than the respective values for INVITTOX Protocol No. 120. As INVITTOX Protocol No. 86 was only analysed for its ability to distinguish irritants from non-irritants and comparisons of predictive capacity were made with protocols that determined different levels of irritancy one would have expected the predictive capacity of INVITTOX Protocol No. 86 to be higher. Due to the different sizes and quality of the data sets analysed for the various protocols, there was relatively little weight to support the findings for INVITTOX Protocol No. 86.

INVITTOX Protocol No. 71

The largest data set available for analyses was for INVITTOX Protocol No. 71 which was comprised of chemicals with a wide range of mechanistic actions and potencies. Subsequently, in comparison to the smaller data sets available for the other protocols greater weighting was given to the results from these analyses. INVITTOX Protocol No. 71 data were available for 59 chemicals tested in four different laboratories. This data set did not contain any values for formulations so only the predictive capacity for chemicals could be determined. Only raw *in vivo* data were available for this study. Mean INVITTOX Protocol No. 71 data for each chemical tested in each laboratory were compared with EU, GHS and EPA classifications. For all three classification systems there was a greater predictive capacity for the irritants rather than the non-irritants.

There was inconclusive evidence to determine if solid materials reduced the predictive capacity of the protocol. The viscosity of the test materials was unknown and therefore the effect of viscosity on the predictive capacity of INVITTOX Protocol No. 71 could not be ascertained. Based on a limited amount of data for the very few coloured materials tested, the findings did seem to indicate that the predictive capacity of INVITTOX Protocol No. 71 for coloured chemicals was reduced. However, there were a large number of chemicals that were not coloured that also had their irritancy over- or under-predicted. There were too few coloured chemicals tested to give weighting to this analysis although the results indicated that the ability of the protocol to test and predict the irritancy potential of materials with these physicochemical properties needs to be investigated further. As there were more solid chemicals tested in relation to the number of coloured test chemicals, there was greater weight-of-evidence for the effect of solid materials on the predictive capacity of the protocol.

INVITTOX Protocol No. 71 with recovery data

Based on the results from the EC/HO study for INVITTOX Protocol No. 71, Balls *et al.*, (1995) proposed a PM that utilised recovery data to distinguish mild and severe irritants. The PM took into account the recovery or deterioration of effects 72h following the chemical exposure. Recovery data from two publications were utilised to generate predicted EU, GHS and EPA classifications. There were no *in vivo* data generated by these studies so *in vivo* classifications were taken from other studies. As the paired *in vitro* and *in vivo* classifications were based on chemical name and concentration only, the results from analyses of these data receive little weighting due to the quality of the 'paired' data. Additionally, little weighting was given to the results due to the size of the data set and the absence of data for formulations. For the EU, GHS and EPA classification systems, the protocol and PM had a good predictive capacity for the severe irritants (R41, Category 1 and Category I) rather than the non-irritants and there were too few mild irritants in the data set to comment on the predictive capacity for materials with this level of irritancy.

INVITTOX Protocol No. 120

INVITTOX Protocol No. 120 data were generated by testing COLIPA surfactants and surfactant-based formulations known to be soluble in HBSS. A modified version of the PM featured in the COLIPA study was used to generate predicted EU, GHS and EPA classifications based on INVITTOX Protocol No. 120 data. Raw *in vivo* data were used to generate the EU, GHS and EPA classifications. The quality of the data set was comparable to that used to analyse INVITTOX Protocol No. 71 but there were less than half the data available for the analysis of INVITTOX Protocol No. 120. The test

materials for INVITTOX Protocol No. 120 were limited to surfactants only, but both chemicals and formulations were tested. Overall, this data set was considered to have less weighting than the result for INVITTOX Protocol No. 71, but greater weighting than the results for the other INVITTOX Protocols analysed for this BRD.

The predictive capacity of INVITTOX Protocol No. 120 for the EU, GHS and EPA classifications did not vary greatly. For all three classification systems, INVITTOX Protocol No. 120 had a better predictive capacity for surfactants in comparison to INVITTOX Protocol No. 71 for 'surfactants only' for analyses of 'non-irritants versus the rest.' Both protocols had a higher predictive capacity for the mild and severe irritants rather than the non-irritants as determined by analyses of 'non-irritants versus the rest' although there were fewer R36 irritants in the INVITTOX Protocol No. 120 data set in comparison to the data set for INVITTOX Protocol No. 71. The data set for INVITTOX Protocol No. 120 was; larger than INVITTOX Protocol No. 71 for 'surfactants only', generated by two laboratories and comprised of data for both chemicals and formulations. The INVITTOX Protocol No. 71 data for 'surfactants only' was generated by four laboratories and consisted of fewer test surfactants which were limited to chemicals. Therefore, INVITTOX Protocol No. 120 had greater weighting in support of its better predictive capacity for surfactants than INVITTOX Protocol No. 71. The greater predictive capacity of INVITTOX Protocol No. 120 for surfactants was supported by the publication of Zanvit *et al.*, (1999). Results for the nine surfactants common to the COLIPA and the EC/HO study were compared to *in vivo* MMAS data by Pearson's correlation and Spearman's rank coefficients. INVITTOX Protocol No. 120 (COLIPA) was found to show a greater correlation to the *in vivo* data in comparison to INVITTOX Protocol No. 71 (EC/HO study protocol) (Zanvit *et al.*, 1999).

The results from the literature where surfactants or surfactant-based formulations were tested (Zanvit *et al.*, (1999), Cottin and Zanvit (1997), (Southee, 1998)), ascertained the predictive capacity of INVITTOX Protocol No. 120 by comparing the *in vitro* data with Draize MAS, MMAS scores or classifications. The results from these analyses supported the findings for the concordance analyses performed here; that the assay had a greater tendency to over-predict rather than under-predict the irritancy of surfactants. These findings from multiple studies support and give extra weighting to the findings regarding the predictive capacity of the assay for EU, GHS and EPA classifications.

There was insufficient information about the physical state of the test materials to determine if the predictive capacity of INVITTOX Protocol No. 120 was affected by materials that were coloured, solid, etc.

INVITTOX Protocol No. 82.

Based on concordance values for analyses of 'non-irritants versus the rest,' the predictive capacity of INVITTOX Protocol No. 82 was greater than for the other INVITTOX Protocols analysed in this section. EU classifications were reported in the literature and used for the concordance analyses featured in this BRD. Although the 21 test chemicals covered a range of different chemical classes, little weighting can be given to the results due to; the relatively few test chemicals, the absence of data for formulations, and the quality of the historical *in vivo* data.

