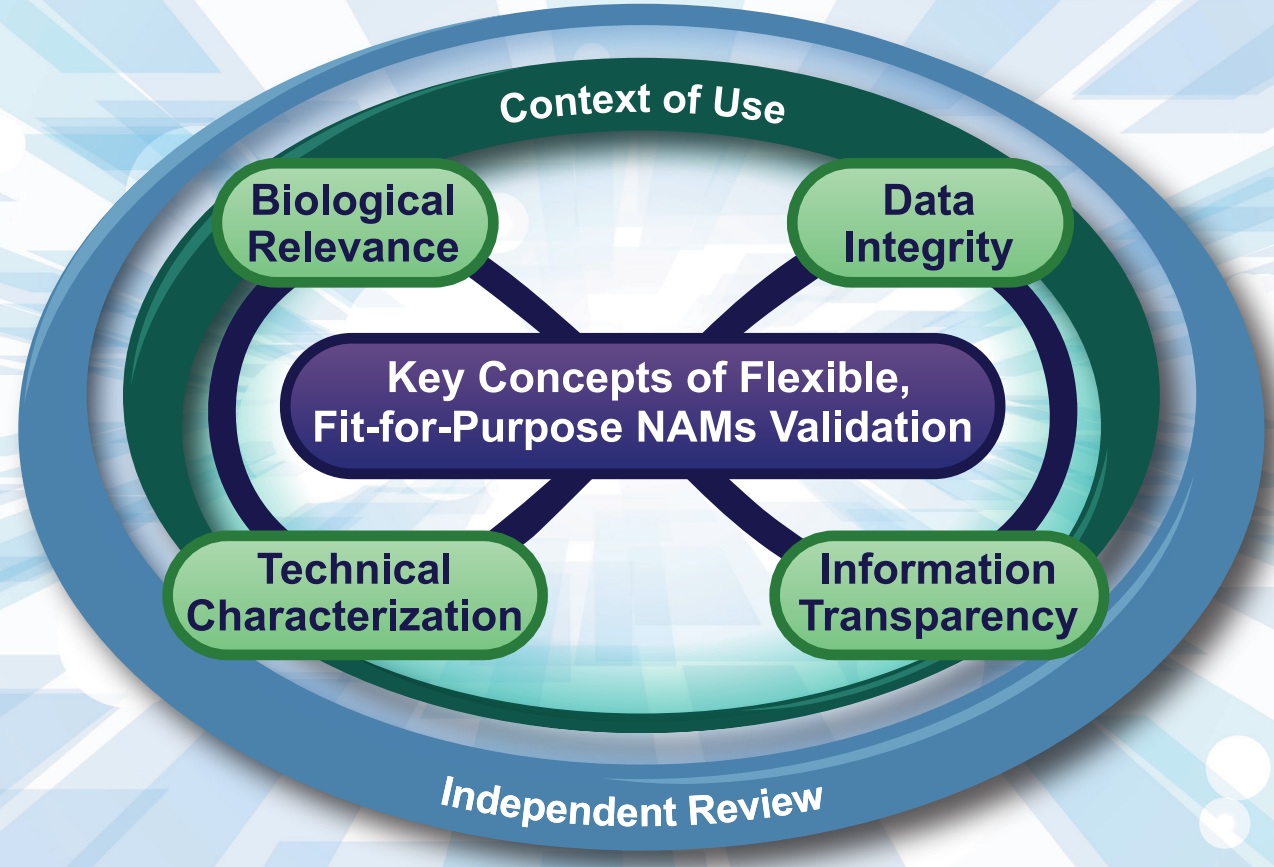
UNITED STATES ICCVAM Advancing Alternatives to Animal Testing

Interagency Coordinating Committee on the Validation of Alternative Methods

Validation, qualification et acceptation réglementaire des nouvelles méthodes d'approche (NAM)

Mars 2024



doi:10.22427/NICEATM-2

Table des matières

Liste des tableaux v

Liste des figures vi

Abréviations et acronymes vii

Groupe de travail sur la validation de l'ICCVAM ix

Représentants des agences auprès de l'ICCVAM xi

Avis de non-responsabilité 1

Résumé analytique 1

1.0 Présentation 2

2.0 Principaux concepts d'une validation flexible et adaptée à l'objectif des NAM 5

3.0 Appliquer les concepts clés pour renforcer la confiance dans les NAM 7

3.1 Contexte d'utilisation 7

3.2 Pertinence biologique 9

3.2.1 Compréhension mécanistique 9

3.2.2 Composés de référence 13

3.2.3 Comparaison avec les méthodes existantes sur les animaux de laboratoire 16

3.3 Caractérisation technique 19

3.3.1 Intégration d'outils de qualité sélectionnés 19

3.3.2 Meilleures pratiques en matière de contrôle qualité 22

3.3.3 Documentation 28

3.4 Intégrité des données 31

3.5 Transparence de l'information 31

3.6 Examen indépendant 32

4.0 Acceptation des NAM par des agences fédérales des États-Unis 33

4.1 Comprendre les besoins réglementaires et les contextes décisionnels 33

4.2 Considérations relatives au contexte d'utilisation 34

4.3 Évolution de la confiance en fonction de l'expérience acquise 34

5.0 Harmonisation aux États-Unis et à l'échelle internationale 35

5.1 Harmonisation aux États-Unis : rôle de l'ICCVAM et du NICEATM 35

5.2 Harmonisation aux États-Unis : autres collaborations fédérales promouvant les principes des 3R 37

5.3 Harmonisation internationale 37

6.0 Communication et formation visant à encourager l'utilisation des NAM 39

7.0 Conclusion et mise en œuvre 40

Références 41

ANNEXE A : GLOSSAIRE 52

ANNEXE B : OUTILS DE QUALITÉ 58

1.0 Organigrammes 58

2.0 Analyse des causes et des effets 59

3.0 Cartes de contrôle 60

4.0 Feuilles de relevé 62

5.0 Diagrammes de dispersion 62

ANNEXE C : ÉVALUATION DE LA MÉTHODE ANALYTIQUE 64

1.0 Limites de détection et de quantification 64

2.0 Identification des interférences 64

3.0 Évaluation de la précision analytique 64

4.0 Stabilité des matériaux 65

5.0 Test de robustesse 65

6.0 Analyse du taux de recouvrement 66

7.0 Analyse technique du domaine d'application 67

8.0 Témoin positif 67

9.0 Nomes de référence pour l'étalonnage des instruments 67

10.0 Établissement de spécifications 68

Liste des tableaux

Tableau 1. Exemples de documents américains et internationaux existants relatifs à la validation, à la qualification et à l'utilisation réglementaire des NAM 3

Tableau 2. Manuscrits produits par les groupes de travail de l'ICCVAM détaillant les besoins des agences en matière d'essais dans des domaines sélectionnés..... 8

Tableau 3. Exemples de critères d'évaluation pour lesquels la pertinence biologique et mécanistique des NAM a été démontrée 10

Tableau 4. Utilisations sélectionnées de composés de référence..... 14

Tableau 5. Outils de qualité souvent utilisés pour la caractérisation technique des NAM..... 21

Tableau 6. Évaluation des méthodes analytiques 24

Liste des figures

Figure 1. Concepts clés lors de l'élaboration et de la mise en œuvre de stratégies de validation flexibles et adaptées à l'objectif des NAM. 6

Figure 2. Cadre pour le développement de NAM robustes..... 20

Figure supplémentaire 1. Organigramme décrivant le protocole modifié relatif au MTS..... 58

Figure supplémentaire 2. Diagramme des causes et des effets 59

Figure supplémentaire 3. Données des cartes de contrôle 61

Figure supplémentaire 4. Corrélation entre les valeurs de la CE_{50} du sulfate de cadmium..... 63

Abréviations et acronymes

3R : principes de remplacement, de réduction ou de réajustement de l'utilisation des animaux

AD : approche définie

AOP : adverse outcome pathway (voie conduisant à des effets indésirables)

BPL : Bonnes pratiques de laboratoire

C&E : causes et effets

CIH : Conseil international d'harmonisation des exigences techniques pour l'enregistrement des médicaments à usage humain

CPSC : Consumer Product Safety Commission (Commission américaine pour la sécurité des produits de consommation)

CU : contexte d'utilisation

DO : densité optique

EPA : Environmental Protection Agency (Agence américaine de protection de l'environnement)

EURL ECVAM : European Union Reference Laboratory for the Validation of Alternative Methods (Laboratoire de référence de l'Union européenne pour la promotion des méthodes de substitution à l'expérimentation animale)

FDA : Food and Drug Administration (Secrétariat américain aux produits alimentaires et pharmaceutiques)

GD : guidance document (document d'orientation)

GIVIMP : Guidance Document on Good In Vitro Method Practices (Document d'orientation sur les bonnes pratiques relatives aux méthodes in vitro)

IATA : Integrated Approaches to Testing and Assessment (Approches intégrées en matière d'essai et d'évaluation)

ICATM : International Cooperation on Alternative Test Methods (Coopération internationale relative aux méthodes de substitution à l'expérimentation animale)

ICCVAM : Interagency Coordinating Committee on the Validation of Alternative Methods (Comité de coordination interagences pour la validation des méthodes alternatives)

IQ/OQ/PQ : Installation Quality/Operation Quality/Performance Quality (Qualité d'installation/Qualité opérationnelle/Qualité de performance)

JaCVAM: Japanese Center for the Validation of Alternative Methods (Centre japonais pour la validation des méthodes alternatives)

MTS : (3-(4,5-diméthylthiazol-2-yl)-5-(3-carboxyméthoxyphényl)-2-(4-sulfophényl)-2H-tétrazolium)

NAM : new approach methodologies (nouvelles méthodes d'approche)

NCATS : National Center for Advancing Translational Sciences (Centre national pour l'avancement des sciences translationnelles)

NICEATM : National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (Centre interagences du programme national de toxicologie pour l'évaluation des méthodes toxicologiques alternatives)

NIEHS : National Institute of Environmental Health Sciences (Institut national des sciences en santé environnementale)

OCDE : Organisation de coopération et de développement économiques

SGH : Système général harmonisé de classification et d'étiquetage des produits chimiques

SOP : standard operating procedure (procédure opérationnelle standard)

TG : test guideline (ligne directrice pour les essais)

Tox21 : Toxicology in the 21st Century (La Toxicologie au 21e siècle)

TSAR : Tracking System for Alternative methods towards Regulatory acceptance (Système de suivi de méthodes d'essais de substitution jusqu'à leur acceptation dans la réglementation de l'Union européenne)

Groupe de travail sur la validation de l'ICCVAM

Agency for Toxic Substances and Disease Registry (Agence du registre des substances toxiques et des maladies)
Moiz Mumtaz, PhD

Consumer Product Safety Commission (Commision pour la sécurité des produits de consommation)
John Gordon, PhD (coprésident)

Department of Defense (Département de la Défense)
Donald Cronce, PhD
Natalia Garcia-Reyero Vinas, PhD
Matthew Johnson, DVM, DACLAM (jusqu'en juin 2022)
Emily N. Reinke, PhD (jusqu'en avril 2022)

Department of Veterans Affairs (Département des Anciens Combattants)
George Lathrop, Jr., DVM, MS, DACLAM

