

## Alternativer til dyreforsøg for øjenirritation

Eva Selzer Rasmussen  
Fødevaredirektoratet.  
Institut for Fødevarerikkerhed og Toksikologi

Miljøstyrelsen vil, når lejligheden gives, offentliggøre rapporter og indlæg vedrørende forsknings- og udviklingsprojekter inden for miljøsektoren, finansieret af Miljøstyrelsens undersøgelsesbevilling.

Det skal bemærkes, at en sådan offentliggørelse ikke nødvendigvis betyder, at det pågældende indlæg giver udtryk for Miljøstyrelsens synspunkter.

Offentliggørelsen betyder imidlertid, at Miljøstyrelsen finder, at indholdet udgør et væsentligt indlæg i debatten omkring den danske miljøpolitik.

# Indhold

<b>FORORD</b> .....	4
<b>SAMMENDRAG</b> .....	5
<b>ENGLISH SUMMARY</b> .....	6
<b>1 INDLEDNING</b> .....	7
<b>2 REKONSTRUEREREDE VÆVSMODELLER</b> .....	9
2.1 REKONSTRUKTION AF VÆV .....	9
<b>3 COLIPA'S VALIDERING AF ALTERNATIVER TIL DRAIZE ØJENIRRITATIONSTESTEN</b> .....	12
3.1 UNDERSØGELSENS DESIGN .....	13
3.2 TESTSTOFFER .....	13
3.3 ALTERNATIVE METODER .....	14
3.4 ETABLERING AF FORUDSIGELSESMODELLER .....	17
3.5 INDSAMLING AF IN VITRO RESULTATER .....	19
3.6 STATISTISKE METODER .....	19
3.7 YDERLIGERE ANALYSER AF COLIPA UNDERSØGELSENS RESULTATER .....	20
3.8 RESULTATER .....	20
3.9 DISKUSSION .....	27
<b>4 TEST FOR RECOVERY FRA ØJENIRRITATION</b> .....	30
4.1 METODE .....	30
4.2 RESULTATER .....	31
4.3 DISKUSSION .....	32
<b>5 VALIDERING AF ALTERNATIVER TIL ØJENIRRITATIONSTESTEN</b> .....	34
<b>6 EVALUERING AF ALTERNATIVER TIL ØJENIRRITATIONSTESTEN</b> .....	40
<b>7 FORKORTELSER</b> .....	43
<b>8 LITTERATURLISTE</b> .....	46

# Forord

Miljøstyrelsen har støttet projektet "Validering af en tredimensionel model for øjen- og hudirritation". Projektet er gennemført som et samarbejdsprojekt mellem Miljøstyrelsen og Veterinær- og Fødevaredirektoratet, Institut for Fødevaresikkerhed og Toksikologi. Projektarbejdet med vævsmodellen SKIN<sup>2</sup> ZK1200 har givet anledning til udarbejdelse af fire videnskabelige artikler, der danner grundlaget for denne rapport:

Brantom, P. G. et al., A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. *Toxicology in Vitro*, 1997, 11, 141-179.

Espersen, R. J., Olsen, P., Nicolaisen, G., Jensen, B. L. & Rasmussen, E. S., Assessment of recovery from ocular irritancy using a human tissue equivalent model. *Toxicology in Vitro*, 1997, 11, 81-88.

Rasmussen, E. S., Brug af rekonstruerede humane vævsmodeller i toksikologiske og farmakologiske undersøgelser. *Dansk Veterinærtidsskrift*, 1/11, 1996, 943-948.

Southee, J. A., McPherson, J. P., Osborne, R., Carr, G. J. & Rasmussen, E. S., 1997: The performance of the tissue equivalent assay using the SKIN<sup>2</sup>™ ZK1200 model in the COLIPA international validation study on alternatives to the Draize eye irritation test. *Toxicology in Vitro*, 1999, 13, 355-373.

Projektet har været fulgt af en styringsgruppe med følgende medlemmer:

Miljøstyrelsen:

Lone Mikkelsen (tidligere formand)

Lars Nørgaard (nuværende formand)

Anne Marie Linderstrøm

Veterinær- og Fødevaredirektoratet:

John Christian Larsen

Eva Selzer Rasmussen (projektleder)

Projektet er gennemført af en projektgruppe på Institut for Fødevaresikkerhed og Toksikologi med følgende medlemmer:

René Espersen

Preben Olsen

Bo Lund Jensen

Gert Maarløw Nicolaisen

Karen Roswall

Martin Bach

Eva Selzer Rasmussen

Både styringsgruppen og projektgruppen takkes for godt samarbejde.

# Sammendrag

Modeller af menneskeligt væv kan konstrueres ud fra enkelte celler. Teststoffer kan påføres direkte på vævet, og stoffers lokalirriterende egenskaber kan undersøges ved målinger af cellernes vitalitet. En vævsmodel for øjets hornhinde, SKIN<sup>2</sup> ZK 1200, er blevet undersøgt som mulig erstatning for test for øjenirritation med kaniner. Vævsmodellen blev anvendt til at bedømme 55 kosmetiske indholdsstoffer og færdige produkter. Testningen indgik i en stor valideringsundersøgelse arrangeret af den europæiske kosmetikindustri (COLIPA). Vævsmodellen blev vist at være meget god til forudsigelse af øjenirritation i kaninforsøg (Draize test MMAS værdier). Desuden kunne metoden forudsige individuelle vævsreaktioner i øjet med stor sikkerhed. Forudsigelser af øjenirritation var både holdbare for kosmetiske ingredienser og færdige produkter. Med SKIN<sup>2</sup> ZK1200 blev der yderligere etableret en model, der kunne forudsige, om øjenirritation ville være vedvarende i et forsøg med 9 teststoffer. De opnåede resultater tyder på at SKIN<sup>2</sup> ZK1200 modellen potentielt ville kunne erstatte dyreforsøg for øjenirritation. Vævsmodellen er udgået af produktion efter undersøgelsens afslutning, men de opnåede resultater er relevante for vurdering af mulighederne for at kunne erstatte dyreforsøg for øjenirritation med alternative metoder.

COLIPA studiet bekræftede i øvrigt resultaterne af andre større undersøgelser med validering af alternativer til dyreforsøg for øjenirritation. Ingen af de øvrige alternative metoder i COLIPA studiet var egnede til forudsigelse af øjenirritation for alle teststoffer. Flere alternative metoder fandtes dog velegnede til at forudsige øjenirritation forårsaget af ingredienser, specielt af tensider. Tests med fosterhinder fra hønseæg havde en ringe reproducerbarhed mellem laboratorier, mens reproducerbarheden af de øvrige alternative metoder mellem forskellige laboratorier var god.

# English Summary

Models of human tissues can be reconstructed from single cells. Test substances can be applied directly to the tissues, and the potency for local irritancy can be evaluated by measurements of the viability of the cells. A corneal model, the SKIN<sup>2</sup> ZK1200 tissue equivalent, was evaluated as a possible replacement of tests for ocular irritation using rabbits. The tissue equivalent was used to evaluate 55 cosmetic ingredients and finished products in a large validation study arranged by the European cosmetics industry (COLIPA). The SKIN<sup>2</sup> ZK1200 model was shown to be very good in predicting ocular irritation (Draize test MMAS values). Moreover, the tissue equivalent assay gave a good prediction of individual tissue reactions in the eye. This was both obtained for cosmetic ingredients and formulations, and all types of materials could be tested. In addition, a new model developed with the tissue equivalent showed promise for prediction of recovery from ocular irritation in a study on nine test substances. The results obtained indicate that the SKIN<sup>2</sup> ZK1200 tissue equivalent showed promise as a replacement of animal experiments for ocular irritancy testing. After the study was completed, the production of the SKIN<sup>2</sup> ZK1200 model has ceased. The results obtained with the model are, however, considered to be relevant for the assessment of the prospects for replacement of animal experiments for ocular irritancy with alternative methods.

The results of the COLIPA study did moreover confirm the results of other large validation studies on alternatives to animal experiments for eye irritancy testing. None of the other alternative methods in the COLIPA study were suitable to predict ocular irritation caused by the mixed group of ingredients and formulations. Several alternative methods were, however, very good in predicting ocular irritation caused by surfactants. Tests using hens eggs showed a low interlaboratory reproducibility, whereas the reproducibility of the other alternative methods was good.

# 1 Indledning

På verdensplan er der en stærkt øget interesse for brug af alternative metoder til dyreforsøg, specielt med henblik på vurdering af sikkerheden af kosmetiske produkter. Der har været anvendt flest ressourcer på at udvikle alternative metoder til vurdering af øjenirritation, hvor de eksisterende dyreforsøg med kaniner er blevet intensivt kritiseret udfra videnskabelige og etiske overvejelser.

Der er udviklet en lang række *in vitro* tests, som mulige alternativer til *in vivo* øjenirritationsforsøg. Disse metoder omfatter simple fysisk-kemiske *in vitro* tests med planteproteiner, metoder med kulturer af celler fra mennesker eller pattedyr, systemer med isolerede øjne fra køer, kaniner eller høns, og tests med fosterhinder i hønseæg. En ny type *in vitro* tests med menneskelige væv, der rekonstrueres udfra enkelte celler, er under udvikling. Brug af sådanne vævsmodeller har givet lovende overensstemmelser med resultater fra dyreforsøg og humane data på en lang række områder. Denne rapport indledes med et afsnit om konstruktion af humane vævsmodeller specielt med henblik på test for lokalirriterende effekter.

Rapporten omhandler tillige forsøg med en vævsmodel for øjets hornhinde, SKIN<sup>2</sup> ZK1200, der er blevet vurderet som mulig erstatning for test for øjenirritation med kaniner. 55 kosmetiske indholdsstoffer og færdige produkter blev testet i en valideringsundersøgelse arrangeret af den europæiske kosmetikindustri (COLIPA), hvor 10 alternative metoder blev vurderet. Institut for Fødevarerikkerhed og Toksikologi deltog i valideringsprojektet støttet af Miljøstyrelsen som et ud af tre laboratorier, der anvendte SKIN<sup>2</sup> ZK1200 modellen.

COLIPA er en europæisk association af producenter af kosmetik, toiletvarer og parfumeartikler. Organisationen blev etableret i 1962, og den repræsenterer 95% af den kosmetiske industri i EU landene med ca. 2000 firmaer. COLIPA omfatter udover de nationale associationer af firmaer fra 15 medlemsstater i EU, 6 grupper af firmaer fra ikke-EU lande og 21 større internationale firmaer.

Kosmetikindustrien er en af de væsentligste drivkræfter bag udvikling og validering af alternativer til dyreforsøg til kosmetiktestning, og COLIPA er medarrangør af en lang række programmer om validering af alternative tests. COLIPA nedsatte i 1992 en komité om alternativer til dyreforsøg (SCAAT), der har det overordnede ansvar for organisationens arbejde på dette område. SCAAT's medlemmer kommer fra firmaerne Biersdorf, L'Oreal, Procter & Gamble, SmithKline Beecham, Unilever og Wella, men alle organisationens internationale firmaer bidrager til arbejdet. COLIPA organiserer og deltager i validering af alternativer til dyreforsøg, primært indenfor områderne øjenirritation, fotoirritation fremkaldt af UV-stråling, hudabsorption, og produkters forenelighed med hud. COLIPA har opnået en stor erfaring med alternative metoder, og organisationen er en aktiv medspiller ved EU Kommissionens vurderinger af metodernes fremskridt.

Interessen for udvikling af alternative metoder er bl.a. forårsaget af EU's kosmetikdirektiv (EU, 1993). Det fremgår af direktivet, at medlemsstaterne

skal forbyde markedsføring af kosmetiske produkter indeholdende bestanddele eller forbindelser heraf, der er blevet afprøvet på dyr efter den 1 januar 1998 (EU, 1993). Fristen er blevet udskudt til 1. januar år 2000 for kosmetiske ingredienser, da der ikke er etableret alternative metoder endnu. EU Kommissionen overvejer dog at gennemføre et helt eller delvist forbud mod testning af færdige kosmetiske produkter i dyreforsøg inden denne frist.

I COLIPA's valideringsundersøgelse for øjenirritation blev der vist en meget god overensstemmelse mellem forudsigelser foretaget ud fra data fra vævsmodellen SKIN<sup>2</sup> ZK1200 og akut irritation i kaninforsøg. Derimod var den anvendte metode med vævsmodellen i COLIPA studiet ikke velegnet til at forudsige, om øjenirritation vil være vedvarende. I et selvstændigt indledende forsøg med 9 teststoffer blev der derfor etableret en metode til vurdering af bedring (recovery) af akut øjenirritation med SKIN<sup>2</sup> ZK1200 modellen. I rapporten er den nye metode med SKIN<sup>2</sup> ZK1200 modellen til vurdering af recovery fra øjenirritation beskrevet efter COLIPA studiet.

Rapporten slutter med en diskussion af de fremtidige muligheder for erstatning af dyreforsøg for øjenirritation med alternative metoder. De fleste alternative metoder er mere reproducerbare end Draize testen. Derfor bør en vurdering af metodernes evne til at forudsige øjenirritation tillægges større vægt end gentagen dokumentation af resultaternes reproducerbarhed mellem forskellige laboratorier. SKIN<sup>2</sup> ZK1200 modellen er den første *in vitro* metode, der har vist et overbevisende potentiale for fuldstændigt at kunne erstatte Draize testen. Vævsmodellen er udgået af produktion, men det vil være muligt at anvende andre kommercielle vævsmodeller, eller at udvikle nye ikke-kommercielle vævsmodeller til øjenirritationstestning. Andre etablerede alternative metoder har ikke vist sig egnede til en generel forudsigelse af øjenirritation, men flere af disse metoder er formentlig velegnede til test af tensidbaserede produkter.



## 2 Rekonstruerede vævsmodeller

Der er udviklet en lang række *in vitro* modeller med væv, der er rekonstrueret ud fra celler fra menneskelig hud. Der anvendes f.eks. øjenmodeller, der fungerer som en model for hornhindens epithelvæv eller er et kombineret epithel og støttevæv (stroma). En række hudmodeller er opbygget på tilsvarende måde, men her er også opbygget et hornlag (stratum corneum). Modellernes lighed med normalt væv giver mulighed for en lang række praktiske anvendelser indenfor toksikologien og farmakologien (Rasmussen, 1996). Forsøg med rekonstruerede vævsmodellerne har foreløbigt givet gode overensstemmelser til resultater fra dyreforsøg ved testning af mange forskellige typer stoffer og produkter for øjen- og hudirritation, ætsning og fototoksicitet. Vævsmodeller anvendes tillige ved undersøgelser af hudens udvikling og sårheling samt i vurdering af anti-aldnings midler og solbeskyttende præparaters effekt.

### 2.1 Rekonstruktion af væv

Rekonstrueret hud er udviklet til behandling af store brandsår, hvor det ikke er muligt at anvende patientens egen hud til transplantation. Både rent kemisk fremstillede systemer, som skumplader fremstillet af makromolekyler fra huden, og biologiske modeller med levende hudceller er blevet brugt. Dyrkning af overhudsceller (keratinocytter) direkte på sårbunden ved større forbrændinger kan føre til sårsmammentrækning, og forudgående brug af en kunstig underhud (dermis) har vist sig at give en betydeligt bedre sårheling. På den kunstige dermis udsås 4-6 cellelag tykke kulturer (væv) af keratinocytter. Hormoner og andre vækstfaktorer tilføjes fra patientens krop, og keratinocytterne danner et fuldt udviklet epithel med hornlag.

