6. GENERAL DISCUSSION

Significant progress has been achieved in the development of *in vitro* methods during the past ten years. Some of the *in vivo* toxicological assays, e.g. *in vivo* phototoxicity tests, *in vivo* pyrogenicity test or *in vivo* skin corrosion test are already replaced by reliable *in vitro* assays. However, regulatory acceptance of the new *in vitro* methods is still slow, mainly due to long-running validation trials and acceptance process, which itself may take several years. Moreover, the requirements of regulators to compare validation outcomes of highly standardised *in vitro* systems with animal data (usually of low level quality), complicate the successful development and validation of new alternative methods.

The aim of presented thesis was to critically review the problems and obstacles connected with the process of development and validation of the new test methods based on the use of reconstructed human skin models. Experiments conducted should demonstrate their usability in topical toxicity testing. Structurally, the thesis comprises of the following four tasks:

TASK 1: Evaluation of present-day quality of several commercially available reconstructed human skin models applying recommendations and criteria described by the OECD Test Guideline 431.

Currently, the only regulatory accepted method based on the use of the reconstructed human skin models is the *in vitro* skin corrosion test - OECD Test Guideline 431. The guideline broadly specifies criteria, which the reconstructed human skin model should fulfil if used within its scope. Besides correct prediction of the 12 reference chemicals given in Annex I of the guideline, evaluation of tissue morphology, presence of functional barrier and sufficient tissue viability are required. However, the guideline does not describe or give an appropriate reference how these test should be performed. Therefore, it opens the door to many possibilities but also to many interpretations.

In this thesis, several basic methods were proposed which could be applied for all reconstructed human skin models as a standard test battery. Three commercially available, and most frequently used reconstructed human skin models were selected as benchmark models: EpiDerm (MatTek Corporation, USA), EPISKIN (EPISKIN SNC, France) and SkinEthic (SkinEthic Laboratories, France). Although their characteristics were extensively investigated several years ago (e.g. Roguet *et. al*., 1998, Boelsma 1999, Ponec *et al*., 2000; Ponec *et al*., 2002), evaluation of their present-day quality was of interest, as it is known, that there were ongoing barrier improvements (e.g. by adding vitamin C into the culture medium).

Obviously standardisation of the manufacturing process improved inter-batch variability, described by Boelsma in 1999 and in Ponec *et al*. in 2002. Some differences in lipid profiles were identified when comparing current results with the results of Boelsma (1999). An increased level of ceramide 6 in EPISKIN and SkinEthic cultures was quite obvious. With regard to the tissue morphology, SkinEthic cultures, revealed higher number of viable cell layers as it was observed in 1999 (Boelsma 1999) and 2003-2004 (personal observation).

Probably the most challenging proofs for the three models were the TEER and ET 50 assay with Triton X-100. The OECD TG 431 states that*: "the stratum corneum should be sufficiently robust to resist the rapid penetration of certain cytotoxic marker chemicals (e.g. 1% Triton X-100). This property can be estimated by the exposure time required to reduce cell viability by 50% (ET50) (e.g. for the EpiDerm and EPISKIN models this is > 2 hours)".* By definition, it is assumed, that other reconstructed human skin models should provide similar results. However, it has to be noted, that the ET 50 values can be easily manipulated to obtain desired outcome. E.g. decreasing of the application volume, leads to higher ET 50. Moreover, it seems to be of importance which solvent (PBS or deionised water) is used for the preparation of 1% Triton solution as this may influence on the surfactant activity. In present study, standardised volume of 160 µl/cm^2 of 1% Triton (water solution) was used as this dose sufficiently covers the tissue surface without a need of additional spreading support. The experiment was based on MatTek quality assurance procedure, where 100µl of water solution of Triton is applied on 0.63 cm^2 (dose of 160 µl/cm^2) and was combined with measurements of TEER at each exposure. Significant differences were identified between the three reconstructed human skin models. Based on viability results, the lowest sensitivity to the surfactant revealed the EPISKIN model. EpiDerm and SkinEthic models showed more or less similar ET 50 values. This effect has been observed also during the skin irritation studies with another surfactant – sodium dodecyl sulphate (SDS). The different sensitivity to surfactants (and certain group of lipophilic substances used in later studies) is most probably linked to the combination of the lipid composition and architecture of the stratum corneum.

It seems, that measurement of the TEER might allow to detect impaired barrier of the reconstructed human skin models. However, to evaluate this hypothesis systematic investigations with “qualified” and “non-qualified” tissues should be performed. This is however challenging issue as manufacturers release only occasionally such a type of products.

In summary, all three reconstructed human skin models revealed good performance in all investigated items and reflected the standards required by the OECD TG 431. Certain improvements were obvious by all three reconstructed human skin models, however for better understanding of accomplished refinements, additional immunohistology investigations

are needed. In addition to the ET 50 experiments with Triton X-100, a lipophilic substance should be investigated for evaluation of the models quality, as the surfactant-based assay as a stand-alone test does not provide sufficient information about the barrier resistance.