Environmental Protection Agency (Agence de protection de l'environnement)
Office of Pesticide Programs (Bureau des programmes sur les pesticides)
Anna Lowit, PhD
Scott Lynn, PhD
Monique Perron, ScD
Office of Research and Development (Bureau de la recherche et du développement)
Kelly Carstens, PhD
Alison Harrill, PhD
Nisha Sipes, PhD

Food and Drug Administration (Secrétariat américain aux produits alimentaires et pharmaceutiques)
Center for Devices and Radiological Health (Centre pour les dispositifs et la santé radiologique)
Jennifer Goode, BS
Center for Drug Evaluation and Research (Centre d'évaluation et de recherche sur les médicaments)
Paul C. Brown, PhD

Center for Food Safety and Applied Nutrition (Centre pour la sécurité alimentaire et la nutrition appliquée)
Suzanne Fitzpatrick, PhD (coprésidente)
Anneliese Striz, PhD
Center for Tobacco Products (Centre pour les produits du tabac)
Jueichuan (Connie) Kang, PhD
Office of the Chief Scientist (Bureau de l'expert scientifique en chef)
Tracy Chen, PhD

National Institute of Environmental Health Sciences (Institut national des sciences en santé environnementale)
Warren Casey, PhD
National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM ; Centre interagences du programme national de toxicologie pour l'évaluation des méthodes toxicologiques alternatives)
Helena Hogberg, PhD
Nicole Kleinstreuer, PhD

National Institutes of Health (Instituts nationaux de la santé)
National Center for Advancing Translational Sciences (Centre national pour l'avancement des sciences translationnelles)
Menghang Xia, PhD

National Institute of Standards and Technology (Institut national des normes et de la technologie)
Elijah Petersen, PhD (coprésident)

Occupational Safety and Health Administration (Administration de la sécurité et de la santé au travail)
Janet Carter, MS

Personnel contractuel d'appui du NICEATM (Inotiv)
David Allen, PhD (jusqu'en décembre 2023)
Michaëla Blaylock
Amber Daniel, MTox
Agnes Karmaus, PhD (jusqu'en avril 2023)

Représentants des agences auprès de l'ICCVAM

Composition du comité de l'ICCVAM au moment de la publication de ce rapport
* Représentant principal de l'agence + représentant principal suppléant de l'agence

Agency for Toxic Substances and Disease Registry (Agence du registre des substances toxiques et des maladies)

* Moiz Mumtaz, PhD
Patricia Ruiz, PhD

U.S. Consumer Product Safety Commission (Commission américaine pour la sécurité des produits de consommation)

* John Gordon, PhD
\+ Kristina Hatlelid, PhD, MPH
\+ Eric Hooker, MS
\+ Joanna Matheson, PhD

U.S. Department of Agriculture (Département de l'Agriculture des États-Unis)

* Jessie Carder
\+ Patrice Klein, MS, VMD, DACPV, DACVPM
Erika Edwards
Ben Green, PhD
Katherine Horak, PhD

U.S. Department of Defense (Département de la Défense des États-Unis)

* Shannon Marko, DVM, DACLAM
\+ Natalia Garcia-Reyero Vinas, PhD (coprésidente)
\+ Saber Hussain, PhD, boursier ATS, boursier AFRL
\+ Elaine Merrill, PhD
Matthew Grogg, PhD
Donald Cronce, PhD

U.S. Department of Energy (Département de l'Énergie des États-Unis)

R. Todd Anderson, PhD

U.S. Department of the Interior (Département de l'Intérieur des États-Unis)

* Barnett A. Rattner, PhD
\+ Jessica K. Leet, PhD

U.S. Department of Transportation (Département des transports des États-Unis)

* Steve Hwang, PhD
\+ Rebecca Rothhaas
\+ Ryan Vierling, PhD

U.S. Department of Veterans Affairs Office of Research and Development (Bureau de la recherche et du développement du département des Anciens Combattants des États-Unis)

* Holly Krull, PhD
\+ George Lathrop, Jr., DVM, MS, DACLAM

U.S. Environmental Protection Agency (Agence de protection de l'environnement des États-Unis)

Office of Chemical Safety and Pollution Prevention (Bureau de la sécurité chimique et de la prévention de la pollution)

Charles Kovatch

Office of Pesticide Programs (Bureau des programmes sur les pesticides)

Monique Perron, PhD
William (Bill) Eckel, PhD
Cecilia Tan, PhD

Office of Pollution Prevention and Toxics (Bureau de la prévention de la pollution et des produits toxiques)

* Anna Lowit, PhD
Louis (Gino) Scarano, PhD

Office of Research and Development (Bureau de la recherche et du développement)

\+ Alison Harrill, PhD
Kelly Carstens, PhD
Grace Patlewicz, PhD

Food and Drug Administration (Secrétariat américain aux produits alimentaires et pharmaceutiques)

Center for Biologics Evaluation and Research (Centre d'évaluation et de recherche sur les produits biologiques)

Leslie Wagner
Allen Wensky, PhD

Center for Devices and Radiological Health (Centre pour les dispositifs et la santé radiologique)

\+ Jennifer Goode, BS
Simona Bancos, PhD
Rakhi M. Dalal-Panguluri, PhD

Center for Drug Evaluation and Research (Centre d'évaluation et de recherche sur les médicaments)

Paul C. Brown, PhD
Nakissa Sadrieh, PhD

Center for Food Safety and Applied Nutrition (Centre pour la sécurité alimentaire et la nutrition appliquée)

* Suzanne Fitzpatrick, PhD (coprésidente)
Omari J. Bandele, PhD
Patrick Crittenden, PhD
Shruti Kabadi, PhD

Center for Tobacco Products (Centre pour les produits du tabac)

Jueichuan (Connie) Kang, PhD

Center for Veterinary Medicine (Centre de médecine vétérinaire)

M. Cecilia Aguila, DVM
Li You, PhD

National Center for Toxicological Research (Centre national de recherche toxicologique)

Mugimane (Manju) Manjanatha, PhD

Office of the Chief Scientist (Bureau de l'expert scientifique en chef)

Tracy Chen, PhD
Chad P. Nelson, PhD, MSPH

National Cancer Institute (Institut national du cancer)

* Brian Cholewa, PhD
\+ Ron Johnson, PhD
Gary Robinson, PhD

National Institute for Occupational Safety and Health (Institut national pour la sécurité et la santé au travail)

* Stephen Leonard, PhD

National Institute for Environmental Health Sciences (Institut national des sciences en santé environnementale)

* Warren Casey, PhD
\+ Stephen Ferguson, PhD
\+ David M. Reif, PhD
Nicole Kleinstreuer, PhD

National Institute of Standards and Technology (Institut national des normes et de la technologie)

* John Elliott, PhD
\+ Elijah Petersen, PhD

National Institutes of Health (Instituts nationaux de la santé)

* Nicolette Petervary, VMD, MS, DACAW
Michael Eichner, DVM, DACLAM

National Library of Medicine (Bibliothèque nationale de médecine)

* Dina N. Paltoo, PhD, MPH, CPI

Occupational Safety and Health Administration (Administration de la sécurité et de la santé au travail)

* Deana Holmes, MT
\+ Janet Carter, MS

Avis de non-responsabilité

Ce rapport a été développé en tant que ressource pour les agences fédérales américaines et les parties prenantes qui cherchent à établir la confiance dans les nouvelles approches destinées à remplacer, réduire ou réajuster le recours aux tests sur les animaux. Les principes décrits dans ce rapport ont été élaborés avec la contribution du personnel de 17 agences fédérales, de plusieurs groupes de travail interagences, du public et du Scientific Advisory Committee on Alternative Toxicological Methods (SACATM ; Comité consultatif scientifique sur les méthodes toxicologiques alternatives). En tant que tel, ce rapport ne reflète pas nécessairement les opinions ou les politiques de l'ensemble des agences ou groupes de travail. Il ne crée aucun droit pour des personnes ou parties et ne doit pas être considéré comme un engagement de la part d'une quelconque agence fédérale. Ce rapport n'établit aucune responsabilité juridiquement exécutoire. Il a pour but de décrire la réflexion actuelle du Comité de coordination interagences pour la validation des méthodes alternatives (ICCVAM) sur un sujet précis et ne doit être considéré que comme une série de recommandations, à moins que des exigences réglementaires ou légales spécifiques soient citées. L'utilisation du mot « devrait » signifie qu'une mesure est suggérée ou recommandée, mais pas obligatoire.

Résumé analytique

De nouvelles méthodes d'approche (NAM) sont développées de plus en plus fréquemment et sont utilisées pour fournir des évaluations réglementaires ou non réglementaires des effets toxiques potentiels de substances chimiques et de produits sur la santé humaine et l'environnement. Ces NAM sont utilisées pour étudier les mécanismes biologiques qui sous-tendent les processus toxicologiques, faciliter l'évaluation de produits nouveaux et existants, et produire des informations sur l'identification des dangers et les relations dose-réponse à des fins d'évaluation des risques et de classification des dangers pour la santé et l'environnement. Ce rapport a été élaboré par l'ICCVAM pour aider les élaborateurs et les utilisateurs finaux à renforcer la confiance dans les NAM. Cette confiance peut être instaurée grâce à la mise en œuvre de stratégies de validation flexibles et adaptées à l'objectif qui tiennent compte de l'application prévue de chaque NAM. Le présent rapport contribue à renforcer la confiance en décrivant différents concepts, notamment le contexte d'utilisation, la pertinence biologique et la caractérisation technique des NAM. Le rapport examine l'acceptation réglementaire des NAM par les agences fédérales américaines et les possibilités d'une harmonisation internationale, et passe en revue les meilleures pratiques en matière d'outils de qualité et d'évaluation technique des NAM. Il souligne également la nécessité d'une communication entre élaborateurs, utilisateurs finaux et organismes de réglementation. La stratégie uniforme de validation est obsolète et n'est pas à même de répondre aux besoins d'un domaine de pointe tel que celui des NAM. L'ICCVAM est une organisation dont la capacité de conseil peut favoriser cette transition vers des approches plus modernes, en coordonnant les efforts de validation et en créant des opportunités de collaboration en matière de recherche. C'est la raison pour laquelle l'ICCVAM a élaboré ce rapport, en coopération avec 17 organismes de réglementation et laboratoires de

recherche fédéraux, pour aider élaborateurs, utilisateurs finaux et organismes de réglementation à renforcer la confiance dans les NAM, afin qu'elles puissent être mises en œuvre pour compléter ou remplacer les tests sur les animaux à des fins réglementaires ou autres.

1.0 Présentation

Le Comité de coordination interagences pour la validation des méthodes alternatives (ICCVAM) est composé de représentants de 17 organismes fédéraux de réglementation et de recherche des États-Unis qui demandent, utilisent, génèrent ou diffusent des informations sur les essais de toxicologie et de sécurité. L'ICCVAM réalise des évaluations techniques de stratégies d'essai intégrées et de méthodes d'essai de sécurité nouvelles, révisées et alternatives applicables dans le domaine réglementaire. L'ICCVAM encourage également la validation scientifique et l'acceptation ou la qualification réglementaire de méthodes d'essai qui évaluent avec exactitude la sécurité chimique et les dangers des produits concernés dans le but de remplacer, de réduire ou de réajuster (en améliorant le bien-être des animaux et en atténuant ou en évitant la douleur et la détresse) l'utilisation des animaux.