Jones *et al.*, (2001) used a protocol similar to INVITTOX Protocol No 82 to test hair care formulations. The protocol had a tendency to over-predict rather than under-predict the irritancy potential of the test shampoos and conditioners. Little weighting can be given

to this finding as; there was no information regarding the formulations compositions, the data set was small, and the *in vivo* data used was not transparent. In view of the contradictory findings for these two studies featuring INVITTOX Protocol No 82, overall there was little evidence in support of the predictive capacity of this protocol.

In the following sections, additional factors were investigated in further detail to determine how the different data sets potentially affected the predictive capacity calculated for each protocol.

Potency of test materials

For each FL assay protocol assessed, the potency ranges of the test materials were similar, as the aim of most studies was to establish the predictive capacity of the FL assay protocols for test materials with mild or moderate ocular irritation potentials. Quantitative scores rather than qualitative classifications were used to compare the potency ranges. Greater weighting could be given to the findings for INVITTOX Protocol No. 71 that featured in the EC/HO study as the test materials covered the entire range of MMAS values equally. In comparison, the COLIPA study tested materials that covered the entire range of MMAS values although the majority had MMAS values below 50. The range of MAS values for INVITTOX Protocol No. 86 data was also limited to a similar range (e.g. 10-43 MAS). The Fixed Dose FL assay (INVITTOX protocol No. 82) was only evaluated in a single study where the majority of test chemicals were classified as non-irritants or as R41 irritants (Clothier *et al.*, 1994). Jones *et al.*, (2001) used a similar protocol but used an in-house qualitative classification scheme for classifying the formulations. Subsequently there was little weight in support of this protocol as there were very few materials with mild irritancy that the FL assay was designed to measure.

Mechanisms

The *in vitro* data collected for this BRD were predominately produced by testing chemicals that exert toxicity via membrane lysis, and to a lesser extent coagulation; some materials had both mechanisms of action. Saponification was a mechanism of toxicity for only a few test chemicals. There was no indication that any of the test materials specifically affected tight junctions. There are chemicals that are known to affect tight junction formation, i.e. Calphostin C, and this has been tested with human corneal cells and MDCK cells but preliminary experiments showed no direct effects on fluorescein leakage (Clothier and Limb, personal communication). If a test material was capable of damaging the tight junctions, this would tend to occur before membrane damage. FL assay results alone cannot distinguish damage to the tight junctions from damage to the cell membranes. Therefore, it would seem that appropriate materials, i.e. those that cause membrane damage, have been tested in the FL assay in order to determine its predictive capacity. If a cell viability assay was also used to test the materials, the difference between specific tight junction damage and cell membrane damage could be potentially ascertained (Clothier and Samson, 1997)

To conclude, there was insufficient data to equally analyse the four INVITTOX Protocols. The proposed PM for INVITTOX Protocol No. 71 performed relatively well for predicting EU, GHS and EPA classifications for ocular irritation. The predictive capacity for these classification systems agreed with the findings reported in the literature where the performance was measured against other types of *in vivo* data. This indicated that the PM defined specifically for this BRD and used for these analyses was appropriate. Overall, there was greater evidence for the predictive capacity of INVITTOX Protocol No. 71 in comparison to the other INVITTOX Protocols analysed in this BRD. The PM

applied to INVITTOX Protocol No. 120 data for the BRD analyses was a modified version of the one featured in the COLIPA study. The predictive capacity of INVITTOX Protocol No. 120 for the EU, GHS and EPA classification systems was not dissimilar to that reported in the literature where the predictive capacity was measured against other types of *in vivo* data. The predictive capacity of INVITTOX Protocol No. 120 was better than INVITTOX Protocol No. 71 when comparing data for surfactants and surfactant-based formulations. Although it can be stated that INVITTOX Protocol No. 71 was challenged by materials with a wider range of mechanisms, both protocols have been equally challenged by the materials they were designed to test, i.e. materials with membrane lysis as the mode of action. At the low concentrations often employed for formulations, these types of materials can cause disruption to the tight junctions; damage caused by mild irritants can be repaired within hours. Only INVITTOX Protocol No. 120 tested both chemicals and formulations and the protocol and PM appeared to be appropriate for both surfactants and surfactant-based formulations. INVITTOX Protocol No. 86 had high quality *in vitro* and *in vivo* data but the test materials were limited to surfactant-based formulations only. Comparisons with the results for the other INVITTOX Protocols showed that the concordance of INVITTOX Protocol No. 86 predicted classifications with the *in vivo* classifications was comparable to the other protocols for different data sets.

6.3.2. Modifications to the assay if designed today

The FAL investigated using a transfected human corneal cell line for the FL assay. Japanese human corneal epithelial cells transfected with the simian virus 40 (J-HCET) were used. Human corneal cell cultures were hypothesised to predict potential human *in vivo* ocular irritation more accurately than cell types originating from different tissue types and species, i.e. MDCK cells. A human corneal cell line is currently used at Gillette Medical Evaluation Laboratories (USA) for the FL assay although few clones of this cell line remain. Subsequently, investigations were conducted to determine the ability of the J-HCET cell line to grow on Nunc 0.45µm polycarbonate inserts and to form tight junctions that cause an impermeable monolayer to sodium-fluorescein dye. In comparison to the cell model used at Gillette Medical Evaluation Laboratories, an advantage of the J-HCET cell line is that it can be cultured for up to 400 passages (Araki-Sasaki *et al.*, 1995) and cells up to 100 passages have been genetically typed by the FAL (Wilkinson, 2006). In comparison, the Gillette Medical Evaluation Laboratories model only expresses properties similar to human corneal cells for up to approximately 20 passages. The J-HCET cell line can be cultured in serum-free medium when supplemented with plant-derived products. Although both cell lines are capable of growing in stratified layers, which model the *in vivo* cornea rather than a monolayer, the J-HCET cells do not form a impermeable layer suitable for the FL assay. Therefore, the J-HCET cell line was used as a monolayer for the FL assay; as is used with the MDCK cells. Further research is required to find a human corneal cell line capable of forming stratified layers which models the *in vivo* situation more closely as stocks of the Gillette Medical Evaluation Laboratories clone which does form stratified layers, are being depleted. Work would be required to interpret the effect of stratified layers on FL assay results as 3D skin models using keratinocytes grown in multiple layers have revealed that cells grown in stratified layers can affect FL even in the absence of tight junctions (personal communication, R Clothier). In addition, multiple layers can cause problems with the resazurin assay which has been found to be a valuable addition for interpreting FL assay results. The resazurin assay has difficulty to penetrate all of the layers and can be difficult to remove which causes inaccurate results. Further work is necessary to evaluate a 3D corneal model with the FL assay. Preliminary experiments have shown that if the calcium concentration is high enough the tight junctions are correctly formed.