Rekonstruerede hudmodeller indeholder typisk enten udelukkende keratinocytter eller kombinationer af fibroblaster og keratinocytter, og de repræsenterer derfor et meget forenklet system sammenlignet med *in vivo* situationen. Den store udfordring ved konstruktion af hudmodellerne har været at frembringe en flerlaget epidermis med stratum corneum uden hjælp fra en levende organisme. Keratinocytter udsås på filtre eller andre substrater, og danner herpå et sammenhængende lag af basalceller. Hvis processen stoppes her, er der konstrueret en øjenmodel. Hvis cellekulturerne bagefter udsættes for luft, dannes der et flerlaget epidermalt væv, der dækkes af et tyndt hornlag. Herved er der dannet en hudmodel. Den kunstige epidermis består af tilsvarende lag af keratinocytter som den naturlige, og en vellignende ultrastruktur er påvist ved lys- og elektronmikroskopi.

Vævsmodellerne kan bruges til studier af keratinocytternes differentiering, herunder af forandringer af profiler af proteiner og fedtstoffer. Der er påvist en stor biokemisk lighed mellem modellerne og normal hud, men vævsmodellerne adskiller sig dog også på flere punkter markant strukturelt og biokemisk fra *in vivo* situationen. Der er specielt store forskelle i den relative forekomst af forskellige fedtstoffer og i deres strukturelle organisering og fordeling. Desuden sker der en høj grad af diffusion gennem de ikke fuldt forhornede celler i stratum corneum *in vitro*.

Flere vævsmodeller består udelukkende af epidermalt væv dyrket på døde erstatninger for dermis som f.eks. deepidermiseret dermis, mikroporøse filtre, eller en kollagenmatriks. Tilstedeværelsen af en levende dermis har dog formentlig stor betydning for vævets funktioner. Under sårheling *in vitro* bliver keratinocyternes vækst markant øget ved tilstedeværelsen af epidermal vækstoffaktor, og af stoffer der udskilles af basalmembranen og fibroblaster. Forekomsten af et epithel, en basalmembran og et bindevæv i *in vitro* modellerne kan tillige bruges ved vurdering af, om læsioner vil kunne heles uden varig skadevirkning. Når kun de ydre dele af et epithel beskadiges, er konsekvenserne generelt små fordi vævet er i stand til at regenerere. Hvis basalmembranen og stroma også beskadiges, kan der opstå permanente skader.

En levende dermis kan konstrueres på flere måder. Celler fra dermis (fibroblaster) kan f.eks. indstøbes i en gel af kollagenfibre og vævs-kulturmedium. Cellerne hæfter sig hurtigt fast til kollagenfibre, gelen trækkes herved sammen og mediet udstødes. Efter 2-3 dage i kultur dannes en tæt plade af hvidt, kondenseret væv, der er et egnet grundlag for dyrkning af epidermale celler. Levende dermalt væv kan ligeledes fremstilles ved dyrkning af fibroblaster på nylonnet med mikroskopiske masker. Cellerne hæfter sig fast til nettet, de udskiller forskellige dermale matriksmolekyler, og vævet fylder gradvist nettets masker. Herved dannes der et tre-dimensionelt bindevæv, hvorpå andre celletyper kan dyrkes.

En serie af modeller har haft fibroblastvæv dyrket på nylonnet som basis. En SKIN<sup>2</sup> ZK 1100 model består udelukkende af fibroblastvæv. I SKIN<sup>2</sup> ZK 1200 øjenmodellen er dette bindevæv suppleret med 3-4 lag keratinocyter. Vævet er en model for hornhinden, der har et lag af epidermale celler af tilsvarende tykkelse. SKIN<sup>2</sup> ZK 1300 systemer med hornlag har været anvendt som hudmodeller. En model med humane slimhindeceller fra mundhulen er blevet brugt til test af materialer til tandbehandling. Hudvæv opbygget på biologisk nedbrydelige net er udviklet til brug ved transplantationer.

Vævsmodellerne er metabolisk aktive, og de kan yderligere bruges til at undersøge effekter af samspillet mellem forskellige celletyper. Forsøg med dyrkning af pigmentdannende celler (melanocyter) i flerlagede keratinocytkulturer har vist, at flere karakteristiske træk fra *in vivo* situationen kan genskabes. Melanocyterne etablerer sig i det basale keratinocytlag, og de begynder at producere pigmentet melanin. Efter UV-bestråling øges vævets pigmentering. En sådan model er vurderet som meget velegnet til test af stoffers og produkters fototoksiske eller solbeskyttende virkning.

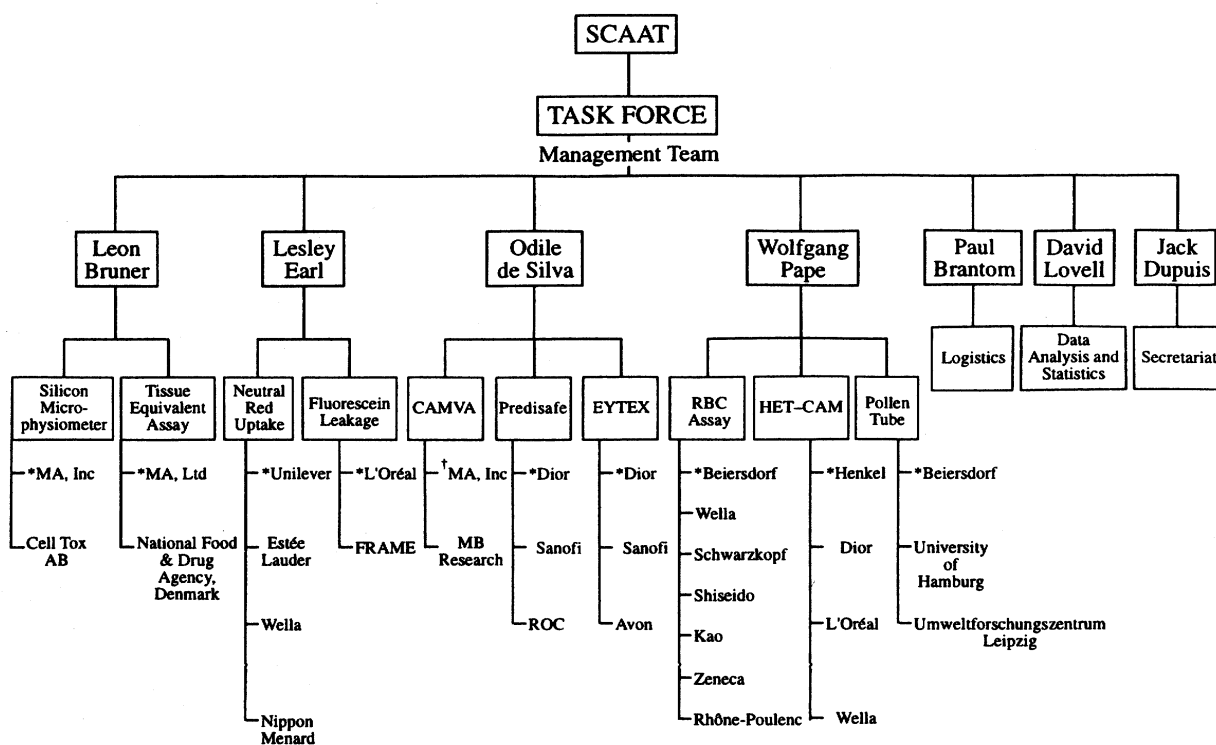
Påvirkning af væv med lokalirriterende stoffer og UV-stråling kan ændre cellernes enzymaktivitet og vækstmønster samt medføre udskillelse af irritationsmarkører som prostaglandiner, leukotriener og cytokiner. I gennem en længere årrække er målinger af sådanne parametre i traditionelt dyrkede cellekulturer blevet vurderet som mulige alternativer til dyreforsøg ved testning af stoffers lokalirriterende effekter. Generelt har resultater fra sådanne systemer været i god overensstemmelse med *in vivo* øjenirritationsdata for vandopløselige forbindelser, men ved en række større blindundersøgelser af blandede stoffer og materialer har metoderne dog vist sig at være fuldstændig utilstrækkelige til forudsigelse af *in vivo* responset. Forsøg med øjenmodeller af rekonstrueret væv er nu blandt de mest lovende alternativer til dyreforsøg. En række procedurer med vævsmodeller er også udviklet til brug ved testning for

ætsning, hudirritation og fototoksicitet, og sådanne metoder er under yderligere validering.

### 3 COLIPA's validering af alternativer til Draize øjenirritationstesten

I 1993 tog COLIPA initiativ til at organisere et program for validering af alternative metoder til Draizetesten for øjenirritation. Programmet var specielt rettet mod test af kosmetiske indholdsstoffer og formuleringer. Studiet blev designet til at bygge på erfaringer fra en tidligere valideringsundersøgelse på dette felt arrangeret af EU Kommissionen og British Home Office (EC/HO studiet) (Balls et al., 1995), hvor 40 % af de deltagende laboratorier kom fra kosmetikindustrien. 20 teststoffer var fælles for begge undersøgelser, og flere af de alternative metoder blev også brugt i begge studier.

Figur 3.1



Opbygning af COLIPA's øjenirritationsstudie.

COLIPA øjenirritationsstudiet var designet til at afgøre, om data fra de alternative metoder kunne give: 1. en acceptabel overensstemmelse med Draize testens modificerede maksimale gennemsnitlige scoringsværdier (MMAS), 2. en acceptabel overensstemmelse med Draize testens individuelle vævsscoringer og tid for recovery, og/eller 3. en rimelig forudsigelse af øjenirritationspotentialer i Draize testen ud fra en forudsigelsesmodel (de Silva, 1996).

### 3.1 Undersøgelsens design

Studiet blev overordnet sponsoreret og organiseret af COLIPA, men enkelte uafhængige forskningsgrupper deltog med egen sponsorering. Veterinær- og Fødevaredirektoratet, Institut for Fødevaresikkerhed og Toksikologi, deltog f.eks. med sponsorering af testningen fra Miljøstyrelsen i et samarbejdsprojekt. COLIPA nedsatte en arbejdsgruppe med ansvar for gennemførelsen af øjenirritationsstudiet, og udpegede et managementudvalg til at forestå ledelsen af projektet (se figur 3.1). Hver test fik tildelt et ledelaboratorium, der forestod koordineringen mellem COLIPA og deltagerne, udarbejdede testprotokol og forudsigelsesmodel for metoden, og varetog den daglige ledelse af den samlede projektgruppes arbejde. Projektet blev planlagt med mindst to laboratorier pr. metode, og kun tre metoder blev påbegyndt i fire eller flere laboratorier (Brantom et al., 1994).

### 3.2 Teststoffer

COLIPA studiet omfattede 55 testmaterialer, der dækkede et bredt spektrum af kosmetiske indholdsstoffer og formuleringer. Materialerne omfattede ikke bare vandopløselige tensidbaserede produkter, men også f.eks. faste stoffer, pulvere, aerosoler, viskøse væsker, cremer, farvede produkter og alkoholbaserede materialer. De 55 teststoffer dækkede hele spektret af Draize test MMAS scoringens 110 points skala.

23 af teststofferne var ingredienser (se tabel 3.1). *In vivo* øjenirritationsdata for disse ingredienser kom fra ECETOC's databank (ECETOC, 1992). Alle stoffer i denne databank er testet for øjenirritation i Draize testen efter OECD guideline 405, og banken omfatter både MMAS værdier og individuelle vævsscoringer for alle forsøgsdyrene. 20 af disse stoffer var fælles for COLIPA studiet og EU/Home Office undersøgelsen.

**Tabel 3.1**  
**Ingredienser anvendt som teststoffer i COLIPA studiet**

<i>Benzalkoniumklorid 1% *</i>	<i>Propylenglycol</i>
<i>Benzalkoniumklorid 5% *</i>	<i>Natriumhydroxid 1% *</i>
<i>Benzalkoniumklorid 10% *</i>	<i>Natriumhydroxid 10% *</i>
<i>n-butylacetat *</i>	<i>Natriumlaurylsulfat 3% *</i>
<i>Cetylpyrimidinbromid 6% *</i>	<i>Natriumlaurylsulfat 15% *</i>
<i>Cetylpyrimidinbromid 10% *</i>	<i>Natriumlaurylsulfat 30%</i>
<i>Ethylacetat *</i>	<i>Trikloreddikesyre 30% *</i>
<i>Glycerol *</i>	<i>Triton X-100 1%</i>
<i>Imidazol *</i>	<i>Triton X-100 5% *</i>
<i>Isopropanol *</i>	<i>Triton X-100 10% *</i>
<i>Methylethylketon *</i>	<i>Tween 20 *</i>
<i>Polyethylenglycol 400 *</i>	

\*: Fælles teststoffer i EU/Home Office og COLIPA studiet

De resterende 32 teststoffer var færdige kosmetiske produkter (se tabel 3.2). Produkterne var baseret på formuleringer, der har været undersøgt i en valideringsundersøgelse arrangeret af den amerikanske kosmetikindustri (CTFA), og produkterne var tidligere testet *in vivo*. Resultaterne fra disse forsøg var dog ikke sammenlignelige med ECETOC's data, da der var anvendt lokalbedøvelse

ved CTFA's *in vivo* testning. Nye Draize tests efter OECD guideline 405 blev derfor udført for COLIPA for de 32 produkter af Agencé du Medicament. *In vivo* testningen var udført med fra 1 til 6 dyr pr. teststof, og forsøgene var kun gentaget for 2 stoffer. Her blev gennemsnittet i scoringsværdier fra de udførte forsøg anvendt. Resultaterne fra dyreforsøgene kunne derfor ikke bruges til at bedømme variationen mellem forskellige forsøg i opnåede scoringsværdier.

**Tabel 3.2**  
**Formuleringer anvendt som teststoffer i COLIPA studiet**

<i>Cologne</i>	<i>Mascara</i>
<i>Emulsion antiperspirant</i>	<i>Fugtighedscreme med solfaktor</i>
<i>Eye liner</i>	<i>Mundvask</i>
<i>Hair conditioner</i>	<i>Shampoo - normal</i>
<i>Hair styling lotion</i>	<i>Brusegel</i>
<i>Hudrensemiddel</i>	<i>Tandpasta</i>
<i>Håndrensemiddel</i>	<i>Solcreme SPF 15</i>
<i>Håndsæbe</i>	<i>Solcreme</i>
<i>Hårfarvebasis formulering 1</i>	<i>Shampoo - 2 i 1</i>
<i>Hårfarvebasis formulering 2</i>	<i>Shampoo - antiskal</i>
<i>Hårfarvebasis formulering 3</i>	<i>Shampoo - baby</i>
<i>Rensegel</i>	<i>Pumpe-deodorant</i>
<i>Renseskum</i>	<i>Flydende sæbe</i>
<i>Rouge</i>	<i>Salve (hydrofil)</i>
<i>Øjenmakeup fjerner</i>	<i>Parfumeret hudlotion</i>
<i>Øjenskygge</i>	<i>Skrubbemiddel</i>

COLIPA studiet omfattede 2 faser: De 10 første stoffer blev testet i 1994, og de resterende 45 stoffer i 1995. Alle teststoffer blev testet blindt, og de var individuelt kodede, så de deltagende laboratorier ikke kunne sammenligne deres resultater før studiets afslutning. Med prøverne medfulgte et telefonnummer til en engelsk giftinformationscentral, så laboratorierne i tilfælde af uheld kunne blive informeret om prøvernes giftighed.