Peu après sa création en tant que comité permanent en 1997, l'ICCVAM a publié un rapport intitulé Validation and Regulatory Acceptance of Toxicological Test Methods (« Validation et acceptation réglementaire des méthodes d'essai toxicologiques ») (ICCVAM, 1997), qui décrivait les critères de validation et d'acceptation réglementaire des méthodes d'essai nouvelles et alternatives. Des directives supplémentaires ont ensuite été fournies dans une publication de 2003, ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods (« Lignes directrices de l'ICCVAM pour la nomination et la soumission de méthodes d'essai nouvelles, révisées et alternatives ») (ICCVAM, 2003). Les principes énoncés dans ces documents ont ensuite servi de base à l'élaboration des critères internationaux décrits dans le document d'orientation (GD) 34 de l'Organisation de coopération et de développement économiques (OCDE) intitulé Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (« Document d'orientation sur la validation et l'acceptation internationale des méthodes d'essai nouvelles ou actualisées pour l'évaluation des dangers ») (OCDE, 2005). Ces ressources décrivaient un modèle de validation flexible en principe, mais qui, en pratique, a montré plusieurs limitations, notamment sa longueur et sa forte consommation de ressources. Dans certains contextes d'utilisation, il peut ne pas être nécessaire de soumettre les méthodes à toutes les étapes de ce processus de validation pour permettre à une agence fédérale de recueillir des données exploitables. De plus, ces documents ne sont pas compatibles avec de nombreuses approches modernes des essais de toxicité, qui mettent l'accent moins sur la substitution des tests *in vivo* par une méthode alternative unique que sur l'intégration des résultats de multiples essais *in vitro* et *in chemico* et d'approches *in silico* (p. ex. modèles informatiques). En 2018, l'ICCVAM a publié A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States (« Feuille de route stratégique pour l'établissement de nouvelles approches destinées à évaluer la sécurité des produits chimiques et des médicaments aux États-Unis ») (ICCVAM, 2018). Ce document fournit un cadre conceptuel

favorisant une meilleure communication entre les agences et les élaborateurs de méthodes d'essai ainsi qu'une plus grande flexibilité dans la manière d'établir la confiance, afin de garantir l'adoption de nouvelles méthodes par les agences fédérales et les industries réglementées, une fois validées pour une application ou un contexte d'utilisation (CU) particulier. Le texte qui suit fournit des informations plus précises sur les moyens de renforcer la confiance dans les nouvelles méthodes d'approche (NAM) sur la base des principes énoncés dans la Feuille de route 2018 de l'ICCVAM.

Dans le contexte de ce rapport, le terme « NAM » fait référence à toute technologie, méthodologie, approche ou combinaison de celles-ci qui peut être utilisée pour fournir des informations sur l'évaluation des dangers et des risques chimiques et favorise le remplacement, la réduction ou le réajustement de l'utilisation des animaux (principes des 3R). Ce rapport s'appuie sur les principes énoncés dans les directives et les documents existants (Tableau 1) acceptés aux États-Unis et à l'échelle internationale, afin de préconiser une approche plus flexible de renforcement de la confiance dans les NAM. Nous présentons ici les concepts clés de validation, de qualification et d'acceptation réglementaire appliqués aux NAM comprenant des essais dans un système biologique (p. ex. essais *in vitro*, certains essais *in chemico*, petits organismes modèles), en reconnaissant que des considérations supplémentaires peuvent être nécessaires pour d'autres types de NAM, tels que les prédictions des modèles informatiques (OCDE, 2007).

Tableau 1. Exemples de documents américains et internationaux existants relatifs à la validation, à la qualification et à l'utilisation réglementaire des NAM

Titre du document	**Référence**
Validation and Regulatory Acceptance of Toxicological Test Methods (« Validation et acceptation réglementaire des méthodes d'essai toxicologiques ») (retiré à la date de publication du présent rapport)	ICCVAM, 1997
ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods (« Directives de l'ICCVAM pour la nomination et la soumission de méthodes d'essai nouvelles, révisées et alternatives »)	ICCVAM, 2003
Série de l'OCDE sur les essais et l'évaluation n° 34 : Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment (« Document d'orientation sur la validation et l'acceptation internationale de méthodes d'essai nouvelles ou actualisées pour l'évaluation des risques »)	OCDE, 2005
Recommended Procedures Regarding the CPSC's Policy on Animal Testing (« Procédures recommandées concernant la politique de la CPSC en matière d'essais sur les animaux »)	CPSC, 2012

Titre du document	**Référence**
FDA Predictive Toxicology Roadmap (« Feuille de route sur la toxicologie prédictive de la FDA »)	FDA, 2017a
Qualification of Medical Device Development Tools: Guidance for Industry, Tool Developers, and Food and Drug Administration Staff (« Qualification des outils de développement des dispositifs médicaux : conseils à l'attention de l'industrie, des développeurs d'outils et du personnel de la Food and Drug Administration »)	FDA, 2017b
EPA Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program (« Plan stratégique de l'EPA visant à promouvoir le développement et la mise en œuvre de méthodes de test alternatives dans le cadre du programme de la loi TSCA »)	EPA, 2018
ICCVAM Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States (« Feuille de route stratégique de l'ICCVAM pour établir de nouvelles approches destinées à évaluer la sécurité des produits chimiques et médicaux aux États-Unis »)	ICCVAM, 2018
Guidance Document on Good In Vitro Method Practices (GIVIMP) (« Document d'orientation sur les Bonnes pratiques de laboratoire appliquées aux études in vitro (GIVIMP) »)	OCDE, 2018
Guidance for Industry and Test Method Developers: CPSC Staff Evaluation of Alternative Test Methods and Integrated Testing Approaches and Data Generated from Such Methods to Support FHSA Labeling Requirements (« Directive à l'attention de l'industrie et des élaborateurs de méthodes d'essais : évaluation par le personnel de la CPSC des méthodes d'essais alternatives et des approches de test intégrées ainsi que des données obtenues par ces méthodes pour répondre aux exigences d'étiquetage de la loi FHSA »)	CPSC, 2020
Qualification Process for Drug Development Tools Guidance for Industry and FDA Staff (« Processus de qualification des outils de développement de médicaments à l'attention de l'industrie et du personnel de la FDA »)	FDA, 2020
EPA New Approach Methods Work Plan (« Plan de travail sur les nouvelles méthodes d'approche de l'EPA »)	EPA, 2021a
EPA Strategic Plan to Reduce the Use of Vertebrate Animals in Chemical Testing (« Plan stratégique de l'EPA visant à réduire l'utilisation d'animaux vertébrés dans les essais chimiques »)	EPA, 2021b

Titre du document	Référence
Advancing New Alternative Methodologies at FDA (« Promotion des nouvelles méthodes alternatives à la FDA »)	FDA, 2021a

2.0 Principaux concepts d'une validation flexible et adaptée à l'objectif des NAM

Les principes sous-jacents de la validation tels qu'ils sont décrits dans le GD 34 de l'OCDE restent fondamentaux, mais les processus de validation devraient permettre de développer efficacement et en temps opportun de NAM adaptées à l'objectif, fiables et fournissant des informations pertinentes pour l'espèce considérée. Le GD 34 de l'OCDE stipule en particulier que « les nouvelles méthodes d'essai sont soumises à une validation afin de garantir qu'elles sont scientifiquement robustes et répondent aux besoins réglementaires », c.-à-d. que les méthodes sont adaptées à l'objectif, ce qui, en plus d'éclairer les décisions réglementaires, pourrait inclure la sélection et la hiérarchisation des cas d'utilisation. La directive indique également que « le processus de validation doit être flexible et adaptable » et que la performance doit être « démontrée à l'aide d'une série de substances chimiques de référence » et « évaluée par rapport aux données de toxicité pertinentes existantes » (OCDE, 2005). Ainsi, la confiance dans les NAM ne devrait pas être considérée comme un statut universel accordé à l'issue d'un processus défini (p. ex. considérer qu'une NAM est « validée » après la réussite d'une étude par essai circulaire) ; elle devrait au contraire être comprise comme un processus évolutif et itératif nécessitant une communication entre les élaborateurs de méthodes, les décideurs en matière de réglementation et les organismes de validation (ICCVAM, 2018).

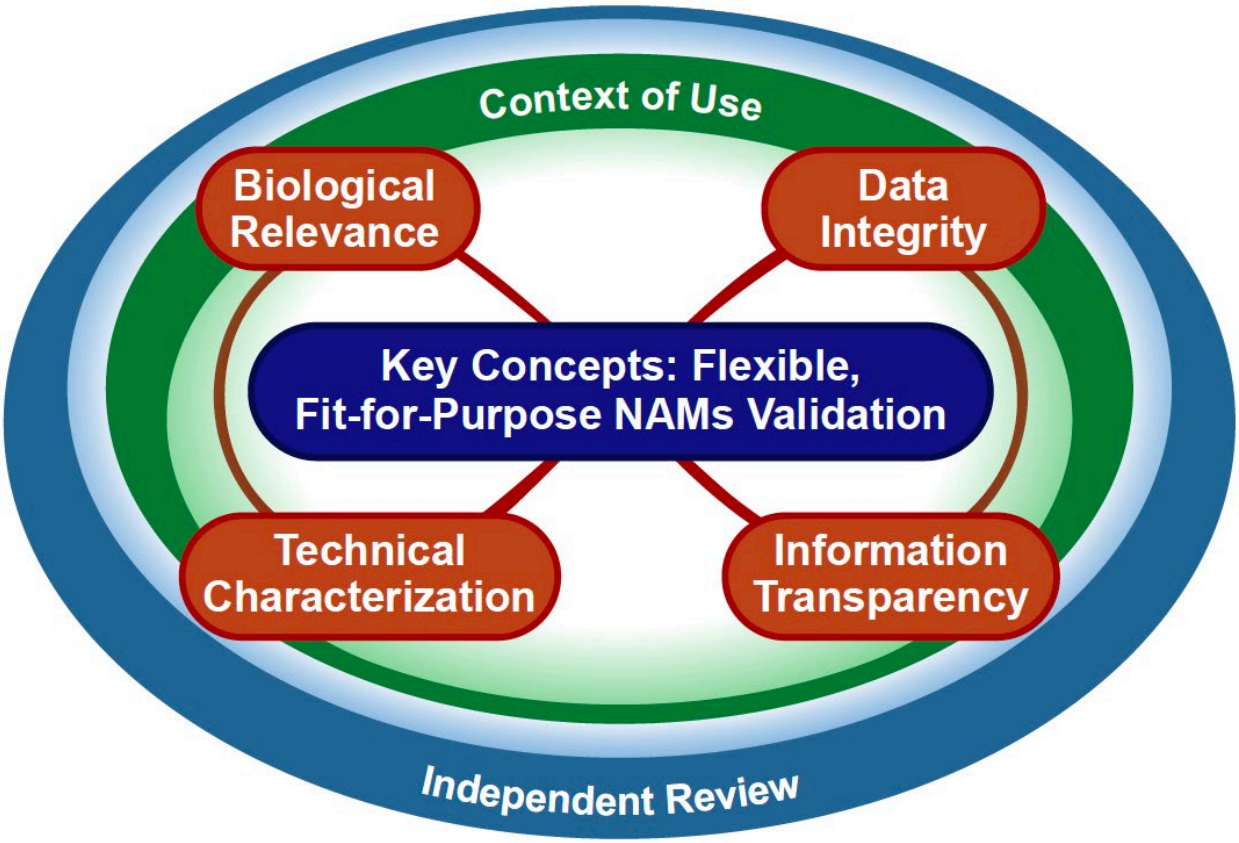


Figure 1. Concepts clés lors de l'élaboration et de la mise en œuvre de stratégies de validation flexibles et adaptées à l'objectif des NAM. D'après van der Zalm et al. (2022).