TASK 2: Prove of transferability of the existing protocol for in vitro skin corrosion to various reconstructed human skin models in order to evaluate the idea of the "common protocol concept"

It has been hypothesised that well developed reconstructed human skin models, showing similar attributes, should also provide similar results in one robust test protocol (Liebsch *et al*., 1997; Liebsch *et al*., 2000). As described above, the SkinEthic model revealed many characteristics similar to EpiDerm and EPISKIN models, already validated for skin corrosion testing. Therefore, when developing the SkinEthic skin corrosion assay, the EpiDerm skin corrosion protocol was taken as a base. After minor technical adjustments, the SkinEthic skin corrosion assay was evaluated in a validation trial between three laboratories and followed all feasible validation principles. The SkinEthic model revealed almost identical outcome as EpiDerm and SkinEthic models in the ECVAM skin corrosion validation studies (Fentem *et al*., 1998; Liebsch *et al*., 2000), Moreover, the assay and model fulfil all criteria required in the OECD TG 431 to obtain regulatory acceptance.

Importantly, the OECD TG 431 requires the correct prediction of the 12 reference test substances. However, one chemical (10 % sulphuric acid) appears not classified correctly in the guideline. Sulphuric acid 10 % should be classified only as irritant (ECB, 2005). Indeed, when tested on SkinEthic model, no corrosive effects were observed after one hour exposure. The pH of Sulphuric acid 10 % is 1.2 (Barrat *et al*., 2000). Following the refined test strategy of the OECD TG 404, the chemical would have been be classified as corrosive without any need for *in vitro* testing. However, this might be an interesting and rare case, where the QSAR rule does not work. The reactivity of low concentrated sulphuric acids is most probably "inactivated" by lipids of the stratum corneum. This situation is however rare and specific only for very small group of chemicals. For instance, 5 or 10% NaOH would cause corrosion effects, as the mechanism of penetration is completely different (NaOH directly digests the tissues, while acids at low concentration penetrate most probably via extra-cellular spaces). Except or this chemical, the SkinEthic reconstructed human skin model provided predictions in concordance with OECD TG 431 and previously performed ECVAM validation trials. The study results were recently published in Toxicology *in Vitro* (Kandárová *et al*., 2006a) and the outcome is currently reviewed by ECVAM Scientific Advisory Committee (ESAC) which should provide a statement about scientific validity and acceptance of this method for regulatory testing.

The transferability of the skin corrosion protocol was further evaluated with the EST-1000 model (Cell Systems, Germany). The model has become commercially available only recently, therefore detailed information about the tissue morphology and lipid content are published by the manufacturer only (Hoffmann *et al*., 2005). The experiments with EST-1000 model at ZEBET applying the "common skin corrosion test protocol" revealed an outcome comparable to previously performed study with SkinEthic. Although the skin model seems to be slightly more sensitive to potassium hydroxide, the assay provided satisfactory results. A validation study, managed by Cell Systems, is currently performed in four laboratories, to assess the interlaboratory reproducibility of the assay with the 12 OECD reference chemicals (Hoffman, personal communication).

In summary, the idea of the common protocol for skin corrosion tests appears to work, under the assumption that the reconstructed human skin models have sufficiently developed barrier.

TASK 3.: Based on the common protocol concept, to optimise and evaluate skin irritation protocols for EpiDerm and SkinEthic models for skin irritation validation studies.

In comparison to skin corrosion, which is very simple phenomenon, skin irritation is highly complicated cascade comprising different effects, actions and reactions. Thus, to develop one single *in vitro* test protocol which could easily replace the *in vivo* test appears to be challenging. The situation is complicated by the fact, that the reconstructed human skin models were developed to resemble human skin, its attributes (e.g. morphology lipid profile and barrier) and reactions. However, most of the regulators require that the reconstructed human skin models should provide results similar to those obtained on skin of animals, which is in case of skin irritation the skin of albino rabbits. However, it is well known that skin of rabbit is far more sensitive to injury than human skin and the response to some chemical groups can be fairly different in both species.

It is in general assumed, that the reconstructed human skin models have less developed barrier than the human skin and the skin of animals (e.g. rabbit). However, it is often forgotten, that the conditions of *in vivo* skin irritation test are far from exposure to chemicals which would happen in realistic conditions e.g. by an accident.

In the *in vivo* test, the primary barrier of the animal skin, the fur, is removed. By this procedure, the stratum corneum is exerted and might be also damaged by shaving. Next, the amount of hair follicles in rabbit skin is not comparable with human skin. While penetration via hair follicles is only negligible in man, the plenty of hair follicles in rabbit skin enhance the

penetration and may transport the compounds directly into the dermis. Finally, the four hour exposure in the *in vivo* test seems to be not adequate for a prediction of the irritation effects of industrial chemicals, as in case of an accident, the skin of injured person would be decontaminated much faster (usually within one hour).

The next factor, which leads to enhanced penetration are the occlusive conditions of the *in vivo* test. This procedure causes swelling of the stratum corneum and permits the penetration via keratinocytes which at normal conditions happens only rarely due to the highly resistant cornified envelope of keratinocytes. Moreover, volatile chemicals which would quickly evaporate from the tissue surface are kept unnaturally on the skin.