- Annex IVa Information available on the chemicals tested to assess FL assay predictivity of *in vivo* eye irritation
- Annex IVb Information available on the products/formulations tested to assess FL assay predictivity of *in vivo* eye irritation
- Annex Vai CTFA Phase III data set containing predicted and actual EU, GHS and EPA classifications for each formulation tested using the TEP assay/ INVITTOX Protocol No. 86. Raw *in vitro* and *in vivo* data were available
- Annex Vaai Example of a data set containing predicted and actual EU, GHS and EPA classifications for each test material. Data sets contained summarised *in vitro* data. -Entire *in vitro* and *in vivo* Data sets provided on CD
- Annex Vb *In vivo* reference data as reported in the literature for each substance

- tested using the FL assay –Example given but all data provided on CD
- Annex Vc Example of the *in vivo* reference data as submitted by Company # 3 for 17 formulations -Entire *in vitro* and *in vivo* data sets provided on CD
- Annex VI Table presenting the studies where FL assay results were used to assess the assay's predictive capacity for *in vivo* ocular irritation
- Annex VII Raw *in vivo* data as entered into ECVAM v6 template -provided on CD
- Annex A CTFA Study Phase III formulation compositions (from draft HET-CAM BRD: Appendix C2 (ICCVAM/NICEATM, 2004))
- Annex B COLIPA Study test chemicals and formulations compositions (from COLIPA)
- Annex C Formulations compositions from Company # 3

7. Applicability Domain

The applicability domain of the FL assay and the various protocols was investigated. The FL assay was developed specifically to model loss of trans-epithelial impermeability of the conjunctiva and cornea (INVITTOX Protocol No. 71), which occurs during chemical-induced ocular irritation. The FL assay models ocular irritation caused by damage to the inter-cellular junctions and the cell membranes of the corneal and conjunctival epithelia. The FL assay is useful for testing mild to moderate irritant materials within a defined range where it can measure mechanistic damage to the adhesion molecules. Various FL assay protocols have featured in studies that evaluated their ability to model and predict *in vivo* ocular irritation. In order to test the predictive capacity of the FL assay protocols for *in vivo* ocular irritation, a range of relevant chemicals and formulations have been tested. These were predominately cosmetic ingredients and formulations that are the materials that are most likely to enter into the human eye as a result of frequent use.

The applicability domain(s) for the various FL assay protocols have been determined by evaluating the predictive capacity of various FL assay protocols for EU, GHS and EPA classifications based on *in vivo* data. The PMs were devised according to descriptions in Section 6 of this BRD. The predictive capacities of all the protocols discussed above did not vary greatly according to whether they were compared to EU, GHS or EPA classifications.

The amount, type, and quality of *in vivo* data used to determine the applicability domain of the different protocols varied for each protocol. For the larger data sets analysed, the predictive capacities of the protocols and PMs for the EU, GHS and EPA classification systems did not vary greatly. For the smaller data sets, a single classification affected to a greater extent the overall predictive capacity of the protocols for the three classification systems. The type and quality of the data for each protocol are discussed below. The applicability domains for the various protocols were then discussed in the final paragraphs of this section.

INVITTOX Protocol No. 71

The EC/HO study evaluated the predictive capacity of INVITTOX Protocol No. 71 by testing a wide range of pure chemicals (Balls *et al.*, 1995). The conclusions formed for INVITTOX Protocol No. 71 were based on the results for which definitive classifications were predicted (i.e., greater-than results which prevented definitive classifications from being assigned were not analysed). The results indicated that the predictive capacity of the assay varied for different types of chemicals and that the assay was more suited to predicting the effects of surfactants rather than alcohols. The predictive capacity of surfactants was similar to when data for all the test chemicals were analysed. No further analyses to determine the predictive capacity for different chemical classes were performed, as there were too few chemicals representative of each chemical class. The protocol and/or PM required further work to increase the overall predictive capacity, and for the non-irritants in particular, for the EU, GHS and EPA classification systems.

INVITTOX Protocol No. 71 with recovery data

Using a limited data set, the use of recovery data generated by INVITTOX Protocol No. 71 to predict potential ocular irritancy was investigated. This data set was of the lowest quality analysed due to the unknown suitability of the paired *in vitro* and *in vivo* data for the test chemicals. Despite this concern, the concordance rates of predicted and actual

EU, GHS and EPA classifications were higher than those for INVITTOX Protocol No. 71 without recovery data, but this was hypothesised to be due to the relatively smaller data set rather than a greater predictive capacity of the PM.

INVITTOX Protocol No. 82

The quality of INVITTOX Protocol No. 82 (Fixed Dose FL assay) data was low in comparison to the data sets used to assess the applicability domain of the other INVITTOX protocols. The modified PM was used to assess the predictive capacity of this assay for the test chemicals featured in the Clothier *et al.*, (1994) publication. The concordance value was the highest one recorded for any data set analysed in this BRD. Sensitivity was not as high as recorded for INVITTOX Protocol No. 120 but specificity was higher than that recorded for any other data set. This result was good in consideration of the nature of the *in vivo* data and that a number of different types of chemicals had been tested. Further work would be required to determine if the predictive capacity was as high for a larger set of test materials that also included formulations.

INVITTOX Protocol No. 86

INVITTOX Protocol No. 86 (TEP assay) was tested in the CTFA study Phase III and this was the only data set available to assess the predictive capacity of this protocol. As the test materials were all surfactant-based formulations, the applicability domain of this protocol can only be ascertained for these types of materials. Based on relatively few data, the predictive capacity of INVITTOX Protocol No. 86 displayed a greater predictive capacity for the irritants rather than the non-irritants.