La Figure 1 montre les concepts clés à prendre en compte lors de la conception et de la mise en œuvre de stratégies de validation flexibles et adaptées à l'objectif. Ces concepts sont largement applicables à l'utilisation des NAM à différentes fins, mais devraient être adaptés en fonction du scénario et de l'application considérés. Un concept général à prendre en compte est le CU, à savoir l'objectif fixé pour la NAM (p. ex. criblage, identification des dangers, évaluation de la puissance ou point de départ pour l'évaluation quantitative des risques). Le CU influera sur la flexibilité du processus de validation et les modalités d'établissement de l'adéquation de la NAM à un CU donné. Les concepts clés appliqués à ce processus de validation flexible et adapté à l'objectif sont la pertinence biologique, la caractérisation technique, l'intégrité des données et la transparence de l'information. Comme le montre la figure, ces aspects sont liés et interdépendants. Enfin, toutes les étapes du processus de validation doivent faire l'objet d'un examen indépendant. Ces concepts clés sont semblables aux éléments essentiels proposés par van der Zalm et al. (2022) dans un cadre visant à établir la confiance scientifique dans les NAM. Toutefois, nous les adaptons ici pour qu'ils se prêtent à un plus large éventail d'applications des NAM. Chacun de ces concepts clés est examiné plus en détail dans la section 3.0.

3.0 Appliquer les concepts clés pour renforcer la confiance dans les NAM

La validation devrait être un processus robuste, mais flexible dans lequel la confiance scientifique est établie en déterminant l'adéquation d'une NAM à un CU particulier. Le cas échéant, des preuves devraient dans la mesure du possible être fournies pour indiquer que l'utilisation d'une méthode alternative fournira des informations équivalentes, voire meilleures que la méthode existante, et qu'elle donnera lieu à un examen réglementaire menant à des décisions qui protègent autant la santé humaine. Le CU détermine les critères détaillés et la mise en œuvre des concepts clés. Afin de déterminer si une NAM est adaptée à un objectif particulier, le CU et la biologie pertinente que la NAM est censée couvrir devraient tous deux être clairement définis. La NAM doit être dûment décrite et fournir des informations techniquement fiables qui sont biologiquement pertinentes et/ou protègent la santé pour le critère d'évaluation biologique ou le processus examiné. La composition des informations fournies par la NAM doit être suffisamment transparente pour permettre un examen indépendant garantissant leur intégrité et leur fiabilité. Si approprié et possible, le renforcement de la confiance dans les NAM peut comprendre une démonstration que celles-ci fournissent des informations d'une qualité et d'une pertinence équivalentes ou supérieures pour les espèces considérées aux fins de la prise de décisions réglementaires, en les comparant quantitativement ou qualitativement aux informations fournies par les méthodes d'essai traditionnelles sur les animaux. Pour éviter de s'en tenir au statu quo, il convient d'expliciter la possibilité que les NAM pourraient fournir des informations d'une qualité et d'une pertinence supérieures aux fins de la prise de décisions réglementaires que les méthodes d'essai traditionnelles sur les animaux. Cela nécessite un cadre de validation prévoyant cette possibilité, mais également les circonstances dans lesquelles une comparaison avec les données des méthodes d'essai traditionnelles sur les animaux pourrait ne pas être praticable.

3.1 Contexte d'utilisation

L'établissement du CU, à savoir l'utilisation prévue d'une NAM, comprend la rédaction d'une déclaration décrivant de manière complète et claire la manière dont la NAM est censée être utilisée et son objectif réglementaire (le cas échéant). Les agences fédérales américaines opèrent conformément à des lois et règlements propres à chaque agence (voir p. ex. Shaffer, 2021). Elles recourent par conséquent à des critères différents pour établir si une NAM est acceptable et applicable à leurs besoins individuels. Dans certains cas, même au sein d'une même agence (p. ex. un centre de la Food and Drug Administration [FDA] des États-Unis ou un bureau de l'Agence américaine de protection de l'environnement [EPA]), les besoins peuvent varier en fonction du domaine de réglementation. La situation est encore plus complexe au niveau international, car la réglementation en vigueur dans d'autres pays peut différer de celle des États-Unis ; les méthodes acceptables dans un pays pour un CU donné peuvent donc ne pas l'être ailleurs. Par conséquent, une NAM validée pour répondre à un critère d'évaluation biologique spécifique pour un (ou plusieurs) CU dans le cadre d'une directive/d'un règlement particulier pourrait ne pas être jugée acceptable en vertu d'une directive/d'un règlement différent pour d'autres agences, juridictions réglementaires et pays. Pour mieux cerner la diversité de ce

paysage réglementaire, l'ICCVAM a recueilli des informations auprès d'une variété d'agences tant nationales qu'internationales, afin de caractériser leurs besoins pour différents critères d'évaluation biologique de la toxicité (NIEHS, 2023a ; exemples présentés dans le Tableau 2).

Tableau 2. Manuscrits produits par les groupes de travail de l'ICCVAM détaillant les besoins des agences en matière d'essais dans des domaines sélectionnés

Cible des essais	**Référence**
Essais de toxicité aiguë	Strickland et al., 2018
Essais d'écotoxicité	Ceger et al., 2022
Extrapolation *in vitro-in vivo*	Chang et al., 2022
Essais de nanomatériaux à l'aide des NAM	Petersen et al., 2022a
Sensibilisation cutanée	Daniel et al., 2018; Strickland et al., 2019
Essais d'irritation cutanée et oculaire	Choksi et al., 2019

Les NAM relevant de différents CU, tels que le criblage/la hiérarchisation et l'identification des dangers, peuvent être différentes de celles conçues pour répondre à des questions spécifiques sur les mécanismes toxicologiques et/ou pour étayer les évaluations quantitatives des risques. Différents critères pourront donc être élaborés et appliqués à l'évaluation et à la validation des NAM pour chaque fin. De nombreuses nouvelles méthodes sont développées dans les milieux universitaires pour répondre à des questions de recherche fondamentale qui peuvent être de nature spécifique ou exploratoire. Parfois, il peut s'avérer que l'une de ces méthodes pourrait être utilisée dans des domaines appliqués sortant du cadre strict de la recherche fondamentale, notamment la prise de décisions réglementaires. Il est important que les élaborateurs de nouvelles méthodes rencontrent à la fois les agences fédérales et/ou les utilisateurs prévus ou les parties prenantes avant de valider une NAM afin de s'assurer que les études de validation conçues pour établir la confiance seront adaptées au CU visé.

D'autres NAM peuvent être plus utiles pour la découverte et le développement de produits. Une NAM peut être adoptée pour le criblage de nouvelles molécules afin de déterminer leurs effets désirables (p. ex. pharmacologiques) ou potentiellement indésirables (toxiques). Ces données peuvent être utilisées pour prendre des décisions concernant les molécules à évaluer plus avant. La standardisation des méthodes d'essai peut promouvoir l'innovation en concentrant les efforts et les ressources sur les découvertes technologiques ou thérapeutiques, plutôt que sur le développement ou les essais de méthodes. Ces utilisations des NAM ont tendance à ne pas relever de l'autorité réglementaire de la plupart des agences fédérales. Cependant, il est toujours recommandé d'appliquer les concepts clés visant à établir la confiance scientifique dans les NAM afin de garantir l'obtention de données de haute qualité. Il est reconnu que ces utilisations des NAM peuvent à la fois favoriser directement les principes des 3R et fournir une expérience

et des données préliminaires propices à la poursuite du développement d'une NAM et susceptibles de permettre son utilisation à des fins réglementaires.

3.2 Pertinence biologique

La pertinence d'une NAM décrit la relation entre l'essai et l'effet chez l'espèce cible, et indique si la méthode d'essai est significative et utile pour un objectif défini, ses limites étant identifiées. Une démonstration adéquate de la pertinence d'une NAM est importante pour contribuer à renforcer la confiance à son égard. La pertinence biologique peut être démontrée de différentes manières en fonction des informations disponibles, comme détaillé ci-dessous.

Les considérations relatives à la pertinence biologique détaillées dans les sous-sections ci-après sont les suivantes :

- Quel type d'information la NAM fournit-elle ? La biologie et les mécanismes menant au résultat/critère d'évaluation biologique sont-ils compris ?
- Quelles sont les données de référence disponibles pour évaluer le résultat que la NAM est censée éclairer ?
- Quels sont les éléments à prendre en compte pour déterminer le bien-fondé et les modalités de la comparaison des résultats d'une NAM à ceux d'une méthode de laboratoire établie ?

D'autres considérations relatives à la pertinence biologique pourraient également être importantes, en fonction des circonstances propres à la demande de validation.

3.2.1 Compréhension mécanistique

Il est important d'examiner la biologie de l'espèce considérée (généralement, mais pas seulement, humaine) lors de l'évaluation de la pertinence d'une NAM. Les comparaisons entre les informations fournies par la NAM et la biologie *in vivo* devraient être étayées, dans la mesure du possible, par les connaissances mécanistiques existantes (p. ex. une voie toxicologique impliquée dans les effets indésirables [AOP] ou un processus biologique ayant une importance toxicologique). L'association d'une NAM à une AOP établie par l'intermédiaire d'un événement déclenchant moléculaire ou d'un ou plusieurs événements clés peut aider à démontrer la pertinence biologique de la NAM. D'autre part, l'absence d'une AOP établie pour le résultat prédit par une NAM n'exclut pas nécessairement que la NAM soit utile, et ces informations complémentaires peuvent provenir d'indications mécanistiques tirées des données *in vivo* et de la compréhension de la biologie de l'espèce cible. Il est important qu'une NAM soit caractérisée de manière exhaustive et décrive clairement quel événement biologique est mesuré et le lien avec l'effet indésirable ou le danger concerné.

La description de la NAM devrait indiquer la plausibilité biologique du modèle destiné à prédire les résultats *in vivo* et/ou fournir un lien mécanistique avec un processus, un mécanisme ou une AOP biologique. Par exemple, si une NAM est censée prédire le risque qu'un produit pharmaceutique humain provoque des malformations fœtales ou une mortalité embryofœtale, les

mécanismes du développement embryofœtal (p. ex. migration cellulaire, différenciation, vasculogenèse, neurulation, gastrulation) et le rapport avec les effets indésirables ultérieurs sur le développement qu'étudie le modèle devraient être compris (FDA, 2021b). Bien que la relation avec l'effet *in vivo* prévu puisse être de nature corrélative, il est plus difficile de renforcer la confiance dans les prédictions de la NAM à partir de relations uniquement corrélatives ou empiriques ; par conséquent, les essais dont la pertinence biologique est clairement établie pour le processus évalué sont préférés. L'absence de compréhension de la pertinence biologique et mécanistique d'une NAM pourrait restreindre son applicabilité aux limites étroitement définies par les données utilisées pour valider la NAM et rendre difficile l'extension des NAM à des classes chimiques autres que celles utilisées pour établir et valider la NAM.