### 3.3 Alternative metoder

COLIPA undersøgelsen omfattede 10 alternative metoder, der anvendes i kosmetikindustrien som screeningstests:

#### *SKIN<sup>2</sup> ZK1200*

I testen anvendtes et kunstigt væv, der blev fremstillet af humane fibroblaster og keratinocytter fra forhud. Fibroblasterne blev dyrket på nylonnet, hvor de i løbet af en måned dannede et sammenhængende bindevæv. Ovenpå dette væv blev der udsæt keratinocytter, der voksede ud til et 3-4 cellelag tykt epithel. Væv på 1 x 1 cm blev fremstillet som et kommercielt testkit, *SKIN<sup>2</sup> ZK1200*, af firmaet Advanced Tissue Sciences i Californien. Kits med 24 vævsprøver blev pakket i bakker, hvor vævene lå under agar med næringsmedie, hvorefter de med fly blev transporteret til de deltagende laboratorier.

Teststofferne blev doserede ufortyndet på epithelsiden af vævene i faste koncentrationer på 25 µl eller 25 mg. Efter eksponering for teststoffer i perioder på op til 60 minutter, blev vævene vasket og cellerne vitalitet opgjort med MTT testen. MTT er en gul tetrazoliumfarve, der reduceres til et violet

formazansalt ved cellulære redoxprocesser. Mængden af violet farve i vævene blev målt fotometrisk efter ekstraktion af farven med isopropanol. For hvert teststof blev der beregnet en  $t_{50}$  værdi. Dette er den eksponeringstid, der medfører en 50% reduktion i MTT metabolismen. Testen havde en matematisk forudsigelsesmodel, der var etableret på basis af historiske data for 132 ingredienser og produkter. Alle stoffer og produkter kunne testes med metoden. Modellen er udgået af produktion efter COLIPA studiets afslutning.

#### *Neutralrødoptagelse*

Neutralrødoptagelsestesten kan udføres med en række forskellige celletyper, men den blev i COLIPA studiet udført med 3T3 musefibroblaster. Vitalfarvestoffet bliver tilbageholdt i lysosomerne i levende celler på grund af forskellen i pH mellem lysosomet og det omgivende cytoplasma. Mængden af neutralrødt, der optages i 3T3 celler, er direkte proportionalt med antallet af levende celler i kulturen. Koncentrationen af teststof, der giver en 50% reduktion i optagelsen af neutralrødt bestemmes efter ekstraktion af farvestoffet. Testen havde en matematisk forudsigelsesmodel, der kunne forudsige MMAS værdier op til 60. Modellen var etableret på basis af historiske data for 30 stoffer, hvoraf 29 var tensider. En begrænsning ved metoden er, at der ikke kan testes uorganiske syrer og baser, og at kun vandopløselige stoffer kan testes.

#### *Test med røde blodlegemer*

I testen opgøres skade på cellemembranen ved fotometrisk måling af udsivning af hæmoglobin fra røde blodlegemer. Desuden måles protein-denaturering ved at følge reduktion i mængden af oxyhæmoglobin fotometrisk. Koncentrationen af teststof, der giver 50% hæmolyse i forhold til totalt lysesede kontrolprøver bestemmes. Både den laveste koncentration af teststof, der giver denaturering, og den maksimale procentvise denaturering opgøres også. Testen havde en matematisk forudsigelsesmodel, der var etableret på basis af historiske data for 25 færdige produkter, og en klassifikationsmodel, der var etableret på grundlag af test af ca. 100 stoffer. Vandopløselige ketoner, primære alkoholer og uorganiske syrer og baser var ikke omfattet af den sidstnævnte model. Kun stoffer, der er vandopløselige eller blandbare med vand kan testes med metoden.

#### *Predisafe*

Metoden består af et kommercielt testkit, hvor der måles frigørelse af neutralrødt fra SIRC celler (hornhindeceller fra kanin), der har optaget farvestoffet inden testningen. Der anvendes en kort eksponeringstid (< 1 minut). Koncentrationen af teststof, der giver 50% frigørelse af neutralrødt bestemmes. Testen havde en matematisk forudsigelsesmodel, der var etableret på basis af historiske data for 47 færdige produkter. Alle stoffer og produkter kunne testes med metoden.

#### *Silikonemikrofysiometeret*

Metoden omfatter, at kulturer af fibroblastceller (L929 celler), der dyrkes på porøse filtre. Kulturerne placeres i et sensorkammer på et Cytosensor™ silikonemikrofysiometer, og ændring i cellernes metabolisme følges ved kontinuerlige målinger af den ekstracellulære pH. For hvert teststof konstrueres

der dosis-responskurver, der relaterer cellernes metabolisme til dosis af teststoffet. Koncentrationen af teststof, der giver 50% hæmning af cellernes metabolisme bestemmes. Testen havde en matematisk forudsigelsesmodel, der var etableret på basis af historiske data for 133 tensider og tensidbaserede produkter. Kun vandopløselige stoffer kan testes med metoden.

#### *CAMVA*

Metoden er baseret på analyse af ændringer af blodkar (VAscular changes) i fosterhinden (den chorioallantoide membran, CAM) af befrugtede hønsæg. Æggene inkuberes i 3 dage, og på 4. dagen bores et hul i skallen, og ca. 2,5 ml af æggehviden fjernes. Et rektangulært vindue på 2 x 2 cm skæres i skallen. På 10. dagen placeres en teflonring på fosterhinden, og 40 ml af teststoffet påføres indenfor ringen. Vinduet forsegles og ægget inkuberes igen. CAM responset evalueres efter 30 minutter, og blodophobning, kapillærindvækst og/eller tilstedeværelse af tomme blodkar (ghost vessels) bedømmes som et positivt respons. For hver fortynding af teststoffet beregnes procentdelen af æg, hvor fosterhinderne reagerer. Herefter beregnes den koncentration, der teoretisk giver reaktion i 50% af æggene med 95% konfidensgrænser. CAMVA havde to forskellige matematiske forudsigelsesmodeller. En model for alkoholer, der var opstillet på basis af historiske data for 4 stoffer, og en for andre stoffer, der var opstillet på basis af historiske data for 19 stoffer. Den sidstnævnte model udelukker resultater opnået med polyethylenglycol-fedtsyrer og beslægtede fede amidethanolamider.

#### *EYTEX*

Metoden bestod af et kommercielt testkit med et planteprotein, der kan blive uigennemsigtigt efter reaktion med et teststof. Reaktionen blev aflæst i et colorimeter. Ved brug af kalibratorstoffer var det muligt at sammenligne et resultat med en Draize skala og bestemme en EYTEX Draize ækvivalent. EYTEX testen havde en matematisk forudsigelsesmodel, der kunne forudsige MMAS værdier op til 99. Modellen var udviklet på baggrund af historiske data for 102 stoffer. Materialer med indhold af manganviolet, >5% urinstof, >3% aluminiumklorhydrat, >5% zinkoxid eller >40% tensider kunne ikke testes. EYTEX modellen er udgået af produktion efter COLIPA studiets afslutning.

#### *Pollenrørvækst*

Metoden er baseret på fotometrisk kvantificering af vækst af pollenrør fra tobaksplanter i et kulturmedium. Pollenkorn dyrkes i 18 timer ved tilstedeværelse af forskellige koncentrationer af teststoffet, og massen af nydannede pollenrør bestemmes ved farvning med Alcianblå. Koncentrationen af teststof, der giver en 50% reduktion i dannelsen af pollenrør bestemmes efter ekstraktion af farvestoffet. Testen havde en matematisk forudsigelsesmodel, der var etableret på basis af historiske data for 43 færdige produkter. Alle stoffer og produkter kunne testes med metoden.

#### *HET-CAM testen*

Metoden er i lighed med CAMVA testen baseret på analyse af ændringer af blodkar i fosterhinden af befrugtede hønsæg. Teststoffer påføres fosterhinden i 6 æg pr. dosis på 9. dagen efter befrugtningen. CAM responset evalueres efter 5 minutter for transparente teststoffer og efter 30 sekunder for andre stoffer.



Blødning, lysering af blodkar og koagulering bedømmes som et positivt respons. Resultatet omsættes til et irritationsindeks ved hjælp af et computerprogram. 5% Texapon SVF (et anionisk tensid) bruges som positiv kontrol. Testen havde en klassifikationsbaseret forudsigelsesmodel, der var etableret på basis af historiske data for 97 stoffer. En begrænsning ved metoden er, at der ikke kan testes stoffer, der klæber til fosterhinden eller som er stærkt farvede.

#### *Fluorescein lækage*

I testen anvendes konfluente nyreceller (Madin-Darby Canine kidney cells) der i 15 minutter eksponeres for 5 faste koncentrationer af teststoffet. Størrelsen af en eventuel skade på cellerne bestemmes ud fra mængden af fluorescein, der kan passere gennem cellemembranen gennem 4 timer. Koncentrationen af teststof, der giver anledning til en 20% beskadigelse af cellemembranen beregnes herefter. Testen havde en klassifikationsbaseret forudsigelsesmodel, der var etableret på basis af historiske data for 43 tensider og formuleringer. En begrænsning ved metoden er, at der kun kan testes vandopløselige stoffer.

### **3.4 Etablering af forudsigelsesmodeller**

En væsentlig del af COLIPA studiet var anvendelsen af forudsigelsesmodeller, der blev brugt til at give en klar standard for vurderingen af de alternative metoders påidélighed. En forudsigelsesmodel er en formel, der omsætter resultaterne fra en alternativ metode til en forudsigelse af toksicitet *in vivo*. Hvis en metode ikke har en forudsigelsesmodel, betragtes den ikke som tilstrækkeligt udviklet til at kunne valideres, fordi der endnu ikke er etableret en relation mellem *in vitro* og *in vivo* data (Bruner et al., 1996).

De forudsigelsesmodeller, der blev brugt i COLIPA studiet, blev udviklet på basis af historiske data fra de enkelte testsystemer. Forudsigelsesmodellerne indeholdt fire elementer, der er nødvendige for at kunne forudsige en *in vivo* effekt ud fra et *in vitro* resultat: 1. en beskrivelse af de typer af teststoffer, som forudsigelsesmodellen omfatter, 2. en beskrivelse af de typer af data, der kan indgå i modellen, og 3. en formel for omdannelse af *in vitro* data til en forudsigelse af et *in vivo* respons, og 4. en indikation for, hvor præcis en forudsigelse antages at være.

I COLIPA studiet blev påidéligheden af de alternative metoder vurderet på baggrund af, om resultaterne fra de alternative metoder var reproducerbare mellem laboratorierne, og om de opnåede data passede med den anvendte forudsigelsesmodel. Bedømmelse af alternative metoders evne til at forudsige øjenirritation blev som tidligere nævnt opstillet som det væsentligste formål ved planlægningen af COLIPA studiet. Efter undersøgelsens afslutning blev vurderingen af metodernes relevans anset for at være et separat spørgsmål. I den indledende rapport om studiet blev der derfor fokuseret mere på spørgsmål vedrørende metodernes reproducerbarhed, end på testenes potentiale for generelt at kunne erstatte Draize testen (Brantom et al., 1997).

I studiet blev der anvendt to forskellige typer af forudsigelsesmodeller. En gruppe af modeller var matematiske funktioner, der gav præcise forudsigelser af et bredt spektrum af Draize MMAS værdier ud fra *in vitro* data. En anden gruppe var ikke-kontinuerte klassifikationsmodeller, der kun relaterede *in vitro*

data til forskellige irritationsklasser. Anvendelse af matematiske modeller giver principielt mulighed for at kunne forudsige alle typer af det ønskede *in vivo* respons, f.eks. MMAS værdier fra Draize testen. Begge typer af modeller kan have begrænsninger, f.eks. i form af stofklasser og tilstandsformer, der ikke kan testes eller typer af respons (f.eks. stærk øjenirritation), der ikke kan forudsiges. Der kan ikke opnå en præcis forudsigelse af et stofs *in vivo* toksicitet med klassifikationsmodeller, men de bruges til en grovere opdeling af stoffer i f.eks. mildt, moderat og stærkt irriterende stoffer. Klassifikationsmodeller giver derfor generelt en svagere forudsigelse af *in vivo* responset end matematiske modeller, og baggrunden for at opstille denne modeltype er som regel, at der ved indledende undersøgelser ikke har kunnet påvises en kontinuert sammenhæng (korrelation) mellem *in vitro* data og et *in vivo* respons.

Hidtil har studier vedrørende validering af alternative metoder primært været koncentreret om indledende analyser af korrelationer mellem *in vitro* og *in vivo* data. COLIPA studiet er en af de første undersøgelser hvor forudsigelsesmodeller, opstillet på grundlag af tidligere undersøgelser, bruges til at bedømme metodernes evne til at forudsige *in vivo* responset og reproducere korrelationer opnået med historiske data. Studiet belyser derfor en række fundamentale problemer ved anvendelse af modellerne, der vil kunne bruges ved planlægning af fremtidige valideringsstudier. En sideløbende anvendelse af matematiske modeller og klassifikationsmodeller gør det meget vanskeligt at foretage en overordnet sammenligning af de alternative metoders relevans. Det er f.eks. nødvendigt at anvende forskellige statistiske metoder ved behandlingen af resultater fra de to ovennævnte grupper af modeller, og resultaterne kan ikke umiddelbart sammenlignes (se afsnit 3.6 om statistiske metoder). I COLIPA studiet vanskeliggjorde en række andre forhold også en sammenligning af de forskellige metoder. Grundlaget for sammenligning af *in vitro* og *in vivo* resultater var f.eks. ikke standardiseret. Således blev der ved analyse af HET-CAM testens resultater anvendt individuelle vævsscoringer fra Draize testene, hvorimod MMAS værdier blev brugt for alle de øvrige *in vitro* systemer. Tillige var det tilfældet, at alle undersøgelsens teststoffer kun blev testet i 3 ud af de 10 alternative metoder (SKIN<sup>2</sup> ZK1200 modellen, Predisafetesten og testen for pollenrørvækst). Ved brug af de øvrige metoder blev det tilladt at undlade at teste ingredienser og produkter, der ville være vanskelige at håndtere. 50% af teststofferne måtte f.eks. frasorteres ved brug af silikonemikrofysiometret.

Teoretisk set vil alternative metoder med kontinuerte matematiske forudsigelsesmodeller uden alvorlige begrænsninger have det bedste potentiale for at kunne erstatte et dyreforsøg. Metoder, der kun har kunnet opnå en klassifikationsmodel, vil derfor kunne undværes, med mindre metoden potentielt kan forudsige et helt unikt og væsentligt biologisk respons. Begrundelsen for valg af modeltype forekommer tillige uklar, når biologisk nært beslægtede systemer som f.eks. CAMVA og HET-CAM testen har forskellige typer af forudsigelsesmodeller. Endvidere vil der kunne opstilles et minimumskrav til antallet af stoffer, der danner basis for opstilling af en forudsigelsesmodel. Dette var ikke tilfældet i COLIPA studiet, hvor antallet af stoffer bag modellerne varierede fra 4 (CAMVA testens model for alkoholer) til 132 (SKIN<sup>2</sup> ZK1200 metodens model for generel forudsigelse af øjenirritation). Endelig vil det være relevant at begrænse en valideringsundersøgelse, der designes til at kunne anvendes til sammenligning af forskellige tests, til at omfatte metoder hvor alle eller flertallet af de valgte

teststoffer kan undersøges.