INVITTOX Protocol No. 120

EU, GHS and EPA classifications based on raw *in vivo* data from the COLIPA study were matched against mean INVITTOX Protocol No. 120 values to allow the predictive capacity and applicability domain of this protocol to be ascertained. Surfactants and surfactant-based formulations were tested with this protocol; the applicability domain of this protocol for only these types of test materials could be ascertained. INVITTOX Protocol No. 120 had a better predictive capacity than INVITTOX Protocol No. 71 and INVITTOX Protocol No. 86 for surfactants and surfactant-based formulations. In comparison to INVITTOX Protocol No. 71, INVITTOX Protocol No. 120 has a longer chemical exposure and FL is measured 4h after the exposure rather than immediately. INVITTOX Protocol No. 86 has the same exposure duration as INVITTOX Protocol No. 120 but FL is measured immediately following the exposure and the EC₅₀ (%) is recorded. The longer exposure duration is used for INVITTOX Protocol No. 120 as the cell strain and inserts used cause a more impermeable monolayer that requires a longer chemical exposure for effects to be measured. Zanvit *et al.*, (1999) stated that FL is measured 4h following the chemical exposure to enable the effects of cationic surfactants to be measured correctly. It is stated that cationic surfactants cause the cells to stick together which affects the rate of FL measured immediately following the exposure; the 4h results were reported to more accurately reflect the irritancy of these test materials. This finding was supported by the greater predictive capacity of INVITTOX Protocol No. 120 in comparison to INVITTOX Protocol No. 86.

Potency of test materials

The potency of the materials used to assess the predictive capacity of the various FL assay protocols would have affected the predictive capacities calculated. The GHS classifications for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120 data sets

were compared as there were slightly more GHS classifications in comparison to the number of EU and EPA classifications. For both INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120 there were fewer Category 2 irritants in comparison to the number of No Category and Category 1 test materials; there were a similar number of No Category and Category 1 test materials. Therefore, both protocols were equally challenged by data sets with similar ranges of potency.

The data sets for INVITTOX Protocol No. 82 and INVITTOX Protocol No. 71 with recovery data, consisted of significantly fewer data in comparison to those for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120. The proportion of materials with mild irritancy according to the *in vivo* data was lower for these smaller data sets in comparison to the data sets for INVITTOX Protocol No. 71 and INVITTOX Protocol No. 120. Although one could expect the different potency ranges to affect the predictive capacity calculated for these protocols, differences in the quality of the *in vivo* data were more likely to have affected the predictive capacity to a greater extent.

Physical state of test materials

Various properties of test materials that were known to affect the ability of the assay to accurately measure induced toxicity were investigated to determine how they affected the predictive capacities calculated for the different FL assay protocols. Data for INVITTOX Protocol No. 71 were predominately used as this data set contained the highest number of chemicals for which detailed chemical information could be attained. There was insufficient information regarding the pH of the test materials to determine its effect on the predictive capacity of the FL assay.

-Solid materials

Data evaluations were performed to determine if the predictive capacity of INVITTOX Protocol No. 71 was affected by testing solid chemicals. Solids can be difficult to test as the concentration in contact with the cells cannot be assumed to be equal to that placed on the cell monolayer, nor uniformly distributed. Also, solids cannot be easily removed from the cell surface following the short exposure period (Balls *et al.*, 1995). The concordance of predicted and actual EU classifications did not support the hypothesis that the predictive capacity of the protocol was adversely affected by solid test chemicals, however it was acknowledged that the test set of chemicals contained relatively few solids.

-Viscous materials

Viscous materials can be difficult to remove from the monolayer following the short exposure period. Due to the short exposure period, mildly irritating materials often need to be tested neat in order to produce a response which can be measured. Therefore the problems associated with viscous materials cannot be reduced by dilution. The effect of viscosity on the predictive capacity of INVITTOX Protocol No. 71 could not be ascertained as the viscosity of the test materials was unknown. CTFA study Phase III data did not indicate that viscosity affected the predictive capacity of INVITTOX Protocol No. 86. In comparison to INVITTOX Protocol No. 71, this protocol has a longer chemical exposure and 5x more washing steps following the removal of the test material. One could hypothesise that the predictive capacity of INVITTOX Protocol No. 71 for viscous materials could be lower than that of INVITTOX Protocol No. 86 due to the shorter exposure duration which would be more greatly affected by any difficulties in removing the test chemical. This would impact on protocol reproducibility, and potentially predictive capacity.

-Coloured materials

There were nine EC/HO study test chemicals that were known to have colour. This was considered to be too few to determine the predictive capacity of INVITTOX Protocol No. 71 for chemicals with colour, although observations of the results did seem to indicate that the predictive capacity of the protocol was not good for coloured chemicals.

The results for INVITTOX Protocol No. 86 which tested 23 formulations, of which 19 were coloured, did not indicate that the predictive capacity of the protocol was affected by materials with colour. As the problems associated with measuring the effects of coloured materials would be similar for all protocols, based on this result no FL assay protocols should have their predictive capacities adversely affected by coloured materials if the test materials were fully removed following the exposure period. However, in comparison to other INVITTOX Protocols, INVITTOX Protocol No. 86 has more washing steps following the removal of the test material which could reduce the impact of coloured materials on the predictive capacity of this protocol in comparison to that for INVITTOX Protocol No. 71. Further work is required to determine the effects of the physical states of test materials on the predictive capacity of the various FL assay protocols.

-Solubility of test materials

The FL assay generally had a good predictive capacity for materials that are water soluble and/or the toxic effect is not affected by dilution. It is important that the basic toxic mechanism is not affected by dilution as the formation of micelles by surfactants can unpredictably alter cellular responses over a concentration range that impacts on the predictive ability of *in vitro* assays.

Test materials that are not soluble in HBSS or distilled water can only be reproducibly tested in the FL assay if they form a stable suspension or emulsion. Emulsions and suspensions will not be as homogeneous as a solubilised material and it is not easy to establish that an emulsion or suspension is uniform or stable. However, the short chemical exposure duration provides less time for the emulsion or suspension to degrade. Materials that are insoluble or do not form stable aqueous emulsions can be solubilised or suspended in mineral oil to reduce evaporation and ensure that the test concentrations remain in contact with the cells for the specified exposure period. The PM featured in INVITTOX Protocol No. 120 was only applied to data generated by testing surfactants and surfactant-based formulations that were soluble in HBSS. Although the effect of solid and viscous materials were not definitively determined for INVITTOX Protocol No. 71 it is possible that the predictive capacity of INVITTOX Protocol No. 71 for EU, GHS and EPA classifications could have been affected by unknown solubilities of the chemicals or the use of solvents other than HBSS. The EC/HO study published Pearson's correlation coefficients and Spearman's rank correlation coefficients for *in vitro* data compared with *in vivo* data for; the entire data set, the chemicals soluble in water, chemicals insoluble in water, surfactants, solids, liquids and solutions (Balls *et al.*, 1995). Due to extensive between-laboratory variation for the four participating laboratories, it was not possible to observe any definitive effects of these factors on the predictive capacity calculated, e.g., in comparison to the predictive capacities for the entire data set, the predictive capacity for only those chemicals that were soluble in water was better for one laboratory, only very slightly better for two laboratories and worse for another.