Une AOP est un cadre d'organisation utile pour établir un lien entre les perturbations moléculaires et cellulaires et les effets indésirables sur la santé, et peut être utilisée pour développer et ancrer des NAM représentant des processus toxicologiques importants. Il existe des exemples de critères d'évaluation présentant des AOP bien établies, notamment la sensibilisation cutanée (Kleinstreuer et al., 2018). Pour les critères d'évaluation ne s'appuyant pas sur des AOP bien établies, il est possible d'envisager la pertinence mécanistique de la NAM sur la base de facteurs tels que les voies ou les processus biologiques (Tableau 3).

Tableau 3. Exemples de critères d'évaluation pour lesquels la pertinence biologique et mécanistique des NAM a été démontrée*

Critère d'évaluation	**Résumé**	**Références**
Sensibilisation cutanée	Le critère d'évaluation s'appuie sur une AOP humaine pertinente bien développée par rapport à laquelle des approches définies combinant plusieurs NAM sont cartographiées et décrites dans la ligne directrice n° 497 de l'OCDE.	Kleinstreuer et al., 2018 ; OCDE, 2021a
Perturbation endocrinienne	Des modèles de voies établis utilisant des NAM complémentaires dans le cadre d'une stratégie intégrée sont disponibles pour l'activité des récepteurs des œstrogènes et des androgènes. L'EPA accepte ces NAM pour le criblage de niveau 1 dans le cadre du programme de dépistage des perturbateurs endocriniens.	Judson et al., 2015 ; Kleinstreuer et al., 2017 ; EPA, 2023

Critère d'évaluation	**Résumé**	**Références**
Neurotoxicité développementale	Les AOP disponibles pour ce critère d'évaluation complexe sont limitées. En revanche, une batterie de NAM couvrant les processus critiques du neurodéveloppement humain a été développée. L'OCDE a publié un document d'orientation sur la batterie, qui comprend des études de cas sur les approches intégrées en matière d'essai et d'évaluation (IATA).	Crofton and Mundy, 2021 ; OCDE, 2022a ; OCDE, 2023
Toxicité par inhalation	Une approche alternative utilisant un essai *in vitro* sur cellules humaines et une modélisation informatique a été utilisée pour déterminer un point de départ pour l'évaluation des risques pour la santé humaine par l'EPA. Cette approche a également été publiée dans le cadre d'une étude de cas d'IATA de l'OCDE.	Corley et al., 2021 ; EPA, 2021c ; OCDE, 2022b
Irritation oculaire	Document d'orientation décrivant un cadre d'essai alternatif pour évaluer le pouvoir irritatif oculaire des pesticides et des formulations pesticides. Les méthodes d'essai *in vivo*, *in vitro* et *ex vivo* disponibles ont été examinées en fonction de leur pertinence vis-à-vis de l'anatomie oculaire humaine et des mécanismes de toxicité.	EPA, 2015 ; Clippinger et al., 2021
Irritation cutanée	Document d'orientation proposant une IATA pour la corrosion et l'irritation cutanées sur la base des informations existantes, des propriétés physicochimiques et d'autres méthodes non expérimentales.	OCDE, 2014

* Se reporter aux directives spécifiques à chaque agence pour l'acceptation des différentes NAM.

Dans l'idéal, la description de la relation entre la NAM et l'effet biologique considéré devrait être basée sur les informations disponibles relatives à la biologie ou au mécanisme d'action pertinent pour le critère d'évaluation concerné chez l'espèce considérée. Cela pourrait ne pas être possible dans certains cas, la comparaison ne pouvant être basée que sur les données se rapportant à une espèce différente de celle considérée. Cependant, la physiologie de l'espèce considérée peut différer de celle de l'espèce de substitution existante, ce qui souligne encore

davantage la nécessité d'intégrer une compréhension mécanistique dans l'évaluation de la NAM. Par exemple, les caractéristiques anatomiques et physiologiques de l'œil du lapin diffèrent de celles de l'œil humain (Clippinger et al., 2021) et les fenêtres de sensibilité au cours du développement du cerveau varient selon les espèces (Smirnova et al., 2014 ; Tsuji et Crofton, 2012). Par conséquent, la pertinence biologique vis-à-vis de l'espèce considérée et les principaux aspects relatifs à l'exposition devraient être pris en compte lors de l'évaluation tant de la NAM que de la méthode d'essai de référence existante. Les évaluations de la pertinence biologique et mécanistique pourraient prendre en compte, par exemple, la pertinence du type de cellule utilisé, les caractéristiques physiologiques de l'organe/tissu à l'étude, ou la présence de métabolites propres à l'espèce associés à la substance testée (en notant que les métabolites peuvent varier d'une espèce à l'autre ou être différents de ceux produits dans une NAM). En outre, la capacité d'une méthode à évaluer un mode d'action ou un mécanisme spécifique à une espèce particulière pourrait être examinée (Hartung, 2010 ; Madia et al., 2021 ; Parish et al., 2020).

Au-delà de la translation entre espèces, les variations entre individus (p. ex. donneurs de cellules ou de tissus utilisés dans les essais *in vitro*) peuvent elles aussi générer une variation biologique qui peut affecter la réponse mesurée, voire apporter des éclairages supplémentaires sur la variabilité interindividuelle au sein d'une population (Harrill, 2020). Bien que cela ne soit pas abordé de manière approfondie dans le présent document, les utilisateurs et les élaborateurs de NAM devraient savoir que l'observation d'un effet (en ce qui concerne l'ampleur et, dans certains cas, la présence de l'effet observé) ainsi que le point de départ ou la dose à laquelle l'effet se produit peuvent être modulés par la variation de la séquence génétique présente sur un panel de lignées cellulaires (Chiu et al., 2017). L'utilisation de panels de lignées cellulaires génétiquement diversifiées dans les tests cellulaires peut offrir des avantages pour comprendre la réponse phénotypique et la dynamique de la toxicité dans la population (Harrill et McAllister, 2017), révéler des variants de séquences génétiques qui confèrent une sensibilité (Frick et al., 2015), permettre l'identification de sous-populations particulièrement sensibles (Church et al., 2015) et faciliter la mesure de la variabilité toxicocinétique et toxicodynamique (Rusyn et al., 2022). Bien que des lignées cellulaires dérivées de tumeurs soient parfois utilisées dans des modèles *in vitro*, leur composition génétique peut ne pas représenter de manière adéquate la variation génétique au sein de la population. Les recommandations visant à incorporer cette diversité dans les NAM dépendront du contexte. En conséquence, des plans d'étude adaptés aux objectifs devraient être développés en coopération avec des généticiens, afin de garantir une couverture adéquate de la variation de l'espèce cible, ainsi qu'avec des bioéthiciens et les groupes de parties prenantes concernés si la variation génétique humaine doit être explorée (voir par exemple ICCVAM, 2022). Les recommandations spécifiques à l'évaluation des variations interindividuelles dépassent le cadre de ce rapport et continueront probablement d'évoluer au rythme des avancées scientifiques dans ce domaine.

Dans certains cas, des AOP complexes ou des résultats dérivés de plus d'une AOP pourraient nécessiter l'évaluation de plusieurs aspects du mécanisme de l'effet indésirable afin d'obtenir une prévision adéquate des résultats. Les NAM présentent généralement des limitations qui ne leur permettent pas d'être prédictives ou techniquement applicables pour toutes les classes chimiques. Une NAM peut également être limitée en ce qui concerne la complexité de la biologie qu'elle peut représenter. La combinaison de plusieurs NAM présentant des

performances, une applicabilité et des couvertures biologiques différentes dans le cadre d'approches définies (AD) peut améliorer leur pouvoir prédictif (OCDE, 2017). Les NAM appropriées peuvent pour ce faire être associées aux principaux événements d'une AOP afin de garantir une représentation mécanistique suffisante pour prédire les effets indésirables pertinents chez l'espèce considérée. La ligne directrice nº 497 : Approches définies pour la sensibilisation cutanée (OCDE, 2021a) fournit un exemple d'association des NAM aux événements clés d'une AOP débouchant sur une AD visant à répondre à un critère d'évaluation réglementaire.

3.2.2 Composés de référence

Les composés de référence peuvent avoir un large éventail d'objectifs potentiels, allant de la réalisation d'un essai à l'évaluation de sa qualité technique et de sa pertinence biologique, comme le montre le Tableau 4. Un composé de référence peut être adapté à un objectif, mais inadéquat pour un autre. Par exemple, un composé de référence peut être efficace pour une méthode donnée comme témoin positif (p. ex. essai de neurotoxicité développementale) même s'il n'entraînerait pas l'effet indésirable impliqué chez l'espèce cible (p. ex. en raison d'une incapacité à traverser la barrière hématoencéphalique) et pourrait donc ne pas être approprié comme composé de référence pour un critère d'évaluation biologique. Selon le CU, les composés de référence peuvent inclure des substances chimiques monoconstituantes, des mélanges ou des extraits complexes (p. ex. issus de dispositifs médicaux).

Tableau 4. Utilisations sélectionnées de composés de référence

Ensembles de composés distincts*	**Objectif(s)**	**Critères de sélection des composés**
Témoin(s) positif(s)	Vérifier si la méthode fonctionne selon les attentes.	Composés qui se sont avérés provoquer une modification systématique, mesurable et statistiquement significative d'un résultat d'essai précis dans les conditions de la méthode.
Composés de performance	Évaluer dans quelle mesure une méthode nouvelle ou modifiée peut être comparée à une méthode similaire établie (« me-too »).	Ensemble de composés systématiquement associé à une réponse (ou à une absence de réponse) avec la méthode déjà établie.
Composés d'épreuve de compétence	Évaluer la performance d'un nouveau laboratoire à l'aide d'une méthode établie.	Composés systématiquement associés à une réponse (ou à une absence de réponse) avec la méthode établie ; peuvent être un sous-ensemble des composés de performance.
Composés de référence pour les critères d'évaluation biologiques	1) Évaluer la pertinence biologique d'une méthode. 2) Comparer la concordance entre des méthodes conçues pour mesurer un résultat relevant du même critère d'évaluation biologique *in vivo*, mais qui utilisent un contexte différent (p. ex. espèce, méthode ou modalité de mesure différente).	1) Composés associés de manière démontrée à un effet (résultat positif) ou à une absence d'effet (résultat négatif) *in vivo* pour le critère d'évaluation considéré (dans l'idéal pour l'organisme cible considéré ou une espèce de comparaison appropriée). 2) Un ensemble commun de composés qui ont été testés pour le critère d'évaluation considéré dans au moins une méthode fiable.

***** Les composés admissibles à ces ensembles peuvent se chevaucher.

Pour démontrer la validité scientifique d'une NAM, il est essentiel d'évaluer sa performance par rapport aux méthodes d'essai existantes, souvent en testant des composés de référence pour des critères d'évaluation biologiques dont les activités biologiques sont bien caractérisées et

comprises. Les composés de référence pour des critères d'évaluation biologiques comprennent généralement des composés pour lesquels il a été démontré qu'ils produisaient des effets indésirables, dans l'idéal avec une gamme de puissances, ou ne produisaient pas de tels effets (Browne et al., 2019).

L'identification de composés de référence pour des critères d'évaluation biologiques ayant des effets avérés sur des résultats apicaux particuliers peut être difficile. C'est particulièrement vrai chez l'homme, pour lequel les données font souvent défaut, en raison des efforts consentis dans le domaine de la toxicologie pour prévenir l'exposition humaine à des substances chimiques potentiellement dangereuses. Le succès de ces efforts signifie que les données disponibles chez l'homme sont limitées, ou que ces substances n'ont pas été caractérisées ou mesurées de manière adéquate dans les populations humaines. Lorsque de telles données sont disponibles chez l'homme, elles peuvent provenir d'expositions accidentelles ou faibles, qui génèrent souvent des données quantitatives limitées sur les niveaux d'exposition ou les résultats. Les études de cas peuvent fournir des indications sur le danger, mais ne sont généralement pas adaptées pour établir des corrélations claires entre la cause et l'effet.