### 3.5 Indsamling af *in vitro* resultater

BIBRA stod både for forsendelse af prøver til laboratorierne, og indsamling af data samt den statistiske analyse af resultaterne. Disse funktioner blev udført i henhold til GLP. De deltagende laboratorier indsendte de opnåede *in vitro* resultater på standardiserede datablanketter direkte til BIBRA. Blanketterne indeholdt information om prøvernes koder, en beskrivelse af testmaterialet, *in vitro* rådata og beregnede *in vivo* data. De indsendte datas kvalitet blev kontrolleret af BIBRA's kvalitetssikringsenhed (Brantom et al., 1997).

### 3.6 Statistiske metoder

BIBRA udførte den statistiske analyse af de opnåede data. Der brugtes forskellige grupper af tests afhængigt af, om *in vitro* metoderne havde matematiske eller klassifikationsmæssige forudsigelsesmodeller.

For *in vitro* metoder med matematiske forudsigelsesmodeller blev metodernes relevans bedømt ved analyser af lineære korrelationer mellem forudsagte Draize MMAS værdier på baggrund af *in vitro* data og observerede Draize MMAS værdier. Det blev desuden undersøgt i hvilket omfang de forudsagte Draize MMAS værdier passede med de opstillede forudsigelsesmodeller. Det blev beregnet hvor stor en procentdel af observationerne, der for de enkelte tests faldt udenfor forudsigelsesmodellens 95% og 99% forudsigelsesintervaller. BIBRA beregnede også summen af kvadrerede forskelle mellem forudsagte og observerede Draize MMAS værdier. Disse summer kan bruges til at vurdere, hvor gode forudsigelser der er opnået. *In vitro* metodernes evne til at forudsige *in vivo* respons blev tillige vurderet ud fra Altman/Bland diagrammer af forskellen mellem forudsagte og observerede Draize MMAS værdier for hvert laboratorium.

Med henblik på at vurdere de ovennævnte metoders reproducerbarhed mellem forskellige laboratorier blev der beregnet gennemsnit, standardafvigelser og variationskoefficienter (CV) for utransformerede og logaritmisk transformerede *in vitro* data. Desuden blev der udarbejdet Altman/Bland diagrammer af forskellen mellem forudsagte Draize MMAS værdier for laboratorierne parvist.

Både relevansen og reproducerbarheden af *in vitro* metoder med klassifikationsmodeller blev vurderet med kappa statistik. Udtrykket for overensstemmelse i klassificering, kappa ( $\kappa$ ), har et maksimum på 1, når overensstemmelsen er perfekt. En kappaværdi på 0 betyder, at overensstemmelsen ikke er bedre end hvad der kan være tilfældigt, og negative værdier viser en ringere overensstemmelse end dette. I kappa statistikken kan der både bruges ensartet vægtning af resultaterne eller forskellig vægtning. Ved ensartet vægtning tages der ikke hensyn til graden af uoverensstemmelse mellem resultaterne, idet alle uoverensstemmelser behandles ens. En kappa med lineær vægtning tildeler større betydning til effekter af uoverensstemmelser på mere end 2 klassificeringer, og en kvadratisk vægtet kappa tillægger sådanne uoverensstemmelser endnu større betydning. En kappa-værdi kan ikke sammenlignes med en korrelations-koefficient.

### 3.7 Yderligere analyser af COLIPA undersøgelsens resultater

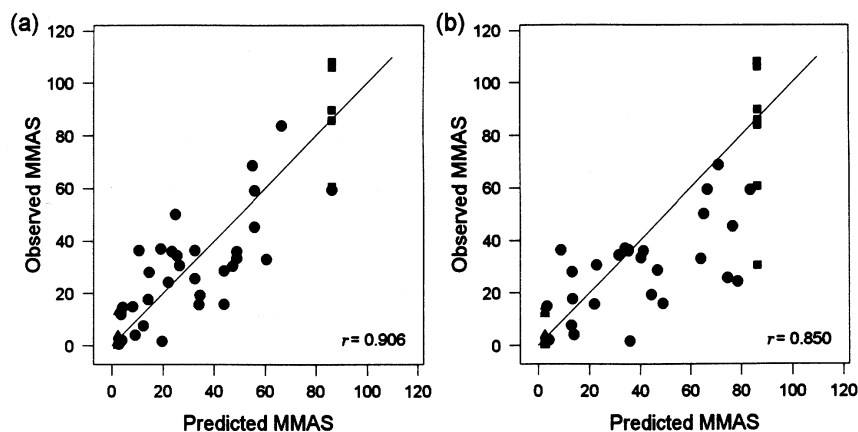
En af konklusionerne i den foreløbige rapport om COLIPA undersøgelsen var, at der burde foretages en yderligere, mere detaljeret undersøgelse af de data, der blev opnået i dette og i EU/Home Office studiet. For SKIN<sup>2</sup> ZK1200 modellen fandtes der yderligere data for 20 af teststofferne fra et tredje laboratorium (Procter & Gamble, Cincinnati, USA). Disse data var opnået i forbindelse med EU/Home Office studiet for kodede teststoffer, og de var indsendt blindt til BIBRA. I den samlede vurdering af data fra SKIN<sup>2</sup> ZK1200 modellen blev testens reproducerbarhed genstand for videre analyse og forudsigelsen af individuelle vævsscoringer i Draize testen blev evalueret. Endvidere blev forskelle i resultater mellem COLIPA laboratorierne og Procter & Gamble bedømt (Southee et al., 1999).

### 3.8 Resultater

#### SKIN<sup>2</sup> ZK1200 modellen

Resultater opnået med SKIN<sup>2</sup> ZK1200 modellen var velegnede til at forudsige Draize test MMAS værdier for alle typer af testmaterialer. Testen blev gennemført af 2 laboratorier: Microbiological Associates, Skotland (laboratorium 21), og Institut for Fødevarerikkerhed og Toksikologi, Danmark (laboratorium 23). Et tredje laboratorium (Lab. Simon, Belgien, laboratorium 22) deltog i projektets første fase med test af 10 stoffer. Der blev opnået meget gode korrelationer mellem forudsagte og observerede Draize MMAS værdier (se figur 3.2 og tabel 3.3), og alle ingredienser og produkter kunne testes i modellen.

Figur 3.2



Sammenhæng mellem forudsagte Draize test MMAS værdier med SKIN<sup>2</sup> ZK1200 i laboratorium 21 (a) og 23 (b) og observerede MMAS værdier. Fra Brantom et al., 1997.

Der fandtes også gode korrelationer mellem forudsagte og observerede Draize MMAS værdier, når teststofferne blev opdelt på ingredienser og formuleringer (se tabel 3.4). Endvidere blev der opnået gode korrelationer ( $r > 0.8$ ) mellem *in vitro* data ( $t_{50}$  værdier) og respons i de individuelle væv i Draize testen. Derimod blev der kun opnået en mådelig god korrelation til tid for recovery af Draize test responset ( $r = 0.66$ ) (Southee et al., 1999).

En god lineær korrelation ( $r = 0.87$ ) mellem logaritmisk transformerede  $t_{50}$  værdier fundet med SKIN<sup>2</sup> ZK1200 modellen og scoringsværdier fra low-volume øjenirritationstesten med kaniner er tidligere opnået ved test af 36 kemikalier og produkter. Desuden var  $t_{50}$  værdier for 20 andre stoffer i god overensstemmelse med Draize test MAS værdier (Osborne *et al.*, 1995).

**Tabel 3.3**

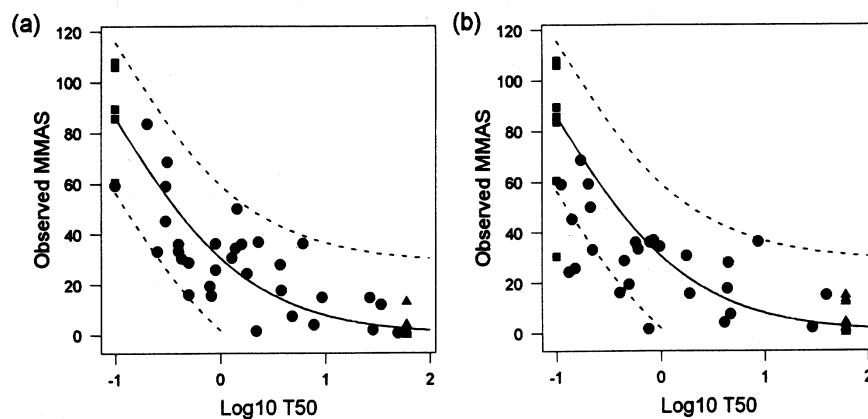
**Korrelationer mellem forudsagte og observerede Draize test MMAS værdier i COLIPA studiet for alle teststoffer**

Metoder	Pearsons lineære korrelations koefficienter*	Antal af testede stoffer**
SKIN <sup>2</sup> ZK1200	$r = 0.88$	55
Neutralrødtoptagelse	$r = 0.29$	52
Test med røde blodlegemer	$r = 0.68$	32
Predisafe	$r = 0.65$	55
Silikone mikrofysiometret	$r = 0.67$	28
CAMVA	$r = 0.61$	47
EYTEX	$r = 0.39$	38
Pollenrørvækst	$r = 0.43$	55

\*: Gennemsnit af korrelationskoefficienter opnået i de deltagende laboratorier.

\*\* : Gennemsnit af stoffer, der blev testet i de deltagende laboratorier.

**Figur 3.3**



Sammenhæng mellem  $t_{50}$  værdier opnået med SKIN<sup>2</sup> ZK1200 i laboratorium 21 (a) og 23 (b) og observerede Draize test MMAS værdier. Den ubrudte linie viser metodens forudsigelsesmodel, og de stiplede linier viser modellens 95% konfidensintervaller. Fra Brantom et al., 1997.

De resultater, der blev opnået med SKIN<sup>2</sup> ZK1200 modellen passede desuden godt med testens forudsigelsesmodel (se figur 3.3). SKIN<sup>2</sup> ZK1200 modellen var den metode i COLIPA undersøgelsen, der bedst var i stand til reproducere forudsigelsesmodellen. Ingen datapunkter faldt udenfor modellens 95% eller 99% intervaller for laboratorium 21. Kun 5.3% af datapunkterne faldt udenfor 95% intervallet for laboratorium 23. Disse punkter repræsenterede 7 stoffer, hvis irriterende potentialer blev let til moderat overvurderede (Brantom et al., 1997).

#### Tabel 3.4

#### Korrelationer mellem forudsagte og observerede Draize test MMAS værdier i COLIPA laboratorierne, der testede SKIN<sup>2</sup> ZK1200 modellen

	Lab 21	Lab 23
Alle materialer (n = 55)	r = 0,906	r = 0,850
Formuleringer (n = 32)	r = 0,786	r = 0,727
Ingredienser (n = 23)	r = 0,896	r = 0,826

En positiv kontrol (natriumlaurylsulfat, 2%) var inkluderet i alle de forsøg, der blev udført med SKIN<sup>2</sup> ZK1200 modellen (se tabel 3.5). Der fandtes et lavere respons med den positive kontrol i laboratorium 23 end i laboratorium 21, og dette bekræfter at der var en tendens til at overvurdere irritationspotentialer i laboratorium 23. Variationen i målingerne af de positive kontrolprøver var markant lavest i laboratorium 21, som også havde udført flest forsøg. Laboratorium 23 havde importeret flere væv pr. forsendelse, og dermed opnået en betydelig besparelse på omkostningerne. De ovennævnte resultater tyder dog på at det vil være bedst at begrænse antallet af væv, der håndteres pr. forsøgsrunde (Southee et al., 1999).

Ved forudsigelsen af Draize MMAS værdier med SKIN<sup>2</sup> ZK1200 modellen anvendes logaritmisk transformeret  $t_{50}$  værdier, og dette mindsker datavariationen betydeligt i forhold til brug af utransformeret data.

Altman/Bland diagrammerne over forskellene mellem forudsagte og observerede MMAS værdier viste standardafvigelser på under 15% i begge laboratorier. Der fandtes 100% overensstemmelse i  $t_{50}$  værdier opnået i de to laboratorier for 23 af teststofferne med værdier på under 0.1 minut eller over 60 minutter, og en lineær korrelation med  $r = 0.97$  mellem de opnåede  $t_{50}$  værdier for de resterende teststoffer. Ved statistisk analyse fandtes dog en marginal signifikant forskel ( $p = 0.06$ ) i de parrede sammenligninger af laboratoriernes  $t_{50}$  værdier. De observerede forskelle skyldes formentlig, at der var en tendens til at overvurdere *in vivo* responset for de mest reaktive teststoffer i laboratorium 23. Det blev konkluderet, at det var muligt at opnå en god konsistens i resultater opnået i forskellige laboratorier med metoden (Southee et al., 1999).

**Tabel 3.5**  
Positive og negative kontrolprøver med SKIN<sup>2</sup> ZK1200 metoden

	Positiv kontrol (% viabilitet)	Negativ kontrol (OD <sub>540</sub> )
Lab 21 (n = 29)	49,0 ± 11,2 (CV: 22,9%)	1.705 ± 0.229 (CV: 13,4%)
Lab 23 (n = 14)	29,4 ± 14,2 (CV: 48,3%)	1.404 ± 0.138 (CV: 9,8%)

Negative kontrolprøver med udoserede væv var også inkluderet i alle forsøgene (se tabel 3.5). De gennemsnitlige MTT målinger for udoserede kontrolvæv var noget højere i laboratorium 21 end i laboratorium 23. Dette tyder på at vævenes viabilitet ved modtagelsen var højest i laboratorium 21. Variationen i MTT målingerne af de negative kontrolprøver var lav i begge laboratorier (< 15% CV) (Southee et al., 1999).

**Tabel 3.6**  
Korrelationer mellem forudsagte og observerede Draize test MMAS værdier i COLIPA studiet for 23 formuleringer

Metoder	Pearsons lineære korrelations koefficienter *	Antal af testede formuleringer **
SKIN <sup>2</sup> ZK1200	$r = 0.86$	23
Neutralrødtoptagelse	$r = 0.63$	21
Test med røde blodlegemer	$r = 0.81$	11
Predisafe	$r = 0.54$	23
Silikone mikrofysiometret	$r = 0.72$	19
CAMVA	$r = 0.62$	23
EYTEX	$r = 0.25$	14
Pollenrørvækst	$r = 0.53$	23

\*: Gennemsnit af korrelationskoefficienter opnået i de deltagende laboratorier.

\*\* : Gennemsnit af formuleringer, der blev testet i de deltagende laboratorier.

Ved analyse af resultater opnået med SKIN<sup>2</sup> ZK1200 modellen opnåedes yderligere information om variationen i systemet. Resultater opnået hos Procter & Gamble undervurderede signifikant irritationspotentialt af de 20 fælles teststoffer i forhold til COLIPA undersøgelsens laboratorier ( $p < 0.001$ ). I Danmark blev der set en mindre tendens til overvurdering af resultaterne i forhold til resultaterne fra ledelaboratoriet i Skotland. Samlet tyder dette på at følsomheden af SKIN<sup>2</sup> ZK 1200 modellen kan være stærkt påvirket af forskelle i transporttid for vævene. Under transporten havde vævene dårlige livsbetingelser, idet de kun kunne optage næring fra den overliggende agar (Southee et al., 1999).

En anden mulig årsag til de fundne forskelle mellem laboratorierne er, at der blev anvendt forskellige doseringsmåder. På Procter & Gamble blev prøverne først sat på et dækglas, der derefter påførtes vævene. COLIPA laboratorierne testede derimod de fleste af de 20 fælles teststoffer ved direkte påføring på vævene. Teststof kan være tabt ved den indirekte doseringsmåde, og toksiciteten af flygtige stoffer er formentlig højest ved den direkte doseringsmåde. Tillige blev de største forskelle i resultater mellem COLIPA laboratorierne fundet for meget reaktive stoffer og ofte for materialer med stort indhold af opløsningsmidler. Småforskelle i doseringsteknik vil kunne have stor indflydelse på resultater for sådanne stoffer (Southee et al., 1999).