All above mentioned facts clearly indicate, that, although the primary aim of the *in vivo* rabbit skin irritation test is to protect the consumer, the test provides inadequate overpredictions. This has been confirmed in studies of Basketter, Robins, Phillips and others. Mainly “mild irritants” present a problem. While the rabbit test usually correctly predicts severe irritants and clear non-irritants, it completely fails in discrimination of mild irritants. There are also cases known where the rabbit test was negative and in human patch test the human volunteers were irritated by the test chemical (e.g. DMSO or Lactic acid).

Therefore, when performing validation studies to replace *in vivo* skin irritation test for industrial chemicals, all these facts should be carefully considered and only those chemicals should be selected for which rigid databases of *in vivo* data (preferably from more than one source) exist. Moreover, well documented toxicology profiles as well as description of physical and chemical attributes of test substances should be available to allow for future examinations e.g. using QSAR or confirmatory studies on human volunteers for non-toxic chemicals.

In this thesis, two studies with reconstructed human skin models EpiDerm and SkinEthic were performed applying the refined protocol of EPISKIN model (Cotovio *et al*., 2005). During the study with the EpiDerm model, *in vivo* data obtained from ECETOC database No. 66 (ECETOC, 1995) were retrospectively evaluated by an BfR expert. It revealed that at least one chemical (Dimethyl disulphide) should not be used in the testing set, due to the incorrect conductance of the *in vivo* test. In some other tests invalid scoring system have been applied using decimal scores. Moreover, many chemicals showed high variability in observed effects between the test animals, ranging from irritant to non-irritant (e.g. Tallow propylene polyamine). As it can be seen from Figure 33, there is no clear border which could discriminate irritants from non irritants reliably. Many of the substances can be found in the in the middle part of scoring scale are therefore classified only as mild irritants in GHS classification system.

In the EpiDerm skin irritation assay, several substances (e.g. methyl palmitate, methyl stearate or linalyl acetate) were predicted as “false negatives” when compared with the rabbit

test. However, these chemicals are non-irritating to human skin. These three chemicals were classified as non-irritants also by EPISKIN model and the first two provided the same outcome when tested on SkinEthic model (linalyl acetate was not tested with SkinEthic model). This prediction is in concordance with the response of human skin. However, because regulators accept as a gold standard only the rabbit test, even substances predicted in concordance with human response are classified as false negative and the sensitivity of the tests is unfairly decreased.

Another very specific case are chemicals able to remove the SC lipids (so called defatting agents). Although in this cases no erythema or oedema formation is observed, based on appearance of slight superficial whitening, regulators classify these chemicals as irritants, too. As already mentioned, volatile chemicals present another very specific problem for *in vitro* studies. Because the skin whitening can not be observed with reconstructed human skin models and the occlusive conditions can not be fully reproduced, defatting agents and highly volatile chemicals tested *in vitro* might be under-predicted in comparison to animal test. On the other hand, the question arises, if the conditions of *in vivo* test are adequate.

Despite the problems described above, based on results presented in this thesis and already published (Kandárová *et al*. 2004 and 2005), the EpiDerm skin irritation test protocol was included into the ECVAM skin irritation validation study. Moreover, the SkinEthic skin irritation assay, which is based on the common protocol should be (in case of successful ECVAM validation study with EpiDerm and EPISKIN models) evaluated in a catch-up validation trial.

TASK 4. Evaluation of applicability of the EpiDerm Phototoxicity assay as an adjunct test for testing of substances "over-predicted" in the validated 3T3 NRU-PT test (OECD TG 432).

It is known that the validated 3T3 NRU PT (using monolayer fibroblast cell culture) may provide over-predictions of phototoxicity of certain chemicals due to a lack of a penetration barrier. These chemicals are subsequently classified as phototoxins. However, if these substances are used only in topical formulations and do not penetrate the stratum corneum in sufficient amount, they can be still used in safe concentrations (Liebsch *et al*., 2005). This information can not be obtained in the 3T3 NRU PT, therefore, due to the presence of barrier function, the reconstructed human skin models may be useful for that purpose.

The experiments described here showed, that the reconstructed human skin model EpiDerm appears to provide the desired information, which is concordant with the results of

the human patch tests. This was demonstrated in studies with UV filters and essential oil *Litsea Cubea. C*lear and correct predictions were obtained with the human skin model. Yet, care must be taken, by complex mixtures like Bergamot oil, as purity, concentration of phototoxic compounds and age of the sample may significantly influence the final result. Moreover, the test assay should be adjusted to real-use conditions. This is of particular importance for cosmetic and pharmaceutical industry, as the base formulations may enhance penetration of the compounds into the viable layers of epidermis, and in rare cases due to the penetration via hair follicles, also into the hypodermis.

The study results were submitted to ECVAM in form of interim report as the human patch tests, are not completed. However, the preliminary results show, that the reconstructed human skin model is able to predict the photopotency of chemicals in concordance with human data.