Malgré ces difficultés, des efforts devraient être déployés pour identifier des composés ayant des effets sur l'homme liés au résultat ou au mécanisme prédit par la NAM en cours de développement. Même si ces données ont une capacité limitée à étayer l'évaluation quantitative de la méthode, elles peuvent fournir un contrôle qualitatif de la pertinence de la NAM. Des données épidémiologiques humaines de haute qualité peuvent également être difficiles à obtenir et à aligner sur les données toxicologiques ; toutefois, les données épidémiologiques sont utiles pour identifier des substances chimiques présentes dans l'environnement qui sont fortement associées à des effets nocifs sur la santé humaine. Des cadres de référence fondés sur des données probantes ont été développés en vue de la compilation de grandes quantités d'informations, puis de la sélection transparente et objective d'ensembles de données plus restreints et pertinents (Wikoff et al., 2020), améliorant ainsi l'accessibilité des données. Ces cadres comprennent des cartes de preuves systématiques, qui fournissent un aperçu général d'un socle de données probantes, et des revues systématiques ultérieures, qui permettent une évaluation plus ciblée et complète d'une question de recherche particulière, et l'intégration plus rapide des données épidémiologiques humaines avec d'autres sources d'informations existantes Wolffe et al., 2019). En combinaison avec des sources de données mécanistiques et des données existantes chez l'animal, ces éléments pourraient servir à définir des composés de référence pour des critères d'évaluation biologiques contribuant à établir la confiance scientifique dans les NAM (voir Krishna et al., 2021 pour un exemple).

Dans la mesure du possible, les composés qui se sont avérés provoquer le ou les effets pertinents chez l'espèce considérée devraient être utilisés comme témoins positifs lors de l'évaluation de la robustesse et de la pertinence de la NAM. Petersen (2021) a décrit les caractéristiques à prendre en compte lors du choix des témoins positifs (p. ex. substances chimiques testées à l'aide de mesures de contrôle en cours de processus chaque fois que l'essai est effectué), et bon nombre de ces considérations seront également pertinentes pour les composés de référence employés dans l'évaluation d'une NAM. Les composés de référence positifs et négatifs pour les critères d'évaluation biologiques, les composés de performance et les composés d'épreuve de

compétence devraient être inclus et sélectionnés avec le même soin. En raison du biais historique de la littérature en faveur des résultats positifs (sachant que les résultats positifs signalés dans la littérature ne sont pas tous de vrais positifs), il pourrait être plus ardu d'identifier des composés de référence fiables qui sont négatifs pour le critère d'évaluation cible. La publication des efforts de préparation des données axés sur les résultats négatifs devrait être sérieusement envisagée ; la réalisation de tests des composés jusqu'à des concentrations relativement élevées sur un large éventail de cibles pourrait fournir des informations précieuses pour aider à surmonter ce problème.

Une liste unique pourrait ne pas permettre de couvrir tous les CU et la sélection des composés de référence devrait tenir compte des besoins réglementaires de chaque CU. Dans certains cas, il peut également être utile de définir des sous-ensembles de composés de référence pour les critères d'évaluation biologiques en fonction du mécanisme évalué (plutôt que du critère d'évaluation apical) pour chaque NAM d'une batterie. Par exemple, peu de composés peuvent être considérés comme des composés de référence pour la neurotoxicité développementale, et chaque processus évalué dans la batterie (p. ex. prolifération ou excroissance d'axones) possède ses propres composés de référence pour les critères d'évaluation biologiques propres à l'essai. Par conséquent, un composé de référence pour un critère d'évaluation biologique peut être positif dans un essai, mais négatif dans un autre essai évaluant des mécanismes ou des processus différents.

Des listes de composés de référence (p. ex. pour évaluer la transférabilité) sont exigées dans le cadre du processus de validation officiel décrit par l'OCDE. Cependant, la compilation de composés de référence fiables est gourmande en temps et en ressources. Les efforts visant à faciliter l'élaboration rapide de listes organisées de composés de référence par le biais de revues systématiques rigoureuses et de processus d'automatisation, lorsque cela est possible, ont permis une évaluation plus robuste de la sensibilité et de la spécificité des NAM (Judson et al., 2019 ; Thomas et al., 2019). Un exemple de liste de composés de référence pour des critères d'évaluation biologiques en vue de la qualification des NAM est disponible dans la révision 2021 par la FDA du document S5(R3) Detection of Reproductive and Developmental Toxicity for Human Pharmaceuticals Guidance for Industry (« Ligne directrice pour l'industrie : Détection de la toxicité reproductive et développementale pour les produits pharmaceutiques utilisés chez l'homme ») du Conseil international d'harmonisation des exigences techniques pour l'enregistrement des médicaments à usage humain (CIH) (FDA, 2021b). La liste comprend des composés donnant des résultats positifs et négatifs. La ligne directrice indique que ces composés, parmi d'autres, peuvent être utilisés pour étayer la qualification d'un essai ou d'une batterie d'essais alternatif dans des CU particuliers.

3.2.3 Comparaison avec les méthodes existantes sur les animaux de laboratoire

La norme visant à établir la confiance scientifique dans une NAM et à obtenir son acceptation réglementaire a généralement consisté à examiner si la NAM pouvait fournir des informations d'une utilité, d'une qualité scientifique et/ou d'une pertinence équivalentes ou supérieures à celles de la méthode d'essai existante utilisée pour la prise de décisions réglementaires (en fonction du cadre réglementaire de chaque agence). Par exemple, l'un des critères du GD 34 de

l'OCDE pour valider toute nouvelle méthode d'essai est le suivant : « la méthode génère des données à des fins d'évaluation des risques qui sont au moins aussi utiles, et de préférence meilleures, que celles obtenues à l'aide des méthodes existantes. Le niveau de protection de la santé humaine ou de l'environnement sera ainsi comparable, voire supérieur » (OCDE, 2005). La loi américaine Toxic Substances Control Act modifiée (section 4(h)(1)(B)) inclut des considérations spécifiques aux NAM et impose à l'EPA d'encourager et de faciliter « l'utilisation de méthodes et de stratégies d'essai scientifiquement valables qui réduisent ou remplacent l'utilisation d'animaux vertébrés, tout en fournissant des informations d'une qualité et d'une pertinence scientifiques équivalentes ou supérieures à même d'étayer les décisions réglementaires » (15 USC §2601, 2016).

Historiquement, le concept « équivalent ou meilleur » repose sur une comparaison directe avec les données des tests traditionnels sur les animaux. Cependant, pour atteindre l'objectif d'une « meilleure » information, il doit être reconnu qu'une NAM peut ne pas fournir les mêmes renseignements que la méthode d'essai traditionnelle sur les animaux et que ses résultats peuvent ne pas correspondre directement aux résultats de cette dernière (voir p. ex. Clippinger et al., 2021 ; Hoffmann et al., 2018, 2008 ; ICCVAM, 2018 ; Kolle et al., 2017 ; Petersen et al., 2022b ; Piersma et al., 2018 ; Prior et al., 2019 ; Sewell et al., 2017). Dans certains cas, la NAM peut fournir des informations biologiquement pertinentes, des informations mécanistiques ou des critères d'évaluation suffisamment sensibles adaptés au processus de prise de décision réglementaire, et une comparaison avec les données des méthodes d'essai traditionnelles sur les animaux peut ne pas être nécessaire. Par exemple, certaines cibles pharmacologiques ou toxicologiques peuvent ne pas exister chez les espèces non humaines, auquel cas les études sur les animaux ne permettent pas d'évaluer les effets humains potentiels médiés par de telles cibles. Des NAM propres à l'homme pourraient fournir des informations biologiquement pertinentes dans ces cas. En outre, les NAM fournissent souvent des informations mécanistiques plutôt que des données sur les critères d'évaluation apicaux mesurés dans les méthodes d'essai sur les animaux (à titre d'exemple, alors qu'une réduction du poids corporel observée chez un animal peut ne pas élucider le mécanisme sous-jacent de la toxicité, ces informations mécanistiques pourraient être fournies par des NAM). Il peut également arriver que le modèle animal permette de mesurer un critère d'évaluation biologique complexe qui est pertinent pour le CU, mais n'est pas couvert de manière adéquate par la NAM en question. Dans de tels cas, une comparaison avec les méthodes traditionnelles d'essai sur les animaux peut être l'option la plus opportune, en particulier lorsque le mécanisme n'est pas entièrement compris. L'objectif de l'évaluation est de démontrer que la NAM fournit des informations qui mènent à une décision réglementaire semblable à celle qui serait prise sur la base des méthodes existantes.

Les comparaisons entre les NAM et les méthodes existantes sur les animaux de laboratoire, le cas échéant, devraient tenir compte de la fiabilité et de la reproductibilité des méthodes d'essai sur les animaux de référence, en clarifiant notamment les raisons de la variance observée dans les deux types de méthodes, si possible. Lorsqu'elles sont disponibles, l'utilisation de données de référence relatives à l'espèce considérée permet d'évaluer une NAM par rapport à la réponse propre à l'espèce. Si des données chronologiques *in vivo* sont disponibles, des NAM telles que des systèmes microphysiologiques peuvent être utilisées pour simuler la toxicocinétique, après quoi l'extrapolation informatique du système microphysiologique peut être comparée aux

données animales ou humaines. La modélisation pharmacocinétique basée sur la physiologie permet d'estimer les concentrations internes à partir de doses de référence et de les comparer aux concentrations d'activité *in vitro*. Ces modèles peuvent également être appliqués dans le cadre d'une approche de dosimétrie inverse pour effectuer une extrapolation *in vitro-in-vivo* (IVIVE) et prédire les doses administrées équivalentes qui entraîneraient des concentrations plasmatiques ou dans les tissus cibles où une bioactivité est observée dans la NMA. Quand l'espèce considérée est celle humaine, des preuves épidémiologiques, cliniques ou observationnelles de haute qualité des effets peuvent être utiles pour fournir des informations sur la manière de bâtir la confiance scientifique en faveur de la NAM. Cependant, de telles données de référence humaines sont indisponibles pour la plupart des critères d'évaluation et des substances chimiques, et les comparaisons doivent souvent reposer sur des données provenant d'une espèce différente. Les données des études sur les animaux peuvent être préparées et comparées pour produire des listes de normes de référence comprenant des résultats reproductibles, robustes et pertinents. Dans certains contextes, tels que les applications écologiques, l'homologie de séquences de cibles moléculaires ou la conservation de mécanismes biologiques peuvent être utiles pour étayer une comparaison avec des données dérivées d'espèces différentes (Farmahin et al., 2013 ; LaLone et al., 2016).