-Test materials binding to the insert membrane

Certain test materials can bind to the insert membrane, thus making their removal very difficult. Chemical binding to the insert membrane, is more common for cationic surfactants, such as benzalkonium chloride, which are attracted to the positively charged membrane (Balls *et al.*, 1995). Negatively charged cell surface proteins can also attract positively charged surfactants. Chemical binding to the insert membrane or cell surfaces, increases the chemical exposure duration but can also physically block the passage of sodium-fluorescein dye through the insert. Also, test materials with high molecular weights that are not fully removed, can physically block the passage of the sodium-fluorescein dye through the insert, which could cause chemical effects to be under-estimated. Ward *et al.*, (1997a) found that the different properties of the insert membranes affected the interactions with various test chemicals.

To counter the effects of coloured, viscous and solid materials and those that bind to the monolayer or insert membrane, on the predictive capacity of the FL assay, sufficient washing steps are needed to ensure the full removal of the test materials at the end of the exposure period. However, many washing steps increases the likelihood of damage caused to the monolayer and/or insert membrane, thus producing erroneous results. INVITTOX Protocols No. 71, 82 and 86 have one or two washing steps after the removal of the test materials whereas INVITTOX Protocol No. 86 has ten washing steps.

Types of test materials

The majority of data used to assess the predictive capacities of the various FL INVITTOX Protocols were generated by testing surfactants and surfactant-based formulations. Most cosmetic formulations and some ingredients are surfactant-based and surfactants respectively. Therefore, the various protocols have been evaluated by testing relevant materials, i.e. those that are humans are likely to be routinely exposed to, and which are potentially able to enter the eye and cause ocular irritation. Cosmetic formulations or ingredients at the concentrations intended for use tend to range from non-irritants to moderate irritants that cause membrane damage as their primary mechanism of irritancy/toxicity. The following chemical classes were proposed by the ECVAM Eye Irritation Meeting (2005) as having membrane lysis as their primary mechanism of eye irritation; surfactants, organic solvents, ketones, alcohols, volatile liquids, ethers, polyethers, esters, aromatics, amines. At lower concentrations than those causing membrane damage, these types of chemicals could also cause damage to the inter-cellular junctions. Therefore, the FL assay has the potential to measure the toxicity exerted by a range of materials, and to predict their *in vivo* ocular irritation effects. A number of chemicals belonging to these various chemical classes were tested using INVITTOX Protocol No. 71 as part of the EC/HO study, but there were too few in each class to determine the predictive capacity of this protocol for each chemical class. Further testing could determine the predictive capacity of the protocols for these different chemical classes. The choice of test chemicals representing these various chemical classes requires careful consideration in order to allow the effects of other physical properties such as viscosity and colour to be distinguished.

Conclusions

There was insufficient data to equally analyse the four INVITTOX Protocols. The largest high quality data sets (i.e. raw *in vivo* data used to calculate EU, GHS and EPA classifications) were available for INVITTOX Protocol No. 71, INVITTOX Protocol No. 120, and to a lesser extent INVITTOX Protocol No. 86. With the exception of INVITTOX

Protocol No. 71, all other FL INVITTOX protocols showed a greater predictive capacity for the irritants rather than the non-irritants. None of the protocols showed a significantly greater predictive capacity for any particular classification system. As the difference in predictive capacities for the various classification systems was slight, the protocols differed as to which classification system they had a slightly greater predictive capacity for. The predictive capacity of INVITTOX Protocol No. 86 for surfactant-based formulations was higher than that calculated for INVITTOX Protocol No. 71 for 'surfactants only.' INVITTOX Protocol No. 120 had the greatest predictive capacity for surfactants and surfactant-based formulations in comparison to INVITTOX Protocol No. 86 and INVITTOX Protocol No. 71 for 'surfactants only.' The greater predictive capacity was hypothesised to be due to the measurement of FL 4h after, rather than immediately after, the chemical exposure. INVITTOX Protocol No. 71 had a greater predictive capacity for the surfactants rather than the alcohols tested.

Preliminary analyses for the effect of coloured test materials on the predictive capacity of the protocols indicated that INVITTOX Protocol No. 86 may be less adversely affected than INVITTOX Protocol No. 71. This was hypothesised to be due to the greater number of washing steps following removal of the test material in INVITTOX Protocol No. 86, in comparison to INVITTOX Protocol No. 71. Increased washing steps in other INVITTOX Protocols could potentially enhance the removal of other types of test materials and increase the reproducibility and robustness of the protocols.

As a result of the findings regarding the predictive capacities of the INVITTOX Protocols analysed for this BRD, two recommendations can be made according to the intended use of the FL assay. Firstly, if the FL assay is to be used to test surfactants and surfactant-based formulations only, it is recommended that INVITTOX Protocol No. 120 is investigated further. If the FL assay is to be used to test a wider range of materials from many different chemical classes, further testing could be carried-out using INVITTOX Protocol No. 71 to increase the existing data set for a wide range of chemicals. As INVITTOX Protocol No. 120 measures FL 4h following the chemical exposure to increase the predictive capacity of the assay for cationic surfactants, further work could investigate the effect of FL measured at this time-point on the predictive capacity of INVITTOX Protocol No. 71. It would also be interesting to discover if modifications to the PM for INVITTOX Protocol No. 71 would increase the predictive capacity of this protocol. For any protocol examined in the future, further work is needed to determine the effects of physical properties of the test chemicals such as colour and viscosity on the predictive capacity.

The principal advantage of the FL assay is that it can measure recovery and delayed effects from the initial acute test material exposure for up to 96h on the same set of cells. Recovery from effects is an important element in the EU, GHS and EPA classification systems for ocular irritation. Use of the FL assay to measure recovery and delayed effects is only stated as part of INVITTOX Protocol No. 82 although other FL assay INVITTOX Protocols have also been used to measure recovery. However, there were no large data sets with high quality *in vitro* recovery data and corresponding *in vivo* data available for analyses for this BRD. Inclusion of recovery data for the classification of some types of chemicals could potentially increase the predictive capacity of the FL assay protocols. The time-point of recovery to be considered for the classification of test chemicals could vary according to chemical class (Cottin and Zanvit, 1997). The authors of this BRD suggest that further work could investigate the use of recovery data for increasing the predictive capacity of either INVITTOX Protocol No. 71 or INVITTOX

Protocol No. 120 which already have existing large data sets for immediate chemical-induced effects.