#### **Tabel 3.7**

#### **Korrelationer mellem forudsagte og observerede Draize test MMAS værdier i COLIPA studiet for 32 ingredienser**

Metoder	Pearsons lineære korrelationskoefficienter*	Antal af testede ingredienser**
SKIN <sup>2</sup> ZK1200	$r = 0.76$	32
Neutralrødoptagelse	$r = 0.21$	31
Test med røde blodlegemer	$r = 0.94$	21
Predisafe	$r = 0.87$	32
Silikone mikrofysiometret	$r = 0.84$	9
CAMVA	$r = 0.64$	24
EYTEX	$r = 0.33$	24
Pollenrørvækst	$r = 0.78$	32

\*: Gennemsnit af korrelationskoefficienter opnået i de deltagende laboratorier.

\*\* : Gennemsnit af ingredienser, der blev testet i de deltagende laboratorier

#### *Andre metoder med matematiske modeller*

Der blev opnået ringe til moderat gode korrelationer mellem forudsagte og



observerede Draize MMAS værdier for de øvrige *in vitro* metoder med matematiske forudsigelsesmodeller (se tabel 3.3). Det fulde sæt af testmaterialer kunne foruden i SKIN<sup>2</sup> ZK1200 modellen kun testes i Predisafe testen og i pollenrøstesten, og derfor giver de opnåede korrelationer formentlig et overestimat af de fleste af metodernes evne til at forudsige Draize MMAS værdier for en blandet gruppe af ingredienser og formuleringer. Generelt var metodernes evne til at reproducere en forudsigelsesmodel relativt dårlig, og betydeligt mere end 5% af datapunkterne faldt udenfor modellernes 95% konfidensintervaller. Der fandtes der markante tendenser til at overestimere *in vivo* responset med CAMVA testen og med silikonemikrofysiometret, mens der med testen med røde blodlegemer var en tendens til at undervurdere *in vivo* responset.

For flere af de øvrige *in vitro* metoder med matematiske forudsigelsesmodeller fandtes der betydeligt bedre korrelationer mellem forudsagte og observerede Draize MMAS værdier, når teststofferne blev opdelt på ingredienser og formuleringer (se tabel 3.6 og 3.7). Specielt fandtes der gode korrelationer for begge typer af testmaterialer med testen med røde blodlegemer, men kun 11 ud af 23 formuleringer kunne testes. Der opnåedes endvidere en god forudsigelse af ingrediensers irriterende potentiale med Predisafe testen og med silikonemikrofysiometret, men med den sidstnævnte metode blev kun 9 ingredienser testet.

**Tabel 3.8**  
**Forskellige mål for reproducerbarheden af alternative metoder med matematiske forudsigelsesmodeller mellem laboratorier.**

Metoder	CV%*	CV% log transformeret	Altman/Bland S. D.
SKIN <sup>2</sup> ZK1200	25,0 ± 32,4	9,6 ± 106,6	13,7
Neutralrødoptagelse	37,3 ± 29,8	7,5 ± 6,8	3,8-6,9
Test med røde blodlegemer	41,7 ± 32,6	6,7 ± 4,9	1,8-6,8
Predisafe	31,8 ± 30,2	49,6 ± 139,9	9,1-9,9
Silikone mikrofysiometret	-	3,0 ± 4,0	7,5
CAMVA	-	114,0 ± 480,7	14,7
EYTEX	9,9 ± 10,3	-	10,7-13,2
Pollenrøst	23,8 ± 17,9	3,9 ± 3,4	3,2-4,5

\*: *variationskoefficient*

Forskellige mål for for metodernes reproducerbarhed mellem laboratorier ses i tabel 3.8. Ved brug af utransformerede variationskoefficienter sås en specielt stor variation (41.7%) for testen med røde blodceller, mens EYTEX testen

havde under 10% variation. Ved brug af logaritmisk transformerede variationskoefficienter så en særdeles stor variation (114%) for CAMVA testen og en stor variation for Predisafe testen (49.6%), mens de øvrige tests havde under 10% variation. Ud fra de opnåede standardafvigelser fra Altman/Bland diagrammerne, havde CAMVA, EYTEX og SKIN<sup>2</sup> ZK1200 metoderne over 10% datavariation, mens de øvrige metoder havde under 10% variation.

I den prelimære rapport om undersøgelsen blev der kun udført en fuld evaluering af reproducerbarheden mellem laboratorier for metoder, der blev udført af 4 laboratorier og derover. Kun neutralrødtoptagelsestesten og testen med røde blodceller kunne opfylde dette krav, og disse tests blev bedømt til at have en rimelig reproducerbarhed (Brantom et al., 1997).

Det er fra flere sider blevet fremhævet som nødvendigt at evaluere reproducerbarheden af de anvendte metoder i mindst 4 laboratorier (Bruner et al., 1996, Earl et al., 1997). Det bør dog nævnes, at flere nye dyreforsøg i de senere år er blevet optaget i OECD guidelines for testning helt uden krav om vurdering af forsøgenes reproducerbarhed.

#### *Metoder med klassifikationsmodeller*

Resultaterne af HET-CAM testens og fluorescein lækage testens evne til at forudsige *in vivo* øjenirritation i forhold til forskellige irritationsklasser er opsummeret i tabel 3.9. HET-CAM testens evne til at forudsige *in vivo* responset var dårlig, og de deltagende laboratorier fejlklassificerede mindst 7 ud af de 55 teststoffer med 2 eller flere kategorier. HET-CAM testen var bedst til at påvise stærkt irriterende stoffer, men metoden undervurderede et betragteligt antal stoffers og materials irriterende potentialer. Resultater opnået med fluorescein lækage testen var bedre til at forudsige Draize test responset. Kun 40 ud af de 55 stoffer og materialer blev imidlertid testet, og kun 4 af stofferne blev klassificeret som moderat irriterende. Dette medfører, at modellen ikke kan evalueres på grund af for få data. Endvidere kunne testens forudsigelsesmodel ikke skelne mellem Draize MMAS værdier på mellem 30 og 110.

**Tabel 3.9**

**Evnen til at forudsige Draize test irritationsklasser for *in vitro* metoder med klassifikationsmodeller**

	HET-CAM testen	Fluorescein lækage testen
$\kappa^*$	0.52	0.69
Vægtet $\kappa^*$ (lineær)	0.65	0.81
Vægtet $\kappa^*$ (kvadratisk)	0.75	0.89

\*: Gennemsnit af kappa værdier opnået i de deltagende laboratorier.

**Tabel 3.10**

**Reproducerbarhed af *in vitro* metoder med klassifikationsmodeller mellem forskellige laboratorier.**

	HET-CAM testen	Fluorescein lækage testen
$\kappa^*$	0.41	0.80
Vægtet $\kappa^*$ (lineær)	0.57	0.88
Vægtet $\kappa^*$ (kvadratisk)	0.57	0.94

\*: Gennemsnit af kappa værdier opnået i de deltagende laboratorier.

HET-CAM testens reproducerbarhed mellem forskellige laboratorier blev vurderet som moderat god i den lavere og højere del af irritationsskalaen, men dårligere i midten af skalaen. Fluorescein lækage testen blev kun anvendt i 2 laboratorier, men reproducerbarheden var bedre end HET-CAM testens (se tabel 3.10). De opnåede kappa-værdier kan som tidligere nævnt ikke sammenlignes med resultaterne af de statistiske metoder, der kan anvendes ved brug af matematiske modeller. Det vil derfor kun være muligt, at opnået reelt grundlag for at sammenligne fluorescein lækage testen med de øvrige alternative tests, hvis testens resultater behandles ved brug af de samme statistiske metoder. Dette er teoretisk muligt, men det blev ikke gennemført i COLIPA studiet.

### 3.9 Diskussion

Kort efter COLIPA studiets afslutning blev det konkluderet, at ingen af de anvendte *in vitro* metoder på baggrund af de preliminære resultater kunne leve op til undersøgelsens kriterier for både reproducerbarhed og relevans. På dette grundlag blev ingen af de alternative metoder betragtet som en pålidelig (valid) erstatning for Draize testen. Tre af de anvendte metoder - fluorescein lækage testen, testen med røde blodlegemer og SKIN<sup>2</sup>ZK1200 systemet - blev dog bedømt som enten reproducerbare eller relevante. Videre undersøgelser af data blev anbefalet med henblik på opstilling af nye forudsigelsesmodeller, som vil kunne anvendes i en fremtidig valideringsundersøgelse (Brantom et al., 1997).

De summariske konklusioner i den indledende rapport skyldes formentlig flere faktorer. COLIPA undersøgelsen blev planlagt med mindst 3 deltagende laboratorier pr. test for 90% af metoderne. I realiteten blev undersøgelsen gennemført med kun 2 laboratorier pr. test for flere af metoderne. Efter studiets afslutning blev det meddelt deltagerne, at metodernes reproducerbarhed kun ville blive vurderet, hvis testen var udført i mindst 4 laboratorier. 70% af de deltagende tests kunne ikke opfylde dette krav. De opnåede resultater vedrørende metodernes relevans blev kun diskuteret summarisk, hvor kriteriet for evaluering af reproducerbarheden ikke blev opfyldt. Derfor blev der kun foretaget en større evaluering af testen med røde blodceller og neutralrødt optagelsestesten. Desuden blev der ikke foretaget en overordnet sammenligning af de deltagende metoders resultater. Dette skyldes bl.a. at studiet var designet til at hvile på brugen af forskellige typer af forudsigelsesmodeller. Endvidere var grundlaget for sammenligning af *in vitro* og *in vivo* resultater ikke standardiseret.

En videre analyse af undersøgelsens resultater med inddragelse af flere data, og med større vægt på metodernes relevans end deres reproducerbarhed har givet anledning til mere detaljerede konklusioner: SKIN<sup>2</sup> ZK1200 modellen blev vist at være meget god til forudsigelse af et bredt spektrum af Draize

MMAS værdier, og desuden kunne testen forudsige individuelle vævsreaktioner både for formuleringer og ingredienser. Metoden må antages at have en meget bred anvendelsesmulighed, fordi alle 55 testmaterialer kunne testes, og tidligere er der opnået lovende resultater ved test af 132 materialer med et meget bredt spektrum af irriterende potentialer. SKIN<sup>2</sup> ZK1200 modellen var den test i COLIPA undersøgelsen, der bedst var i stand til reproducere metodens forudsigelsesmodel. SKIN<sup>2</sup> ZK1200 metodens reproducerbarhed må betragtes som rimelig, og den er i lighed med de fleste andre *in vitro* metoder betydeligt mere reproducerbar end Draize testen (se tabel 3.11).

Ingen af de øvrige alternative metoder i COLIPA studiet var velegnede til forudsigelse af et bredt spektrum af Draize MMAS værdier for blandede stoffer og produkter. Flere metoder fandtes dog velegnede til forudsigelse af *in vivo* responset af ingredienser, specielt af vandopløselige stoffer som tensider. Begge tests med fosterhinder fra hønseæg (CAMVA og HET-CAM testene) havde en meget ringe reproducerbarhed mellem laboratorier. Reproducerbarheden af de øvrige metoder mellem forskellige laboratorier var god. Neutralrødt optagelsestesten og testen med røde blodceller blev udført af mindst 4 laboratorier, og disse tests blev bedømt til at have en rimelig reproducerbarhed (Brantom et al., 1997).

Den mest egnede undersøgelse til belysning af Draize testens variation blev publiceret af Weil og Scala i 1971. 9 stoffer blev testet for øjenirritation i 24 forskellige laboratorier, og der blev gennemført en tilsvarende undersøgelse for hudirritation. Der fandtes generelt en meget stor variation i de opnåede dyreforsøgsresultater, både for resultater opnået med det samme stof i enkelte laboratorier (fra 17 til 117 % CV) og for resultater opnået i forskellige laboratorier (fra 42 til 59 % CV) (Earl et al., 1997). Undersøgelsens konklusion var, at Draize testene for øjen- og hudirritation på grund af den ekstreme variation i resultaterne ikke kan anvendes til en konsistent klassifikation i praksis (Weil og Scala, 1971). Dyreforsøgene er dog fortsat med stort set uændret design, og i dag antages variationen at være af samme størrelsesorden som for 25 år siden (Earl et al., 1997). Den særdeles store usikkerhed drager den gældende regulatoriske praksis med krav om dyreforsøg for øjen- og hudirritation i tvivl, og den vanskeliggør i høj grad en validering af alternative metoder ved sammenligning med data fra dyreforsøgene.

### ***Tabel 3.11***

#### **Historiske data om reproducerbarhed af *in vitro* metoder og Draize testen i individuelle laboratorier (fra Bruner et al., 1996)**

Metode	Positiv kontrol	n	CV %
BCOP	acetone	119	12.0
Microtox	fenol	123	19.4
Silikone mikrofysiometer	SLS	163	15.5
Neutralrødt optagelse	SLS	191	21.7
CORROSITEX	NaOH	44	9.5
SKIN <sup>2</sup> ZK1200	SLS	44	26.0
Draize øjenirritationstest (Weil & Scala, 1971)			40 - 60

*SLS: Natriumlaurylsulfat. NaOH: Natriumhydroxid.*

COLIPA studiets væsentligste resultat var, at studiet var den første store blindundersøgelse af blandede stoffer og produkter, hvor en alternativ test - SKIN<sup>2</sup> ZK1200 metoden - potentielt er vist at kunne erstatte dyreforsøg for akut øjenirritation. Producenten af vævsmodellen, firmaet Advanced Tissue Science i Californien, besluttede dog kort efter COLIPA studiets afslutning at standse produktionen og salget af SKIN<sup>2</sup> modellerne. Beslutningen var begrundet i problemer med at opnå et tilstrækkeligt stort salg af vævsmodellerne på grund af deres høje pris, og lange udsigter for en generel accept af metoderne på trods af meget gode resultater fra flere forskellige valideringsundersøgelser. Firmaet satser nu udelukkende på at fremstille rekonstruerede væv til transplantation.

Standsningen af produktionen af de patenterede SKIN<sup>2</sup> modeller er et stort tilbageslag for arbejdet med udvikling af anvendelige alternative metoder, specielt til påvisning af lokalirriterende stoffer. De allerede opnåede resultater med modellerne er dog ikke blevet uinteressante, idet en stor generel viden om mulighederne i anvendelsen af vævsmodeller i toksikologiske undersøgelser er opnået. Udvikling af ikke-kommercielle vævsmodeller er særdeles påkrævet, idet dette ville kunne løse problemerne vedrørende økonomiske begrænsninger ved anvendelse af patenterede rekonstruerede væv og afhængigheden af få fabrikanter af kommercielle testkits.