Plusieurs publications ont évalué les résultats de méthodes d'essai de référence sur des animaux pour une gamme de critères d'évaluation, montrant qu'ils présentaient des niveaux variables de reproductibilité (Browne et al., 2018 ; Dumont et al., 2016 ; Karmaus et al., 2022 ; Kleinstreuer et al., 2018 ; Luechtefeld et al., 2016 ; Pham et al., 2020 ; Rooney et al., 2021). La variabilité observée pourrait refléter des facteurs biologiques inhérents, des erreurs de transcription, des différences de protocole ou des variations de la pureté de la substance à tester, par exemple. Les résultats contradictoires entre la méthode d'essai de référence sur l'animal et les données des NAM devraient être expliqués dans la mesure du possible, y compris du point de vue de la biologie de l'espèce considérée. Cela est nécessaire pour fournir un contexte réaliste concernant le potentiel des données sur les animaux de laboratoire et, par conséquent, définir des attentes appropriées relatives à la performance maximale des NAM qui sont comparées à la méthode d'essai de référence (Browne et al., 2019).

La valeur relative de la comparaison des NAM avec les tests traditionnels sur les animaux et la pertinence biologique et mécanistique des NAM pour l'espèce considérée devraient être prises en compte sur la base de la qualité des données disponibles. En réalité, dans de nombreux cas, l'établissement de la confiance dans une NAM exige de considérer l'utilisation historique des études sur les animaux et la nécessité de comparaison aux méthodes existantes comme une source importante de données probante. Dans de nombreux cadres légaux et réglementaires, les informations dérivées des NAM ne peuvent pas être moins protectrices que les méthodes existantes. Dans l'idéal, la méthode sera plus prédictive et pourrait permettre de générer de façon plus rapide et complète des données pertinentes sur de nombreuses substances chimiques dans des situations où les données pourraient par ailleurs être rares ou absentes, suscitant ainsi la confiance des communautés des régulateurs et des parties prenantes. Certains CU pourraient autoriser des exceptions à cette règle ; toutefois, l'acceptation de ces NAM en tant que méthodes qualifiées sera soumise aux exigences de l'agence ou de l'organisme de réglementation concerné. L'étendue de ces comparaisons peut dépendre de la quantité et de la qualité des

données disponibles et du degré de compréhension de la pertinence pour l'espèce considérée, tant pour la NAM que pour la méthode de référence. Lorsque la performance de la NAM est évaluée sur la base de la méthode d'essai sur l'animal de référence, l'ampleur de la variance acceptable de la NAM doit être considérée à la lumière de la variance observée des données *in vivo* et du CU prévu.

3.3 Caractérisation technique

La caractérisation technique est un élément clé de la mise au point des NAM destinées à une large utilisation. Elle comprend une évaluation des sources de variabilité de la NAM, la conception de l'essai de manière à inclure des mesures de contrôle pertinentes, l'évaluation de la gamme de substances à tester pour lesquelles l'essai peut être utilisé (c.-à-d. le domaine d'application) et le développement d'une approche d'analyse appropriée des données statistiques. Pour démontrer la validité scientifique d'une NAM, il est essentiel que l'essai soit suffisamment bien caractérisé d'un point de vue technique pour garantir sa robustesse, sa fiabilité et sa reproductibilité. La NAM comprend à la fois le système de test lui-même et la méthode de quantification du critère d'évaluation. Certains aspects de la caractérisation technique recoupent des concepts clés décrits précédemment, tels que la pertinence biologique, qui guide la sélection de composés de référence pour différents objectifs (Tableau 4). L'examen de la performance d'une NAM par rapport à un ensemble bien défini de composés de référence pour des critères d'évaluation biologiques possédant des bioactivités pertinentes a été décrit dans la section 3.2.2. La présente section se concentrera donc sur des aspects techniques tels que les outils de qualité (Ishikawa, 1985), le développement de méthodes, la documentation et les normes.

3.3.1 Intégration d'outils de qualité sélectionnés

D'une manière générale, la caractérisation technique des NAM s'inscrit dans un cadre comportant des étapes principales (Figure 2) : 1) vérification initiale de la pertinence scientifique de la NAM ; 2) évaluation conceptuelle ; 3) évaluation intralaboratoires ; 4) analyse et communication des données statistiques ; et 5) évaluation interlaboratoires (si nécessaire[1]). Les résultats des étapes 2 à 5 sont interdépendants et peuvent donc avoir des répercussions sur toutes les autres étapes. Par exemple, les résultats des tests interlaboratoires pourraient révéler que des tests intralaboratoires supplémentaires devraient être effectués ou que le protocole doit être révisé pour inclure une nouvelle mesure de contrôle.

[1] Il pourrait y avoir des exemples de méthodes pour lesquelles une évaluation interlaboratoires n'est pas nécessaire ou réalisable, comme les essais de criblage robotisés à haut débit et grande échelle et les approches *in silico* basées sur l'apprentissage automatique.

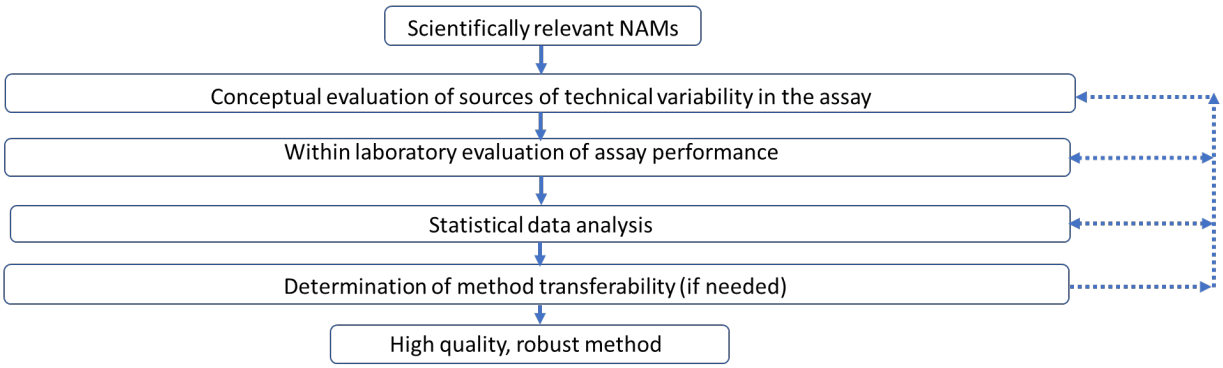


Figure 2. Cadre pour le développement de NAM robustes. Les lignes continues indiquent l'ordre suggéré des étapes à suivre. Les lignes pointillées indiquent les mesures pouvant être envisagées s'il est nécessaire de réévaluer l'une des étapes précédentes. Adapté et réimprimé avec la permission de Petersen et al. (2022b).

La première étape de vérification comprend une évaluation de la pertinence biologique d'un essai, de son CU potentiel et de la possibilité d'atteindre un niveau de qualité technique suffisant. Si la NAM n'est pas associée à un CU clair, manque de pertinence biologique ou présente des problèmes de qualité technique importants et potentiellement insolubles, la NAM pourrait ne pas être adaptée à un usage réglementaire. Au cours de la phase d'évaluation conceptuelle, la NAM est examinée pour évaluer les sources de variabilité attendues et concevoir des mesures de contrôle. Lors de l'évaluation intralaboratoires, des expériences peuvent être réalisées pour évaluer la robustesse de l'essai, établir une plage typique pour les mesures de contrôle (p. ex. témoin négatif, excipient témoin et témoin positif) et déterminer s'il existe des interactions entre les mesures de contrôle et le résultat du test. Les informations issues de l'évaluation conceptuelle et de l'évaluation intralaboratoires permettent de concevoir un modèle statistique pour évaluer le résultat moyen de l'essai, caractériser la variabilité et établir des critères de réponse positive ou négative ainsi que la confiance statistique pour cette détermination.

Enfin, certaines NAM peuvent devoir être soumises à une évaluation de leur transférabilité par le biais d'essais interlaboratoires. Les essais interlaboratoires peuvent révéler que des étapes d'un protocole sont interprétées différemment selon les laboratoires et indiquer les révisions pouvant être apportées pour améliorer la qualité technique de la NAM. Les études de transférabilité peuvent également identifier les problèmes liés à des différences dans l'exécution ou la mise en œuvre technique d'un protocole (p. ex. pipettage dans un laboratoire, procédure automatisée dans un autre) qui nécessitent des modifications.

Le Tableau 5 répertorie plusieurs outils de qualité qui peuvent être utiles pour la caractérisation technique d'une NAM. Un traitement plus détaillé de ce sujet est disponible dans Petersen et al. (2022b). Certaines informations générées par ces outils peuvent ne pas être divulguées par l'élaborateur de la méthode d'essai, d'autres peuvent être demandées par l'agence aux fins de l'examen.

Tableau 5. Outils de qualité souvent utilisés pour la caractérisation technique des NAM*

Outil de qualité	**Descriptif**	**Avantage/utilité**
Organigrammes	Schématisation de chaque étape d'un protocole.	Optimisation de l'identification et de la couverture des étapes du protocole pouvant être soumises à des mesures de contrôle.
Analyse des causes et des effets	Passage en revue de la littérature et du contexte des essais ; schématisation de toutes les sources de variabilité attendues.	Identification des principales sources de variabilité et des aspects d'une méthode potentiellement difficiles à standardiser. Guide les tests de robustesse et la sélection des mesures de contrôle.
Cartes de contrôle	Mesures de contrôle pour l'évaluation de la performance technique, y compris expériences préliminaires ponctuelles, mesures périodiques à une fréquence prédéterminée et mesures de contrôle en cours de processus à chaque réalisation de l'essai. Les cartes de contrôle répertorient les mesures de contrôle au fil du temps.	Test des biais potentiels, évaluation de la performance et de l'étalonnage des instruments et mesure des principales sources de variabilité au fil du temps et entre les tests.
Feuilles de relevé	Enregistrement des données clés, des métadonnées et des mesures de contrôle.	Surveillance des mesures de contrôle en cours de processus et contribution à la résolution des problèmes lors de leur survenue, promotion de l'analyse et de la reproductibilité des données.
Diagrammes de dispersion	Représentation graphique de toutes les données provenant des mesures de contrôle et des résultats de la substance testée.	Détermination de l'existence d'une interaction entre différentes mesures de contrôle en cours de processus ou entre ces mesures de contrôle et les résultats de la substance à tester.

* Des détails et des exemples sont fournis dans l'Annexe B.

3.3.2 Meilleures pratiques en matière de contrôle qualité

Cette section décrit les facteurs techniques que les agences fédérales peuvent considérer comme les meilleures pratiques pour l'évaluation des NAM, notamment les AD et les approches intégrées en matière d'essai et d'évaluation (IATA). Les facteurs énumérés ici peuvent ou non s'appliquer à toutes les NAM. En outre, une agence peut avoir besoin d'autres facteurs non mentionnés dans la ligne directrice, tels que des informations ou des tests supplémentaires, lors de l'évaluation de la NAM. Les élaborateurs et/ou promoteurs sont invités à communiquer directement avec l'agence fédérale à laquelle ils ont l'intention de soumettre des méthodes ou des données afin de déterminer les facteurs supplémentaires éventuellement nécessaires à l'examen.

La présentation d'une NAM à des fins d'évaluation réglementaire devrait inclure une description de toute étude intralaboratoires ou interlaboratoires éventuelle. Elle devrait indiquer si la NAM a été comparée à une méthode de référence : autre NAM, données *in vivo* sur des animaux ou données humaines. Des informations sur la méthode de référence utilisée pour la comparaison (c.-à-d. l'ensemble de données de référence) peuvent également être nécessaires lors de l'évaluation. L'évaluation d'une NAM peut être réalisée au moyen de plusieurs processus différents et implique de documenter, à l'aide d'enquêtes de laboratoire spécifiques, que les caractéristiques de performance d'une méthode sont appropriées et fiables pour les applications prévues. L'acceptabilité des données est directement liée aux critères utilisés pour évaluer la méthode.