8. Supporting materials

8.1. Relevant publications, other scientific reports and review (in chronological order).

Refer to separate volume

8.2. Relevant unpublished data

9. References

- Araki, K., Ohahsi, Y., Kinoshita, S., Hayahsi, K., Kuwayama, Y., Tano, Y., 1994. Epithelial wound healing in the denervated cornea. *Current Eye Research*, 13: 203-211.
- Araki-Sasaki, K., Ohashi, Y., Sasabe, T., Hayashi, K., Watanabe, H., Tano, Y., Handa, H., 1995. An SV40-immortalised human corneal epithelial cell line and its characterisation. *Investigative Ophthalmology and Visual Science*, 36:614-621.
- Atkinson, K. A., 1993. The Evaluation and Validation of Alternative Methods for Ocular Irritation Testing. Ph.D. Thesis. University of Nottingham, UK.
- Bagley, D.M., Botham, P. A., Gardner, J. R., Holland, G., Kreiling, R., 1992. Eye irritation: reference chemicals data bank. *Toxicology In Vitro*, 6: 487-491.
- Baldwin, H.A., McDonald, T.O., Beasley, C.H., 1973. Slit examination of experimental animals eyes. II Grading scales and photographic evaluation of induced pathological conditions. *Journal of Society of Cosmetic Chemistry*, 24: 181-195.
- Balls, M., Clothier, R. H., 1992. Cytotoxicity assays for intrinsic toxicity and irritancy. In: *In Vitro Methods of Toxicology*. Watson, R.R. ed. Florida, CRC Press. pp. 38-52.
- Balls, M., Botham, P.A., Bruner, L.H., Spielmann, H., 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test. *Toxicology In Vitro*, 9(6): 871-929.
- Borenfreund, E., Borrero, O., 1984. *In vitro* cytotoxicity assays. Potential alternatives to the Draize ocular allergy test. *Cell Biology Toxicology*, 1(1):55-65.
- Botham, J. M., Mckillop, C. M., Purchase, I. F. H., 1989. Acute and topical toxicity profiles for substances involved in the in vitro validation of the in vivo ocular irritation model. Report for the commission of the European Communities.
- Botham, P., Osborne, R., Atkinson, K., Carr, G., Cottin, M., van Buskirk, R.G., 1997. IRAG working group 3. Cell function-based assays. Interagency Regulatory Alternatives Group. *Food and Chemical Toxicology*, 35(1):67-77.
- Brantom, P.G., Bruner, L.H., Chamberlain, M., Desilva, O., Dupuis, J., Earl, L.K., Lovell, D.P., Pape, W.J.W., Uttley, M., Bagley, D.M., Baker, F.W., Brachter, M., Courtellemont, P., Declercq, L., Freeman, S., Steiling, W., Walker, A.P., Carr, G.J., Dami, N., Thomas, G., Harbell, J., Jones, P.A., Pfannenbecker, U., Southee, J.A., Tcheng, M., Argembeaux, H., Castelli, D., Clothier, R., Esdaile, D.J., Itigaki, H., Jung, K., Kasai, Y., Kojima, H., Kristen, U., Larnicol, M., Lewis, R.W., Marenus, K., Moreno, O., Peterson, A., Rasmussen, E.S., Robles, C., Stern, M., 1997. A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. *Toxicology In Vitro*, 11(N1-2): 141-179.
- Butler J. M., Hammond B.R., 1980. The effects of sensory denervation on the response of the rabbit eye to prostaglandin E1, bradykinin and substance P. *British Journal of Pharmacology*, 69: 495 -502.

CEC., 1991. Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test. Doc. XI/632/91 - V/E/1/131/91. CEC, Brussels.

Chan, P.K., Hayes, A.W., 1994. Acute toxicity and eye irritancy. In: Principles and Methods of Toxicology. Hayes A. W., ed. 3rd Ed, Chapter 16. New York: Raven Press. pp 548-579.

Clothier, R.H., Morgan, S.J., Atkinson, K.A., Garle, M.J., Balls, M., 1994. Development of a fixed-dose approach for the fluorescein leakage test. *Toxicology In Vitro*, 8: 883-884.

Clothier, R. H., Morris, J., Lankford, W. T., 1995. The evaluation of pesticide ingredients and formulations in vitro and correlations with *in vivo* data. *ATLA*, 23: 667-675.

Clothier, R., Sansom, R., 1996. Effects of surfactant retreatment *in vitro*: a method to evaluate changes in cell junctions and in cell viability. *ATLA*, 24: 859-865.

Clothier, R.H., Starzec, G., Stipho, S., Kwong, Y.C., 1999. Assessment of initial damage and recovery following exposure of MDCK cells to an irritant. *Toxicology In Vitro*, 13:713-717.

Clothier, R., Starzec, G., Pradel, L., Baxter, V., Jones, M., Cox, H., Noble, L., 2002. The prediction of human skin responses by using the combined *in vitro* fluorescein leakage/Alamar Blue (resazurin) assay. *ATLA*, 30(5):493-504.

Cottin, M., Zanvit, A., Rougier, A., Dossou, K.G., 1992. Loss of trans-epithelial permeability: a key role in the ocular toxicity of surfactants? Poster presented at SPTC meeting, Paris.

Cottin, M., Zanvit, A., 1999. Fluorescein leakage test: A. useful tool in ocular safety assessment. *Toxicology In Vitro*, 11:399-405.

Courtellemont, P., Herbert, P., Biesse, J. P., Castelli, D., Friteau, L., Serrano, J., Robles, C., 1999. Relevance and reliability of the PREDISAFE assay in the COLIPA eye irritation validation program (phase 1). *Toxicology In Vitro*, 13: 305-312.

Curren, R. D., Harbell, J. W., 1998. In vitro alternatives for ocular irritation. *Environmental Health Perspectives*, 106 Suppl. 2: 485-92.

Dearman, R.J., Cumberbatch, M., Kimber, I., 2003. Cutaneous cytokine expression: Induction by chemical allergen and paracrine regulation. *Journal of Toxicology-Cutaneous and Ocular Toxicology*, 22: 69-86.

Draize, J. H., Woodard, G., Calvery, H. O., 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *The Journal of Pharmacology and Experimental Therapeutics*, 82: 377-390.