## 4 Test for recovery fra øjenirritation

I Draize testen bliver det kliniske respons i cornea (hornhinden), conjunctiva (øjets bindehinder) og iris (regnbuehinden) observeret indtil øjnene igen er normale eller op til 21 dage. Den lange observationsperiode medfører, at det kan bestemmes om teststoffer kan give permanente øjenskader. Dette forhold er kritisk i klassificeringen af øjenirriterende stoffer. I EU anvendes de gennemsnitlige scoringsværdier fra Draize testen for observationer foretaget 24, 48 og 72 timer efter doseringen med teststof som grundlag for klassificering af stoffer med R36 (irriterer øjnene) eller R41 (risiko for alvorlig øjenskade). Varigheden af øjenirriterende effekter indgår i bedømmelsen, således at tilstedeværelse af respons i øjets væv ved 72 timers observationen eller vedvarende farvning af øjet udløser R41 klassifikationen (EU, 1996). Varigheden af øjenirriterende effekter indgår også i klassificeringen af øjenirriterende stoffer i USA, hvor tilstedeværelse af effekter på øjets væv i mere end 7 eller 21 dage tildeles afgørende vægt (Gupta et al., 1993).

Stort set alle *in vitro* systemer for øjenirriterende effekter har hidtil kun været designet til at bedømme den akutte effekt af teststoffer, og tidligere har det ikke været muligt at demonstrere en overensstemmelse mellem effekten af teststoffer over tid i *in vitro* systemer og resultater fra øjenirritationsforsøg. Institut for Fødevarerikkerhed og Toksikologi har som led i projektet udviklet en model til vurdering af bedring (recovery) fra øjenirritation ved brug af SKIN<sup>2</sup> ZK1200 væv (Espersen et al., 1997). Der blev testet 9 materialer med *in vivo* data fra ECETOCs øjenirritations databank (ECETOC, 1998) (se tabel 4.1).

### 4.1 Metode

Forsøgene blev udført med SKIN<sup>2</sup> ZK1200 væv. Recovery forsøgene blev udført før identiteten af COLIPA undersøgelsens teststoffer var kendt. For alle teststoffer blev der derfor udført et indledende forsøg for at etablere omtrentlige eksponeringstider for nedsættelse af cellernes viabilitet med 35-65%. Vævene blev i 1 til 3 replika påført en fast dosis på 25 ml af hvert teststof, og viabiliteten blev bestemt med MTT testen. I hovedforsøget blev triplika af væv påført 25 ml af hvert test-stof, og de eksponeringstider, der blev fastlagt i forforsøget, blev anvendt (se tabel 4.1).

Både udoserede og doserede væv blev MTT testet umiddelbart efter eksponeringen og efter inkubationsperioder, der svarede til Draize testens observationstider. Tiden for recovery af *in vitro* responset blev defineret som antallet af inkubationsdage, der var nødvendige for at opnå MTT test værdier i de eksponerede væv, som ikke var signifikant forskellige fra de tilsvarende kontrolvæv.

Udoserede væv og væv, der blev eksponeret for 1% benzalkoniumklorid, 3% natriumlaurylsulfat og 5% Triton X-100, blev undersøgt histologisk. Vævene blev fikseret i formalin, indstøbt i paraffin, snittet og farvet med haematoxylin og eosin. Vævene blev derefter undersøgt for antal af tilstedeværende lag af epithelceller og for dermal/epidermal degeneration og nekrose.

**Tabel 4.1**

Teststoffer	Dage for recovery <i>in vitro</i>	Dage for recovery <i>in vivo</i>	Median af dage for recovery <i>in vivo</i>	MMAS værdier
Benzalkoniumklorid 1%, 5 sekunder	>14	>14.6*	>15.5*	45.3*
Benzalkoniumklorid 10%, 1 sekund	>21	>21	>21	108.0
Natriumhydroxid 1%, 5 minutter	7	7.8	7	25.8
Isopropanol 100%, 1 minut	7	10	10	30.5
Methylacetat 100%, 30 minutter	7	15.5	15,5	39.5
Natriumlaurylsulfat 3%, 20 sekunder	1	5	5	16.0
Triton X-100 5%, 20 sekunder	1	7.8*	8.5*	33.1*
Cetylpyrimidinbromid 0.1%, 60 minutter	0	2.2	2	2.7
Glycerol 100%, 60 minutter	0	1.8	2	1.7

*Inkubationsperioder for SKIN<sup>2</sup> ZK1200 væv, der medførte tilbagevenden til kontrol MTT aktiviteter (dage for recovery in vitro) sammen med gennemsnits- og median værdier af dage for recovery in vivo og Draize test MMAS værdier. \*: Gennemsnit af 2 Draize tests.*

Der blev anvendt en ensidig variansanalyse til at sammenligne resultaterne af MTT målingerne. Analyser af lineær korrelation blev brugt til sammenligning af *in vitro* og *in vivo* recovery data.

## 4.2 Resultater

MTT aktiviteten af ueksponerede SKIN<sup>2</sup> ZK1200 væv i kultur var relativt stabil i op til 14 dage, men efter 21 dage sås et markant fald i vævenes viabilitet. Mildt irriterende stoffer som 100% glycerol og 0.1% cetylpyrimidinbromid inducerede ikke signifikante fald i vævenes cellulære viabilitet efter 60 minutters eksponering, men signifikante fald i MTT aktiviteterne blev observeret efter 2 til 3 dage i kultur. Væv, der blev eksponeret for 3% natriumlaurylsulfat og 5% Triton X-100, opnåede MTT aktiviteter på kontrolniveauet efter 1 dag, mens væv eksponeret for 1% natriumhydroxid, 100% isopropanol og 100% methylacetat opnåede kontrol MTT værdier indenfor 7 dage. Væv, der blev eksponeret for forskellige koncentrationer af benzalkoniumklorid, vendte dog ikke tilbage til kontrolniveauets viabilitet på en stabil måde. MTT aktiviteten af væv, der blev påført 1% benzalkoniumklorid i 5 sekunder var upåvirket efter 1 dag, men signifikant lavere MTT værdier blev opnået på dag 3, 7, 10 og 14. Væv, der blev eksponeret for den samme opløsning af stoffet i 10 sekunder opnåede ikke kontrol MTT niveauet gennem en 7 dage lang observationsperiode, og væv eksponeret for 10%

benzalkoniumklorid i 1 sekund opnåede ikke fuld viabilitet efter 21 dage

Antallet af inkubationsdage, der var nødvendige for at opnåen fuld viabilitet *in vitro* og Draize MMAS værdier for teststofferne er vist i tabel 4.1 sammen med gennemsnit og medianværdier af antallet af dage, der skulle til før responset i Draize testene forsvandt. Gode lineære korrelationer blev fundet mellem dage for recovery *in vitro* og dage for recovery *in vivo* ved brug af gennemsnitsværdier ( $r = 0.92$ ,  $p < 0.001$ ) og medianværdier ( $r = 0.91$ ,  $p < 0.001$ ).

Resultaterne af den histologiske analyse er vist i tabel 4.2. Der fandtes moderat til stærk degeneration af epitelceller og stromaceller i alle eksponerede væv efter 7 dage i kultur, og en speciel stærk reaktion fandtes i væv påført 1% benzalkoniumklorid. Der blev ikke observeret konsistent cellulær regeneration i vævene. I kontrolvævene sås en svagere cellulær degeneration på dag 7 og 10.

**Tabel 4.2**

**Histologisk analyse af antallet af lag af epitelceller og af cellulær degeneration og nekrose i SKIN<sup>2</sup> ZK1200 væv.**

	Antal epitelcellelag				
	Dag 0	Dag 1	Dag 3	Dag 7	Dag 10
Kontrol	3	1-3	1-3	0-3 <sup>+</sup>	1-3 <sup>(+)</sup>
Benzalkoniumklorid 1%, 5 sekunder	2-3	1-2	0-3	1-2 <sup>++++</sup>	1-3 <sup>++</sup>
Natriumlaurylsulfat 3%, 20 sekunder	1	1-2	0-2	1-3 <sup>++</sup>	n.d.
Triton X-100 5%, 20 sekunder	1-2	1-3	1-2	1-3 <sup>++</sup>	n.d.

*Histomorfologisk analyse af kontrol og eksponerede SKIN<sup>2</sup> ZK1200 væv.*

*Dermal/epidermal degeneration/nekrose: (+) spor, + mild, ++ moderat, +++ stærk, ++++ meget stærk. n.d.: ikke analyseret.*

### 4.3 Diskussion

Ved de histologiske analyser fandtes ingen konsistente cellulære ændringer i vævet før dag 7 efter doseringen, og dette tyder på at måling af MTT aktiviteten i vævene er en mere sensitiv markør for vævenes viabilitet end de histologiske resultater. Dette er i overensstemmelse med resultater af histologiske analyser og målinger af MTT reduktion i organkulturer med isoleret hud (van de Sandt & Rutten, 1995).

SKIN<sup>2</sup> ZK1200 modellen må på baggrund af de opnåede resultater anses for at have et godt potentiale for forudsigelse af vedvarende øjenirritation. Vævet viabilitet var stabil i 2 uger, og modellen kunne anvendes til at påvises en øget toksicitet af forskellige koncentrationer af benzalkoniumklorid gennem observationsperioden, der er parallel med stoffets *in vivo* effekter i hornhinden. Desuden fandtes der gode korrelationer mellem *in vitro* og *in vivo* recovery data. Dette er ikke tidligere opnået med andre alternative tests (Espersen et al.,



1997).

# 5 Validering af alternativer til øjenirritationstesten

Verden over er der gjort en meget stor indsats for at udvikle pålidelige alternativer til øjenirritationstests. Der er tidligere blevet gennemført en lang række mindre programmer for validering af metoderne, og senest er to store meget ressourcerkrævende studier blevet iværksat af EU/Home Office og COLIPA. Hensigten med de to sidstnævnte undersøgelser var at opnå en endelig afklaring af spørgsmålet, om *in vitro* metoder kan fungere som erstatninger for dyreforsøg til øjenirritationstestning. Denne proces har dog været betydeligt mere kompliceret end først antaget.

## *EU/Home Office undersøgelsen*

Undersøgelsen blev arrangeret af EU kommissionen og det britiske Home Office. Formålet med studiet var at undersøge, om 9 alternative metoder enten fuldstændigt eller delvist kunne erstatte Draize øjenirritationstesten (se tabel 5.1). I studiet fandtes der ingen signifikante overensstemmelser mellem resultater fra *in vitro* metoderne og Draize test MMAS værdier for 60 blandede kemiske stoffer. Der fandtes heller ikke pålidelige forudsigelser af Draize test resultater, når teststofferne opdeltes i vandopløselige stoffer ( $n = 30$ ) og ikke-vandopløselige stoffer ( $n = 18$ ). De alternative metoder kunne tillige ikke generelt identificere stærkt irriterende stoffer. Forudsigelser af øjenirritation foretaget med de alternative metoder havde en såring præcision, at resultaterne ikke skønnes at kunne blive af praktisk betydning. Det eneste positive resultat af undersøgelsen var, at der for flere metoder opnåedes relativt gode overensstemmelser til øjenirritationsdata for 12 tensider (Balls et al., 1995).

### **Tabel 5.1** **EU/Home Office studiets alternative metoder til test for øjenirritation**

---

<i>Test med røde blodlegemer</i>
<i>EYTEX metoden</i>
<i>BCOP (bovin corneal opacitet og permeabilitet) testen</i>
<i>HET-CAM testen</i>
<i>Fluorescein lækage testen</i>
<i>Test med isolerede kyllingeøjne</i>
<i>Test med isolerede kaninøjne</i>
<i>Silikone mikrofysiometret</i>
<i>Neutralrødoptagelses testen</i>

---

### *BGA studiet*

Udfaldet af EU/Home Office studiet var ikke fuldstændig uventet. Det tyske sundhedsministerium (BGA) arrangerede tidligere en valideringsundersøgelse med 136 blandede stoffer. Konklusionen var, at der ikke fandtes nogen overensstemmelse mellem *in vivo* øjenirritationsdata og resultater fra en celledet med neutralrødoptagelse. En test med fosterhinden fra hønseæg (HET-CAM testen) kunne dog identificere visse stærkt irriterende stoffer (Spielmann et al.,

1993). En database med resultater fra test af 200 stoffer er senere blevet etableret. Resultater for 9 stoffer blev ekskluderet på basis af en uacceptabel kvalitet af deres *in vitro* data. Tillige blev resultater for 48 kemikalier udelukket med den begrundelse, at originale Draize test resultater ikke var tilgængelige. Meget præcise individuelle data fra Draize tests udført efter OECD guideline 405 i op til 21 dage (ECETOC, 1998) er dog publiceret for mere end 25% af disse stoffer, og det ville være interessant at genoptage disse stoffer i BGA databasen. Ved analyse af data for de resterende stoffer blev det fundet, at ingen af de alternative tests kunne identificere R-41 stoffer med mere end 50% sensitivitet, og en acceptabel specificitet på mere end 80% kunne kun opnås med HET-CAM testen. For 129 af de 143 resterende kemikalier var resultater fra begge *in vitro* tests tilgængelige, og en lineær diskriminant analyse blev foretaget på kombinerede endpoints fra de to tests. Ved brug af denne procedure fandtes en falsk-negativ rate for R-41 stoffer på 29%, og en falsk-positiv rate for andre stoffer på 22%. Klassifikationen af R-41 stoffer blev dog forbedret ved hensyntagen til stoffernes opløselighed i vand og olie (Spielmann et al., 1996).

#### *CTFA studiet*

USA's kosmetikindustri (CTFA) gennemførte fra 1990 til 1996 et valideringsprogram med ca. 25 alternative metoder til øjenirritationstests. MAS værdier fra en LVET test blev brugt i de to første faser af programmet, og Draize MAS værdier blev brugt i den sidste fase. I begge tests bruges albino kaniner, men i LVET testen installeres der kun 10 µl af teststoffet i øjet (mod 100 µl i Draize testen). De datatransformationer, der passede bedst til *in vitro/in vivo* overensstemmelserne var generelt 2 eller 3 parameter logistiske modeller. HET-CAM testen, neutralrødtfrigørelsestesten og EYTEX™ testen gav relativt gode forudsigelser af 10 formuleringer med vand/alkohol (Gettings et al., 1996a). Ingen af de alternative tests kunne forudsige *in vivo* responset af 18 olie/vand emulsioner (Gettings et al., 1998). Flere tests gav dog gode forudsigelser af Draize MAS værdier for 25 tensidbaserede produkter (Gettings et al., 1996b). Variabiliteten af HET-CAM testen og neutralrødtfrigørelsestesten overskred konsistent *in vivo* forsøgenes variation, men de øvrige alternative tests var mere reproducerbare.

#### *IRAG evaluering*

EU/Home Office studiets resultater var også tidligere delvist bekræftet af en undersøgelse, der i 1993 blev foretaget af en arbejdsgruppe af eksperter fra forskellige regulatoriske myndigheder (IRAG). IRAG foretog en samlet vurdering af eksisterende data fra en lang række *in vitro* metoder på øjenirritationsområdet. 41 laboratorier verden over indsendte 55 datasæt opnået med 23 forskellige *in vitro* metoder til IRAG. Datasættene omfattede fra 9 til 133 teststoffer, der var anvendt i de deltagende laboratorier.

I IRAG studiet viste tests med fosterhinder fra hønseæg dårlige til moderat gode overensstemmelser til Draize test resultater for op til 93 blandede teststoffer. Med HET-CAM testen opnåedes dog gode overensstemmelser til *in vivo* data for tensidbaserede stoffer og produkter, og CAMVA testen viste den bedste forudsigelse for alkoholbaserede produkter (Spielmann et al., 1997). En test med øjne fra kaniner og en test med isolerede hornhinder fra koøjne (BCOP-testen) blev begge bedømt af IRAG til at have et potentiale for at identificere stærkt irriterende stoffer, men metoderne kunne ikke generelt forudsige Draize

test resultater. En test med isolerede øjne fra høns og en test med isolerede lenser fra koøjne viste begge gode overensstemmelser til *in vivo* data generelt. Datasættene for de to sidstnævnte tests var dog for begrænsede til, at metoderne fuldt ud kunne bedømmes (Chamberlain et al., 1997). IRAG bedømte tillige en række tests med cellekulturer til at have et potentiale for forudsigelse af Draize test resultater for vandopløselige stoffer ved normale pH værdier (Botham et al., 1997; Harbell et al., 1997). EYTEX testen viste en dårlig overensstemmelse til *in vivo* data for 454 kosmetiske indholdsstoffer og formuleringer, men forudsigelsen var god for enkelte grupper af materialer, f.eks. petrokemikalier og opløsningsmidler. Et Microtox assay med bakterier fandtes at være potentielt anvendeligt ved test af tensidbaserede produkter. SKIN<sup>2</sup> ZK1200 vævsmodellen blev vurderet som meget anvendelig til screening af kosmetiske produkter og husholdningsartikler, når måling af vævets viabilitet med MTT testen blev anvendt. Måling af udskillelse af prostaglandin E<sub>2</sub> som irritationsmarkør blev bedømt som problematisk. Yderligere udvikling af SKIN<sup>2</sup> ZK 1200 modellen blev anbefalet, mens brug af en vævsmodel udelukkende med fibroblaster ikke blev fundet hensigtsmæssig (Curren et al., 1997).

#### *JMHW/JCIA studiet*

Det japanske ministerium for sundhed og velfærd (JMHW) startede i 1991 en valideringsundersøgelse sammen med den japanske kosmetikindustri (JCIA), nationale forskningsinstitutioner, universiteter og test kit leverandører. 12 alternative metoder blev evalueret med 38 kosmetiske indholdsstoffer, og der blev gennemført Draize tests på de samme stoffer.

To SKIN<sup>2</sup> modeller blev anvendt, begge med en MTT test protokol. En ZK1100 fibroblastmodel blev brugt i 6-8 laboratorier. ZK1200 modellen (TEA) blev anvendt i 2 laboratorier for alle teststofferne, og i 6-7 laboratorier for 13 teststoffer. Ved brug af fibroblastmodellen blev vævene nedsænket i kulturmediet og eksponeret for teststoffer, opløst i mediet. Ved anvendelse af ZK1200 modellen, blev teststofferne påført epitelets overflade. Protokollen adskilte sig dog fra den, der blev brugt i COLIPA studiet, fordi teststofferne blev opløst eller suspenderet i mediet. T<sub>50</sub> værdier blev afledt fra MTT tidsrespons grafer. En relativt høj interlaboratorie variabilitet blev observeret med begge vævsmodeller med en gennemsnitlig CV på 44.5% (n=30) med ZK1100 modellen, og en CV på 61.9 % (n=9) på eksakte t<sub>50</sub> værdier opnået med ZK1200 modellen i 6-7 laboratorier. Der var dog en total overensstemmelse mellem de to laboratorier, der testede det fulde sæt af teststoffer for 17 stoffer med cut-off værdier, og en meget god korrelation (r=0.84) blev opnået ved reanalyse af log-transformede t<sub>50</sub> værdier for de 16 øvrige stoffer. Relativt dårlige lineære korrelationer blev fundet til Draize test MAS values både med *in vitro* data fra ZK1100 modellen (r=0.71) og ZK1200 modellen (r=0.63) (Kurishita et al., 1999). En reanalyse af data opnået med ZK1200 modellen i to laboratorier ved brug af den forudsigelsesmodel, der anvendtes i COLIPA studiet øgede forudsigelsen af *in vivo* resultaterne: Relativt gode korrelationer (r>0.78) blev herved opnået mellem observerede og forudsagte Draize test MMAS værdier. Udelukkelse af teststoffer, der var uforenelige med cellekulturmediet (syrer, baser og alkoholer med lav molekylvægt) øgede dog markant *in vitro/in vivo* korrelationen opnået med ZK1200 modellen (r=0.84) (Ohno et al., 1999).

En MATREX vævsmodel med humane fibroblaster dyrket i en kollagenmatrix blev også anvendt i undersøgelsen. Teststoffer blev påført ufortyndede til vævets

overflade i 24 timer, og koncentrationer, der gav 50% nedsættelse af MTT responset blev bestemt (EC50 værdier). Alternativet blev der anvendt en MATREX scoring med den laveste koncentration, der reducerede vævets viabilitet med 20-80%. 12 laboratorier brugte i alt modellen, men det fulde sæt af teststoffer blev kun testet i 3 laboratorier. Den interlaboratorielle reproducerbarhed var betydeligt bedre ved brug af MATREX scoringen (CV=9.6%, n=39) end ved brug af EC50 værdier (CV=34.6%, n=33)(Ohno et al., 1999). Ensartede korrelationer ( $r=0.67$ ) blev opnået mellem EC50 værdier, MATREX scoringer og Draize MAS værdier (Ohuchi et al., 1999). Det vil dog være muligt at øge metodens forudsigelsesevne betydeligt ved anvendelse af en non-lineær logistisk forudsigelsesmodel.

En dårlig korrelation ( $r=0.31$ ) til Draize MAS værdier blev opnået med EYTEX testen (Matsukawa et al., 1999). Tillige blev der opnået moderate *in vitro/in vivo* korrelationer med en test med måling af denaturering af isoleret hæmoglobin, og de interlaboratorielle CV'er overskred 240%. En god korrelation til Draize MAS værdier ( $r=0.91$ ) blev opnået ved brug af 50% denaturering som endpoint (Hatao et al., 1999). Korrelationen var dog kun baseret på test af 8 stoffer, og skyldtes sammenfald af data. Store interlaboratorielle variationer (CV's >50%) blev også observeret med forskellige tests med hønseæg. Moderate korrelationer til Draize MAS værdier blev opnået med en CAM tryphanblå absorptionstest ( $r=0.69$ , n=52) og HET-CAM testen ( $r=0.72$ , n=55) (Hagino et al., 1999). Det vil dog formentlig være muligt at øge HET-CAM testens forudsigelsesevne ved anvendelse af en non-lineær logistisk forudsigelsesmodel. Resultater fra forskellige cytotoxicitets-tests var moderat til godt korrelerede til Draize MAS værdier: Normale hornhindeceller fra kaniner ( $r=0.53$ , n=28), hæmolysetest med røde blodlegemer ( $r=0.63$ , n=17), mammale cellelinier ( $r>0.71$ , n=29), og SIRC hornhindeceller fra kaniner ( $r>0.81$ , n=29-30). For disse tests var den interlaboratorielle reproducerbarhed acceptabel med CV'er fra 24% til 37% (Ohno et al., 1999).

#### *BCOP workshop, 1997*

BCOP testens evne til at forudsige øjenirritation er blevet evalueret af en gruppe af forskere fra laboratorier med en stor erfaring i brug af testen. En database med *in vitro* resultater for mere end 200 teststoffer har vist overensstemmelser til Draize test irritationsklasser på 80-85%, og assayet har en god reproducerbarhed (Sina and Gautheron, 1998). Test af et stort antal positive kontrollet har vist en udmærket intralaboratorie reproducerbarhed med CV'er på totale BCOP scoringer fra 12% til 16% (Harbell and Curren, 1998). Fluorescein lækage testen med MDCK celler har demonstreret en meget ringe evne til at skelne mellem Draize test irritationsklasser, og målinger af fluorescein lækage gennem hornhinder er ikke vist at være prædiktivt for øjenirritation (Sina and Gautheron, 1998).

#### *COLIPA studiet*

Resultaterne af COLIPA's valideringsundersøgelse af alternativer til øjenirritationstests bekræftede generelt resultaterne fra de tidligere undersøgelser. Flertallet af de alternative metoder i COLIPA studiet fandtes ikke egnede til forudsigelse af akut øjenirritation for blandede stoffer og produkter (Brantom et al., 1997). Dette stemmer overens med konklusionen på EU/Home Office studiet, hvor ingen af de anvendte metoder viste sig egnede

til erstatning af Draize testen (Balls et al., 1995). Flere *in vitro* metoder fandtes i COLIPA studiet egnede til at forudsige øjenirritation forårsaget af kosmetiske ingredienser, specielt af tensidbaserede produkter. I EU/Home Office studiet fandtes flere alternative tests også gode til forudsigelse af tensiders øjenirriterende effekter, og dette er tillige observeret i en række tidligere valideringsundersøgelser (Rasmussen, 1993). I COLIPA studiet havde tests med fosterhinder fra hønseæg en meget ringe reproducerbarhed. Resultater opnået med de øvrige metoder var relativt godt reproducerbare mellem forskellige laboratorier.

COLIPA studiet havde det væsentlige nye resultat, at en *in vitro* model, SKIN<sup>2</sup> ZK1200, blev vist at være meget god til forudsigelse af et bredt spektrum af øjenirritationsdata. Metodens reproducerbarhed må betragtes som rimelig, og den er i lighed med de fleste andre *in vitro* metoder betydeligt mere reproducerbar end Draize testen (Southey et al., 1999). COLIPA studiet viste dermed som den første større blindundersøgelse af blandede stoffer og produkter, at en alternativ test vil kunne erstatte dyreforsøg for akut øjenirritation. SKIN<sup>2</sup> ZK1200 metoden har tillige som den første *in vitro* test vist et godt potentiale for brug til vurdering af vedvarende øjenirritation. Testen har derfor vist et meget stort potentiale for at kunne fungere som en fuldstændig erstatning for dyreforsøg for øjenirritation. På grund af standsningen af produktion og salg af SKIN<sup>2</sup> modellerne er udvikling og anvendelse af alternative vævsmodeller særdeles påkrævet, hvis de eksisterende dyreforsøg for lokalirritation skal kunne erstattes med mere pålidelige alternative metoder.

#### *Alternative vævsmodeller*

Der er udviklet flere kommercielt tilgængelige øjenmodeller med væv af keratinocytter, der vokser på microporøse membraner. Modellerne omfatter en EpiOcular<sup>TM</sup> model og en REC model, der begge har vist potentiale for at kunne erstatte øjenirritationsforsøg. I begge modeller bliver teststofferne påført ufortyndede på overfladen af vævene, og toksiciteten bedømmes på basis af tids-respons relationer opnået med MTT assayet. Ved anvendelse af EpiOcular<sup>TM</sup> modellen var t<sub>50</sub> værdier for 28 kemikalier fra ECETOC's databank meget godt korrelerede ( $r=0.90$ ) til Draize test scoringsværdier. Ved brug af en forudsigelsesmodel udviklet på baggrund af denne undersøgelse, fandtes der en god korrelation ( $r=0.87$ ) mellem forudsagte og observerede Draize scoringsværdier for 41 færdige produkter (Sheasgreen et al., 1996). Tillige er der fundet en relativt god overensstemmelse mellem t<sub>50</sub> værdier for 43 prøver omfattende væsker, pulvere og geler opnået med EpiOcular<sup>TM</sup> modellen og Draize test klassifikationer (Stern et al., 1998). Der er senere udviklet en 2 parameter logistisk forudsigelsesmodel for MTT data opnået med EpiOcular<sup>TM</sup> væv på basis af *in vitro/in vivo* data for 19 vandopløselige kemikalier og 41 færdige produkter, og en god korrelation ( $r=0.90$ ) blev opnået mellem forudsagte og observerede Draize MMAS værdier. Ved brug af forudsigelsesmodellen fandtes der tillige en god forudsigelse af Draize MMAS værdier ( $r=0.89$ ) for 11 færdige produkter. Reproducerbarheden af MTT testen med EpiOcular<sup>TM</sup> modellen er god. På basis af test af 132 prøver fandtes gennemsnitlige CV'er på ca. 5% med negative kontroller, og på 25% med en positiv kontrol (0.3% Triton X-100) (Klausner et al., 1999).

REC modellen ligner meget EpiOcular<sup>TM</sup> modellen, og en god korrelation ( $r=0.89$ ) er opnået mellem MTT test data fra REC modellen og Draize test

MMAS værdier for 40 kosmetiske formuleringer, der dækkede et bredt spektrum af irritationskalaen. Tillige er der fundet acceptable CV'er i et reproducerbarhedsstudie med 1% SLS (n=12, CV=18%) og et tensidbaseret produkt (n=15, CV=24%) (Doucet et al., 1998).

## 6 Evaluering af alternativer til øjenirritationstesten

På baggrund af de opnåede resultater af EU/Home Office studiet er det blevet vurderet, at alternative metoder til erstatning af Draize testen for øjenirritation først vil kunne etableres efter år 2005 (Purchase, 1997). Endvidere blev resultaterne af både EU/Home Office studiet og COLIPA studiet karakteriseret som skuffende i EU Kommissionens rapporter om alternativer til erstatning af dyreforsøg til testning af kosmetik (EU Kommissionen, 1995 og 1996b). På dette grundlag er forbudet mod dyreforsøg til kosmetiske produkter i EU's kosmetikdirektiv blevet udskudt til januar, 2000.

Der er dog fornyligt opnået enighed om mere optimistiske synspunkter vedrørende alternativer til øjenirritationsforsøg på workshops arrangeret af COLIPA (Bruner et al., 1998) og EU kommissionens center for validering af alternative metoder (ECVAM)(Balls et al., 1999). På begge workshops blev det understreget, at det er væsentligt at forstå biologiske mekanismer bag øjenirritation, når *in vitro* forudsigelsen af *in vivo* reponset skal forbedres og følgende forskningsområder er foreslået:

- bedre karakterisering af skader på øjets væv, omfattende udvikling af tidlige markører for øjenskade,
- øget forståelse for effekter af kemiske stoffer på tårefilmen, og af konsekvenser af beskadigelse af tårefilmen,
- karakterisering af akut og medium/kronisk øjenbetændelse,
- validering af areal og dybde af hornhindeskader som markører for øjenskade,
- udvikling af metoder til bedømmelse af sårheling, smerter og kinetik i øjets respons, og
- udvikling af metoder til bedømmelse af persistens og reversibilitet af øjenseffekter.

På COLIPA workshoppen var der enighed om, at Draize testens evne til akkurat påvisning af øjenskade for mennesker er begrænset, og at testen har begrænset værdi som basis for udvikling af alternativer. Behovet for anvendelse af data om det humane øjenirritations respons blev derfor understreget. Det blev anbefalet at identificere et reference sæt på 10-20 modelstoffer, om muligt udvalgt på basis af humane øjenirritationsdata, til brug ved udvikling af mekanismebaserede *in vitro* tests. Det blev tillige anbefalet, at revurdere de tilgængelige *in vitro* tests på basis af bedre *in vivo* data, og at udvikle og forfine mekanismebaserede *in vitro* metoder til øjenirritationstesting. De mest relevante tilgængelige *in vitro* metoder blev identificeret som:

- cytotoxicitets assays (inklusive flerlagede vævsmodeller),
- assays for ændringer i epithelfunktioner (fluorescein lakage testen),
- organotypiske tests, som BCOP testen og tests med isolerede øjne fra kyllinger eller kaniner,
- CAM tests som modeller for primære skader på øjets bindehinder, og
- langtidsmodeller for hornhinden (f. eks. hornhinder i kultur).



En ny strategi for validering af *in vitro* tests for øjenirritation baseret på brug af reference standarder (positive kontrolstoffer) blev foreslået på ECVAM workshoppen. ECVAM planlægger videre validering af en test med isolerede kyllingeøjne, BCOP testen, kombineret anvendelse af HET-CAM testen og neutralrødoptagelsestesten, samt EpiOcular™ vævsmodellen. Tillige blev det anbefalet at sammenligne resultaterne fra EU/Home Office og COLIPA studierne med øjenirritationsklassifikationer, og at evaluere potentialet af teststrategier (kombinationer af *in vitro* tests) til at forudsige øjenirritation.