Dua, H. S, Gomes, J. A. P., Singh, A., 1994. Corneal epithelium wound healing. *British Journal of Ophthalmology*, 78: 401-408.

ECETOC, 1992. Eye Irritation: Reference Chemicals Data Bank. Technical Report No. 48. European Chemical Industry Ecology and Toxicology Centre, Brussels.

Eurell, T. E., Sinn, J. M., Gerding, P. A., Alden, C. L., 1991. In vitro evaluation of ocular irritants using corneal protein profiles. *Toxicology and Applied Pharmacology*, 108: 374-378.

Farquhar, M. G., Palade, G. E., 1963. Junctional complexes in various epithelia. *Journal of Cell Biology*, 17: 375-412.

Gautheron, P., Duprat, P., Hollander, C.F., 1994a. Investigations of the MDCK permeability assay as an *in vitro* test of ocular irritancy. *In Vitro Toxicology*, 7:33-43.

Gautheron, P., Giroux, J., Cottin, M., Audegond, L., Morilla, A., Mayordomo-Blanco, L., Tortajada, A., Haynes, G., Vericat, J.A., Pirovano, R., Gillio Tos, E., Hagemann, C., Vanparys, P., Deknudt, G., Jacobs, G., Prinsent, M., Kalweit, S., Spielmann, H., 1994b. Interlaboratory assessment of the bovine corneal opacity and permeability (BCOP) *Toxicology In Vitro*, 8(3): 381-392.

Gettings, S.D., Lordo, R.A., Hintze, K.L., Bagley, D.M., Casterton, P.L., Chudkowski, M., Curren, R.D., Demetrulias, J.L., DiPasquale, L.C., Earl, L.K., Feder, P.I., Galli, C.L., Glaza, S.M., Gordon, V.C., Janus, J., Kurtz, P.J., Marenus, K.D., Moral, J., Pape, W.J.W., Renskers, K.J., Rheins, L.A., Roddy, M.T., Rozen, M.G., Tedeschi, J.P., Zyracki, J., 1996. The CTFA evaluation of alternatives program: an evaluation of *in vitro* alternatives to the Draize primary eye irritation test. (phase III) Surfactant-based formulations. *Food and Chemical Toxicology*, 34(1): 79-117.

Gilleron, L., Coecke, S., Sysmans, M., Hansen, E., van Oproy, S., Marzin, D., van Cauteren, H., Vanparys, P., 1996. Evaluation of a modified HET-CAM assay as a screening test for eye irritancy. *Toxicology In Vitro*, 10: 431-446.

Grant, W.M., 1986. *Toxicology of the Eye*. 3rd ed. Charles C Thomas: Springfield. IL USA.

Guerriero, F.J., Seaman, C. W., Olson, M.J., Whittingham, R., 2004. Retrospective assessment of the rabbit enucleated eye test (REET) as a screen to refine worker safety studies [Abstract No. 1282] *Toxicological Sciences (The Toxicologist Supplement)*, 78(1-S).

Hackett, R. B., McDonald T.O., 1994. Mechanisms of ocular response to irritants. *Dermatotoxicology*, 5th Edition, (ed. Marzulli, F.N., Maibach, H.I.) pp. 299-306. Washington: Taylor & Francis.

Hagino, S., Kinoshita, S., Tani, N., Nakamura, T., Ono, N., Konishi, K., Iimura H., Kojima, H., Ohno, Y., 1999. Interlaboratory validation of in vitro eye irritation tests for cosmetic ingredients. (2) Chorioallantoic membrane (CAM) test. *Toxicology In Vitro*, 13: 99-113.

Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Janusch Roi A, Prieto P, Sabbioni E, Scott L, Worth A, Zuang V. 2004. A modular approach to the ECVAM principles on test validity. *ATLA*, 32: 467-472.

Hogan, M. J., Zimmerman, L. E., 1962. Ophthalmic Pathology: An Atlas and Textbook, 2nd edition. Philadelphia: W.B.Saunders Company.

Hubbard, A. W., Moore, L. J., Clothier, R. H., Sulley, H., Rollin, K. A., 1993. The use of *in vitro* methodology to predict the irritancy potential of surfactants. Toxicology In Vitro, 8: 689-691.

Hubert, F., 1992. The eye (Rabbit/Human): Parameters to be measured in the field of ocular irritation. ATLA, 20: 476-479.

ICCVAM/NICEATM. Current Status of In Vitro Test Methods for Detecting Ocular Corrosives and Severe Irritants Background Review Documents (BRDs): Bovine Corneal Opacity and Permeability (BCOP). Draft version 2004. Documents accessed on the web, June 2006.

ICCVAM/NICEATM. Current Status of In Vitro Test Methods for Detecting Ocular Corrosives and Severe Irritants Background Review Documents (BRDs): Hen's Egg Test - Chorioallantoic Membrane (HET-CAM). Draft version 2004. Documents accessed on the web, June 2006.

ICCVAM/NICEATM. Current Status of In Vitro Test Methods for Detecting Ocular Corrosives and Severe Irritants Background Review Documents (BRDs): Isolated Chicken Eye (ICE). Draft version 2004. Documents accessed on the web, June 2006.

ICCVAM/NICEATM. Current Status of In Vitro Test Methods for Detecting Ocular Corrosives and Severe Irritants Background Review Documents (BRDs): Isolated Rabbit Eye (IRE). Draft version 2004. Documents accessed on the web, June 2006.

Igarashi, I., Katsuta, Y., Matsuno, H., Nakazato, Y., Kawasaki, T., 1989. Opacification test by using the pig isolated cornea and its application to a test corneal opacity induced by befunolol hydrochloride. The Journal of Toxicological Sciences, 14: 91-103.

Jacobs, G. A., Martens, M. A., 1990. Quantification of eye irritation based upon *in vitro* changes of corneal thickness. ATLA, 17: 255-262.

Jones, P. A., Budynsky, E., Cooper, K. J., Decker, D., Griffiths, H. A., Fentem, J. H., 2001. Comparative evaluation of five *in vitro* tests for assessing the eye irritation potential of hair-care products. ATLA, 29(6): 669-92.

Kahn, C.R., Young, E., Lee, I.H., Rhim, J.S., 1993. Human corneal epithelial primary cultures and cell lines with extended life span: *in vitro* model for ocular studies. Investigative Ophthalmology and Visual Science, 34(12): 3429-41.

Katahira, J. H., Sugiyama, H., Inoue, N., Horiguchi, Y., Matsuda, M., Sugimoto, N., 1997. Clostridium perfringens enterotoxin utilizes two structurally related membrane proteins as functional receptors *in vivo*. Journal of Biological Chemistry, 272: 26652-26658.

Kennah, H.E., Hignet, S., Laux, P., Dorko, J.D., Barrow, C.S., 1989. An objective

procedure for quantitating eye irritation based upon changes of corneal thickness. *Fundamental and Applied Toxicology*, 12: 258-268.

Klyce, S. D., Beuerman, R.W., 1988. Structure and function of the cornea In *The Cornea* (ed. H.E. Kaufman) pp. 3-23. New York: Churchill-Livingstone.

Kruszewski, F.H., Walker, T.L., DiPasquale, L.C., 1997. Evaluation of a human corneal epithelial cell line as an *in vitro* model for assessing ocular irritation. *Fundamental and Applied Toxicology*, 36: 130-140.

Larson, E. M., Doughman, D. J., Gregerson, D. S., Obritsch W. F., 1997. A new, simple nonradioactive, nontoxic *in vitro* assay to monitor corneal endothelial cell viability. *Investigative Ophthalmology and Visual Science*, 38: 1929-1933.

Lewis, R.L., Sweet, D.V., (eds.) 1985. *Registry of Toxic Effects of Chemical Substances, 1983-1984 Supplement*. Cincinnati, OH, USA: NIOSH

Lovell, D. P. 1998. Statistical analysis of Phase II and III fluorescein leakage assay results. BIBRA Project No. 3278.

Martin, K.M., Stott, C.W., 1992. The transepithelial permeability assay as *in vitro* assay for predicting ocular irritation of surfactant formulations. In *In Vitro Toxicology: 10th Anniversary of CAAT (Alternative Methods in Toxicology, Vol. 9, ed., A.M. Goldberg)* (In Press). New York: Mary Ann Liebert, Inc. POSTER

Maurer, J. K., Parker, R. D., 1996. Light microscope comparison of surfactant-induced eye irritation in rabbits and rats at three hours and recovery/day 35. *Toxicologic Pathology*, 24: 403-411.

Moore, P., Ogilvie, J., Horridge, E., Mellor, I.R. Clothier, R.H., 2005. The development of an innervated epithelial barrier model using a human corneal cell line and ND7/23 sensory neurons. *European Journal of cell Biology*, 84: 581-592.

O'Brien, J., Wilson, I., Orton, T., Pognan, F., 2000. Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *European Journal of Biochemistry*, 267(17): 5421-6.

Organisation for Economic Co-operation and Development Guidelines for Testing of Chemicals. 1987. No. 45 Acute Eye Irritation/ Corrosion'

Organisation for Economic Co-operation and Development, 2002. OECD TG (405) for testing of chemicals. Acute eye irritation/corrosion.

Page, B., Page, M., Noel, C., 1993. A new fluorimetric assay for cytotoxicity measurements in vitro. *International Journal of Oncology*, 3: 473-476.

Prinsen, M. K., Koeter, B.W.M., 1993. Justification of the enucleated eye test with eyes of slaughterhouse animals as an alternative to the Draize eye irritation test with rabbits. *Food and Chemical Toxicology*, 31: 69-76.

Purchase, I.F.H., Farrar, D.G., Whitaker, I.A., 1993. Toxicology profiles on substances used in the FRAME cytotoxicology research project. *ATLA*, 14:184-243.

Shaw, A.J., Clothier, R.H., Balls, M., 1990. Loss of Trans-epithelial Impermeability of a Confluent Monolayer of Madin Darby canine kidney (MDCK) Cells as a Determinant of Ocular Irritancy Potential. *ATLA*, 18: 145-151.

Shaw, A.J, Balls, M., Clothier, R.H., Bateman, N.D., 1991. Predicting ocular irritancy and recovery for injury using MDCK cells. *Toxicology In Vitro*, 5: 569-571

Southee, J. A., 1998. Evaluation of the Prevalidation Process: The Fluorescein Leakage Assay. Commissioned report for ECVAM.

Spielmann, H., Liebsch, M., Kalweit, S., Moldenhauer, F., Wirnsberger, T., Holzhütter, H., Schneider, B., Glaser, S., Gerner, I., Pape, W.J.W., Kreiling, R., Krauser, K., Miltenburger, H.G., Steiling, W., Luepke, N.P. Müller, N., Kreuzer, H., Mürmann, P., Spengler, J., Bertram-Neis, E., Siegemund, B., Wiebel, F., 1996. Results of a validation study in Germany on two *in vitro* alternatives to the Draize eye irritation test, HET-CAM test and the 3T3 NRU cytotoxicity test. *ATLA*, 24:741-858.

Tchao, R., 1988. Trans-epithelial permeability of fluorescein *in vitro* as an assay to determine eye irritants. In "Progress In In-Vitro Toxicology," 6: 271-283.

UN., 2003. Globally Harmonised System of Classification and Labelling of Chemicals (GHS). New York & Geneva: United Nations Publications.

Van Meer, G., Van Hof, W., Van Genderen, I., 1992. Tight junctions and polarity of lipids In Tight junctions(ed. Cereijido M.) pp. 187-201. London: CRC Press.

Ward, R.K., Mungall, S., Carter, J., Clothier, R.H., 1997a. Evaluation of tissue culture insert membrane compatibility in the fluorescein leakage assay. *Toxicology In Vitro*, 11:761-768.

Ward, S.L., Walker, T.L., Dimitrijevic, S.D., 1997b. Evaluation of Chemically Induced Toxicity Using an *In Vitro* Human Corneal Epithelium. *Toxicology In Vitro*, 11:121-139.

Wilhelmus, K.R., 2001. The Draize eye test. *Surveys in Ophthalmology*, 45: 493-515.

Wilkinson, P., Clothier R. H., 2005. Comparison of an animal Product free medium and normal growth supplements on the growth and barrier integrity of a human corneal epithelial cell line. *ATLA*, 33: 1-10.

Wilkinson, P.J., 2006. Development of an *in vitro* model to investigate repeat ocular exposure. Ph.D. Thesis. University of Nottingham, UK.

Zanvit, A., Meunier, P. A., Clothier, R., Ward, R., Buiatti-Tcheng, M., 1999. Ocular irritancy assessment of cosmetics formulations and ingredients: fluorescein leakage test. *Toxicology In Vitro*, 13: 385-391.