#### *Muligheder for fremskridt med hensyn til erstatning af øjenirritationstests*

Idag er der opnået en meget stor mængde information vedrørende forskellige alternativer til øjenirritationstestning. Forskellige vævsmodeller med direkte påføring af teststoffer har vist sig at være meget gode til at forudsige et bredt spektrum af Draize test MMAS værdier, og modellerne har indtil videre kunnet bruges til test af alle typer af ingredienser og færdige produkter. Foreløbige resultater peger endvidere på at vævsmodeller vil kunne bruges til forudsigelse af øjenskaders reversibilitet. Anvendeligheden af kulturer af celler, der dyrkes i monolag og doseres med teststoffer opløst i kulturmediet, er begrænset til test af vandopløselige stoffer. Med denne type tests er der generelt fundet gode overensstemmelser til øjenirritationsdata for tensider. Resultater fra forskellige metoder med isolerede øjne eller hornhinder har i flere undersøgelser vist sig at give gode overensstemmelser til øjenirritationsdata eller -klassificering. HET-CAM testens anvendelse er begrænset til identifikation af visse stærkt irriterende stoffer, men testen er i flere tilfælde vist at være mindre reproducerbar end dyreforsøgene. EYTEX testen er tillige vist at være uegnet til generel forudsigelse af øjenirritation i flere undersøgelser.

De fleste alternative metoder til øjenirritationstestning er dog vist at være betydeligt mere reproducerbare end *in vivo* øjenirritationstests. På denne baggrund bør der lægges betydeligt mere vægt på vurdering af de alternative metoders evne til at forudsige *in vivo* responset end på evalueringer af metodernes reproducerbarhed. Test af en mindre gruppe af stoffer og produkter udgør en tilstrækkelig basis for vurdering af metodernes reproducerbarhed i de enkelte laboratorier og mellem laboratorier. Rigide krav til valideringen af *in vitro* metoder kan både medføre, at egnede alternativer bliver overset og at uegnede tests bliver accepterede.

På øjenirritationsområdet har evalueringen af de opnåede informationer om alternative tests været betydeligt mere kompliceret end forventet. En ny, uafhængig revurdering af de eksisterende data på linie med IRAG evalueringen i 1997 vil formentlig være meget værdifuld. Det er afgørende, at der etableres gode forudsigelsesmodeller før valideringsstudier påbegyndes, hvis der skal opnås anvendelige forudsigelser af *in vivo* responset. For de fleste alternative til bedømmelse af øjenirritation vil der kunne etableres bedre forudsigelsesmodeller ved brug af avancerede regressionsanalyser af relationerne mellem historiske *in vitro* og *in vivo* data.

Forbuddet mod dyreforsøg til kosmetiktestning i EU's kosmetik direktiv er blevet betragtet som en drivkraft for udvikling og validering af alternative tests. Denne hensigtserklæring udgør dog også en væsentlig hindring for accepten af alternativer: Et EU forbud mod produkter, der har været testet i

bestemte dyreforsøg kan udløse betydelige hindringer for handelen med kosmetik mellem EU's medlemsstater og andre lande. Det specifikke forbud mod dyreforsøg med kosmetiske produkter og deres indholdsstoffer forekommer endvidere at være overflødig, da egnede alternative metoder skal bruges i stedet for dyreforsøg i henhold til EU's direktiv om dyreforsøg. Et effektivt skridt på vejen mod forbedrede muligheder for at opnå en generel accept af alternativer til dyreforsøg til vurdering af sikkerheden af kosmetiske produkter vil formentlig være at fjerne hensigtserklæringen om forbud mod dyreforsøg ved en kommende revision af EU's kosmetikdirektiv.

## 7 Forkortelser

BCOP: Test med isolerede hornhinder fra køjne

BGA: Det tyske sundhedsministerium

CAM: Fosterhinden i hønseæg

CAMVA: Test med fosterhinden i hønseæg

COLIPA: Gruppe af europæiske kosmetikproducenter

CORROSITEX: Fysisk-kemisk test for åtsning

CTFA: Gruppe af kosmetikproducenter i USA

*In vitro*: Reagensglasforsøg

*In vivo*: I den hele organisme

CV: Variationskoefficient

Draize test: Kaninforsøg for øjenirritation

ECETOC: Europæisk center for toksikologi

ECVAM: EU Kommissionens center for validering af alternative tests

EpiOcular<sup>TM</sup>: Vævsmode for øjets hornhinde

EYTEX: Test med planteprotein

GLP: God laboratoriepraksis

HET-CAM: Test med fosterhinden i hønseæg

IRAG: International gruppe af regulerende myndigheder

JCIA: Gruppe af japanske kosmetikproducenter.

JMHW: Japansk ministerium for sundhed og velfærd.

MAS: Maksimal gennemsnitlig scoringsværdi i Draize testen

MATREX<sup>TM</sup>: Vævsmode med bindevævsceller.

MMAS: Modifieret maksimal gennemsnitlig scoringsværdi i Draize testen (skala 0-110).

MTT: Test for cellers viabilitet

NaOH: Natriumhydroxid

Predisafe: Test med frigørelse af farvestoffet neutralrødt

r.: Lineær korrelationskoefficient

SD: Standardafvigelse

SKIN<sup>2</sup> ZK1200: Vævsmodel for øjets hornhinde

SLS: Natriumlaurylsulfat



## 8 Litteraturliste

- Balls, M., Botham, P. A., Bruner, L. H. og Spielmann, H.: The EC/HO international validation study on alternatives to the Draize eye irritation test. *Toxicology in Vitro*, 1995, 9, 871-929.
- Balls, M., Berg, N., Bruner, L. H. et al.: Eye irritation testing: the way forward. *Alternatives to Laboratory Animals*, 1999, 27, 53-77.
- Botham, P., Osborne, R., Atkinson, K., Carr, G., Cottin, M. og van Buskirk, R. G.: Cell function based assays. IRAG Working Group 3. *Food and Chemical Toxicology*, 1997, 35, 67-77.
- Brantom, P. G., Bruner, L. H., Chamberlain, M. et al.: A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. *Toxicology in Vitro*, 1997, 11, 141-179.
- Bruner, L. H., Carr, G. J., Chamberlain, M. og Curren, R. D.: Validation of alternative methods for toxicity testing. *Toxicology in Vitro*, 1996, 10, 479-501.
- Bruner, L. H., De Silva, O., Earl, L. K., Easty, D. L., Pape, W. J. W. og Spielmann, H.: Report on the COLIPA workshop on mechanisms of eye irritation. *Alternatives to Laboratory Animals*, 1998, 26, 811-820.
- Chamberlain, M., Gad, S. C., Gautheron, P. og Prinsen, M. K.: Organotypic models for the assessment/prediction of ocular irritation. IRAG Working Group 1. *Food and Chemical Toxicology*, 1997, 35, 23-37.
- Curren, R. D., Sina, J. F., Feder, P., Kruszewski, F. H., Osborne, R. og Régnier, J.-F.: Other assays. IRAG Working Group 5. *Food and Chemical Toxicology*, 1997, 35, 127-158.
- deSilva, O.: Alternatives to eye irritation evaluation - the industry perspective. From: Lisansky, S. G., Macmillan, R. og Dupuis, J. (eds.): Alternatives to animal testing. Proceedings of an international scientific conference organised by the European Cosmetics Industry, Brussels, Belgium 1995. CPL Press, Newbury, UK, 1996, pp. 67-73.
- Doucet, O., Lanvin, M. og Zastrow, L.: A new in vitro human epithelial model for assessing the eye irritating potential of formulated cosmetic products. *In Vitro & Molecular Toxicology*, 1998, 11, 273-283.
- Earl, L. K., Dickens, A. D. og Rowson, M. J.: A critical analysis of the rabbit eye irritation test variability and its impact on the validation of alternative methods. *Toxicology in Vitro*, 1997, 11, 295-304.
- ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals) (1998). Eye Irritation: Reference chemicals data bank (second edition). Technical Report No. 48 (2). Brussels.

Espersen, R. J., Olsen, P., Nicolaisen, G., Jensen, B. L. og Rasmussen, E. S.: Assessment of recovery from ocular irritancy using a human tissue equivalent model. *Toxicology in Vitro*, 1997, 11, 81-88.

EU (1993): Directive 93/35/EEC. Official Journal of the European Communities, L 151/32, 23.6.1993.

EU Commission (1995): Development, validation and legal acceptance of alternative methods in the field of cosmetic products.

EU Commission (1996 a) Classification, packaging and labelling of dangerous substances in the European Union. Updated version of Directive 67/548/EEC.

EU Commission (1996 b): Development, validation and legal acceptance of alternative methods in the field of cosmetic products.

Gettings, S. D., Lordo, R. A., Demetrulias, J., Feder, P. I., og Hintze, K. L.: Comparison of low-volume, Draize and in vitro eye irritation test data. I. Hydroalcoholic formulations. *Food and Chemical Toxicology*, 1996a 34, 737-749.

Gettings, S. D., Lordo, R. A., Hintze, K. L. et al.: 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*, (1996b), 34, 79-117.

Gettings, S. D., Lordo, R. A., Feder, P. I., og Hintze, K. L.: Comparison of low-volume, Draize and in vitro eye irritation test data. II. Oil/water emulsions. *Food and Chemical Toxicology*, (1998), 36, 47-59.

Gupta, K. C., Chambers, W. A., Green, S. et al.: An eye irritation test protocol and an evaluation and classification scheme. *Food and Chemical Toxicology*, 1993, 31, 117-121.

Harbell, J. W. og Curren, R.: Report from the bovine corneal opacity and permeability technical workshop, November 3-4, 1997, Gaithersburg, MD. Assay performance. *In Vitro & Molecular Toxicology*, 1998, 11, 337-345.

Harbell, J. W., Koontz, S. W., Lewis, R. W., Lovell, D. og Acosta, D.: Cell cytotoxicity assays. IRAG Working Group 4. *Food and Chemical Toxicology*, 1997, 35, 79-126.

Hatao, M., Murakami, N. Sakamoto, K. et al.: Interlaboratory validation of in vitro eye irritation tests for cosmetic ingredients (4) Haemoglobin denaturation test. *Toxicology in Vitro*, 1999, 13, 125-137.

Hagino, S., Kinoshita, S., Tani, N. et al.: Interlaboratory validation of in vitro eye irritation tests for cosmetic ingredients (2) Chorioallantoic membrane tests. *Toxicology in Vitro*, 1999, 13, 99-113.

Klausner, M., Sennott, H. A., Breyfogle, B., Makwana, A., og Kubilis, J.: The EpiOcular prediction model: a reproducible in vitro means of assessing ocular irritancy potential. *The Toxicologist*, 1999, 48, 336.

Kurishita, A. Katoh, T., Ohsawa, H. et al.: Interlaboratory validation of in vitro eye irritation tests for cosmetic ingredients (5) SKIN<sup>2</sup>™ZK1100 and tissue equivalent assay. *Toxicology in Vitro*, 1999, 13, 139-151.

Matsukawa, K., Masuda, K., Kakishima, H. et al.: Interlaboratory validation of in vitro eye irritation tests for cosmetic ingredients (11) EYTEX™. *Toxicology in Vitro*, 1999, 13, 209-217.

Ohno, Y., Kaneko, T., Inoue, T. et al.: Interlaboratory validation of in vitro eye irritation tests for cosmetic ingredients (1) Overview of the validation study and Draize scores for the evaluation of the tests. *Toxicology in Vitro*, 1999, 13, 73-98.

Ohuchi, J., Kasai, Y., Sakamoto, K.: Interlaboratory validation of in vitro eye irritation tests for cosmetic ingredients (6) Evaluation of MATREX™. *Toxicology in Vitro*, 1999, 13, 153-162.

Osborne, R., Perkins, M. A. og Roberts, D. A.: Development and intralaboratory evaluation of an in vitro human cell-based test to aid ocular irritancy assessments. *Fundamental and Applied Toxicology*, 1995, 28, 139-153.

Purchase, I. F. H.: Prospects for reduction and replacement alternatives in regulatory toxicology. *Toxicology in Vitro*, 1997, 313-319.

Rasmussen, E. S: Prospects for use of in vitro methods for assessment of human safety. National Food Agency of Denmark, 1995.

Rasmussen, E. S.: Use of reconstructed human tissue models in toxicological and pharmacological studies. *Dansk Veterinærtidsskrift*, 1/11, 1996, 943-948.

Sheasgreen, J. E., Kubilus, J., Sennott, H., Ogle, P. og Klausner, M.: Reproducibility and correlation of EpiOcular™, a three-dimensional tissue culture model of human corneal epithelium. *Alternatives to Laboratory Animals*, 1996, 24, Special Issue, 284.

Sina, J. F. og Gautheron, P.: Report from the bovine corneal opacity and permeability technical workshop, November 3-4, 1997, Gaithersburg, MD. An historical perspective. *In Vitro & Molecular Toxicology*, 1998, 11, 316-326.

Southee, J. A., McPherson, J. P., Osborne, R., Carr, G. J. og Rasmussen, E.: The performance of the tissue equivalent assay using the SKIN<sup>2</sup>™ ZK1200 model in the COLIPA international validation study on alternatives to the Draize eye irritation test. *Toxicology in Vitro*, 1999, 13, 355-373.

Spielmann, H., Kalweit, S., Liebsch, M. et al.: Validation study of alternatives to the Draize eye irritation test in Germany: cytotoxicity testing and HET-CAM test with 136 industrial chemicals. *Toxicology in Vitro*, 1993, 7, 505-510.

Spielmann, H., Liebsch, M., Kalweit, S. et al.: Results of a validation study in Germany of two in vitro alternatives to the Draize eye irritation test, the HET-CAM test and the 3T3 NRU cytotoxicity test. *Alternatives to Laboratory Animals*, 1996, 24, 741-858.



Spielmann, H., Liebsch, M., Moldenhauer, F. et al.: CAM based assays. IRAG Working Group 2. *Food and Chemical Toxicology*, 1997, 35, 39-66.

Stern, M., Klausner, M., Alvarado, R., Renskers, K., og Dickens, M.: Evaluation of the EpiOcular™ tissue model as an alternative to the Draize eye irritation test. *Toxicology in Vitro*, 1998, 12, 455-461.

van de Sandt, J. J. M. og Rutten, A. A. J. J. L.: Differential effects of chemical irritants in rabbit and human skin organ cultures. *Toxicology in Vitro*, 1995, 9, 157-168.

Weil, C. S. og Scala, R. A.: Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. *Toxicology and Applied Pharmacology*, 1971, 19, 276-360.



Filnavn: Alternativer til dyreforsøg for øjenirritation.doc  
Bibliotek: G:  
Skabelon: C:\Programmer\Microsoft Office\Skabeloner\Normal.dot  
Titel: Working Report  
Emne:  
Forfatter: Pipse  
Nøgleord:  
Kommentarer:  
Oprettelsesdato: 31-05-00 11:51  
Versionsnummer: 2  
Senest gemt: 31-05-00 11:51  
Senest gemt af: Lisbeth Ramsgaard Carlsen  
Redigeringstid: 1 minut  
Senest udskrevet: 13-06-00 09:45  
Ved seneste fulde udskrift  
Sider: 2  
Ord: 112 (ca.)  
Tegn: 643 (ca.)