En plus de servir à documenter la performance et à caractériser l'applicabilité de la NAM, de bonnes pratiques scientifiques, techniques et de qualité garantissent que le processus d'évaluation global est efficient et efficace, ce qui accroît la confiance dans la méthode proposée. Les élaborateurs devraient conserver toutes les informations nécessaires à l'étalonnage, au fonctionnement et à la maintenance de tout équipement, telles que les manuels des équipements et des logiciels, les certificats et garanties de confirmation de qualité et de sécurité, la documentation des versions logicielles, les algorithmes de modélisation, les bases de données sélectionnées et les kits de formation. Ils devraient également conserver la documentation des fournisseurs de matériaux, de cellules et de réactifs si ces informations sont pertinentes pour l'évaluation d'une NAM (voir discussion détaillée à la section 3.3.3). Pour plus de détails sur les pratiques en matière de qualité, les procédures relatives aux équipements et la documentation à conserver, voir le Document d'orientation de l'OCDE sur les bonnes pratiques en matière de méthodes in vitro (GIVIMP ; OCDE, 2018). Un rapport sur la qualité d'installation/qualité de fonctionnement/qualité de performance (IQ/OQ/PQ) pour chaque instrument utilisé est recommandé ; il est nécessaire si les études sont censées être conformes aux Bonnes pratiques de laboratoire (BPL).

3.3.2.1 Informations concernant les méthodes cellulaires/tissulaires

La documentation sur l'origine des cellules et des tissus utilisés dans toute méthode d'essai doit être conservée. Les informations conservées pour examen devraient idéalement inclure entre autres :

- Considérations éthiques, juridiques et de sécurité
- Espèce/souche/sexe
- Données démographiques (le cas échéant)
- Source/fournisseur
- Nombre de donneurs
- Organe/tissu d'origine
- Type(s) de cellule(s) isolé(s)
- Séquences de cibles moléculaires modifiées, le cas échéant
- Technique et date d'isolement
- Classification de biosécurité
- Identification et authentification des lignées cellulaires
- Tests sérologiques pour les agents infectieux (p. ex. dépistage des mycoplasmes)
- Nombre de sous-cultures cellulaires/doublement de la population
- Temps de doublement
- Stabilité génétique
- Informations sur l'expression génique et des protéines (le cas échéant)
- Prétraitement

Pour des informations plus détaillées, voir la publication intitulée Guidance document on Good Cell and Tissue Culture Practice 2.0 (« Document d'orientation sur les bonnes pratiques de culture cellulaire et tissulaire 2.0 ») (Pamies et al., 2022) et le document GIVIMP de l'OCDE (OCDE, 2018).

Une documentation devrait être fournie concernant la mise en quarantaine des nouvelles cellules et des nouveaux tissus dans des conditions de conservation appropriées en laboratoire avant leur utilisation pour les essais. Les essais utilisant des cellules ou des tissus vivants devraient inclure un essai de viabilité cellulaire.

Tous les laboratoires participants devraient tenir des dossiers et conserver une documentation couvrant la performance de tous les équipements de laboratoire (lecteurs de plaques, incubateurs, réfrigérateurs/congélateurs, etc.).

L'enregistrement des cellules dans des banques de cellules peut contribuer à la traçabilité et à la disponibilité à long terme. Les banques de cellules devraient inclure une documentation de la densité cellulaire et du nombre de sous-cultures cellulaires. Des dossiers sur la préparation des réactifs devraient être tenus (p. ex. en consignant l'utilisation de tout réactif ou consommable dans des feuilles de relevé). Cela s'applique également aux méthodes *in vitro* n'utilisant pas de cellules/tissus. En voici quelques exemples :

- Fournisseur
- Numéro de catalogue
- Numéros de lots
- Dates de préparation
- Dates de péremption

- Nom de l'analyste

Le promoteur ou l'élaborateur de l'étude devrait fournir toutes les informations de sécurité relatives à la méthode ou à la méthodologie en cours de développement et préciser tous les règlements relatifs à l'utilisation, au transport et à l'élimination de toutes les matières dangereuses.

3.3.2.2 Évaluation de la méthode analytique utilisée dans la NAM

L'élaborateur devrait fournir des données démontrant clairement que la substance détectée ou quantifiée est la substance chimique ou l'analyte recherché. Le Tableau 6 répertorie et décrit les meilleures pratiques de contrôle qualité relatives aux méthodes analytiques.

Tableau 6. Évaluation des méthodes analytiques*

Évaluation de la méthode analytique	**Descriptif**	**Avantage/utilité**
Limites de détection et de quantification	Quantité ou concentration la plus faible d'un analyte qui peut être détectée ou quantifiée de manière fiable au-dessus du blanc de réactif ou dans la courbe standard. La concentration la plus élevée sur la courbe standard correspond à la limite supérieure de quantification.	Détermination de l'intervalle d'une méthode analytique.
Identification des interférences	Mise en évidence des cas où les composants de la méthode faussent le signal détecté.	Identification d'interactions entre les composants de la méthode et prévention de la communication de résultats artéfactuels.

Évaluation de la méthode analytique	**Descriptif**	**Avantage/utilité**
Évaluation de la précision analytique	Caractérisation de la précision de la méthode analytique utilisée et de tout autre test de précision, tel que ceux évaluant la variabilité de la performance lorsque la méthode proposée est employée par différents membres du personnel ou réalisée avec différents instruments. Des études de comparaison interlaboratoires peuvent être utilisées pour ce faire.	Renforcement de la confiance dans la fiabilité de la méthode analytique et évaluation des sources de variabilité en laboratoire.
Stabilité des matériaux utilisés dans les NAM	Capacité des matériaux utilisés dans la NAM (p. ex. substances à tester, appareil d'essai, système d'exposition, réactifs et analytes) à produire des résultats similaires et acceptables sur une période donnée dans un environnement donné.	Caractérisation des matériaux de la méthode et obtention systématique de données fiables pour une méthode particulière.
Test de robustesse	Vérification de la capacité d'une méthode à être reproduite dans conditions ou circonstances variables sans que les résultats obtenus montrent des différences inattendues.	Détermination de l'intervalle de paramètres dans laquelle l'essai fonctionne de manière acceptable.
Analyse du taux de recouvrement	Évaluation d'une méthode d'extraction en comparant les résultats des échantillons extraits avec ceux d'échantillons enrichis d'une matrice similaire et/ou de blancs enrichis.	Vérification de l'efficacité et la reproductibilité d'une méthode d'extraction.

Évaluation de la méthode analytique	**Descriptif**	**Avantage/utilité**
Analyse technique du domaine d'application	Fourniture de données adéquates sur les méthodes d'essai pour des substances chimiques et/ou produits représentatifs de ceux correspondants au CU pour lequel le test est proposé et description claire des propriétés physicochimiques du domaine d'application.	Réduction de l'incertitude concernant la performance de l'essai lors d'utilisation avec des substances chimiques et/ou produits différents et fourniture de méthodes et critères permettant de déterminer quand une substance chimique entre dans le domaine d'application.
Témoin positif	Identification du ou des composés témoins positifs pertinents pour le critère d'évaluation, dans la fenêtre de détection de l'essai.	Base cohérente et fiable pour la comparaison des résultats des substances à tester.
Normes de référence pour l'étalonnage des instruments	Étalonnage des instruments à l'aide de normes d'étalonnage et/ou d'échantillons de contrôle qualité.	Garantie de la fiabilité des mesures et identification des sources potentielles d'incertitude.
Établissement des spécifications	Définition de spécifications pour les mesures de contrôle en cours de processus sur la base des résultats de tests intralaboratoires et/ou interlaboratoires en utilisant une approche statistique.	Garantie de critères suffisamment stricts pour les mesures de contrôle, afin d'obtenir des résultats fiables pour les substances à tester et d'éviter tout biais.

***** Des détails et des exemples sont fournis dans l'Annexe C.

3.3.2.3 Évaluation du taux de reconnaissance global et/ou de la concordance de la NAM avec les normes de performance

Le taux de reconnaissance global et la concordance sont deux concepts très similaires souvent considérés comme interchangeables, en fonction du contexte d'un document ou d'un passage donné. En général, la concordance se rapporte à la comparaison de deux méthodes ou tests sur la base des résultats obtenus et le taux de reconnaissance global à la comparaison d'une méthode ou d'un test à une méthode de référence ou à un résultat de test. Une évaluation du taux de

reconnaissance global et/ou de la concordance est souvent utilisée dans l'évaluation statistique des méthodes d'essai et des données associées. Dans ce contexte, le taux de reconnaissance global est défini comme la proportion de prédictions correctes par rapport au nombre total de résultats. Les autres paramètres statistiques employés lors de la détermination du taux de reconnaissance global ou de la concordance des méthodes comprennent la sensibilité, la spécificité, la prédictivité positive et négative et les taux de faux positifs et de faux négatifs.

Dans la mesure du possible, des normes de performance bien définies (p. ex. ensemble équilibré de substances de référence connues pour donner des résultats positifs et négatifs) peuvent être utilisées pour vérifier la validité de la réponse et de la méthode. Les mesures quantitatives de concordance (c.-à-d. sensibilité, spécificité, prédictivité positive et négative, taux de faux positifs et négatifs) devraient être communiquées. Lors de la comparaison d'une méthode d'essai proposée à une méthode ayant des normes de performance établies (p. ex. méthodes me-too pour les lignes directrices pour les essais [TG] de l'OCDE) qui produit des données fonctionnellement et mécaniquement similaires, la concordance (y compris les données discordantes) entre les deux méthodes et par rapport à une méthode de référence devrait être évaluée.

3.3.2.4 Procédures opérationnelles standard et détails des méthodes

Il est recommandé que la méthode d'essai proposée soit assortie de procédures opérationnelles standard (SOP) bien documentées afin de garantir une performance constante de la méthode et des activités de laboratoire connexes, telles que la manipulation du système d'essai et l'étalonnage de l'équipement. Les SOP devraient couvrir tous les aspects des tests et des analyses, et inclure :

- Systèmes de comptabilisation qui garantissent l'intégrité des articles de test (p. ex. tenue des dossiers, sécurité et conservation de la chaîne d'échantillons).
- Outils de préparation et d'analyse des échantillons, tels que méthodes, réactifs (en précisant, le cas échéant, le fabricant, le numéro de catalogue, le numéro de lot, etc.), équipement et instrumentation.
- Procédures de contrôle qualité et de vérification des résultats.
- Détails de la méthode, y compris description et formulation complètes du produit, système d'exposition, volume/poids/solubilité/protocole de dosage de la substance à tester, et intervalle d'exposition/de dose approprié.

Le demandeur devrait également fournir une liste des caractéristiques de fonctionnement et des critères opérationnels permettant d'évaluer la performance et les résultats des essais. Les informations opérationnelles et les critères relatifs aux stèmes techniques qui constituent la NAM peuvent varier, mais les critères peuvent inclure des graphiques de contrôle qualité ou d'autres normes de performance pour l'ensemble des témoins, des normes, des expositions (dose et durée) et des groupes expérimentaux. Une description des méthodes statistiques utilisées pour évaluer les données devrait être incluse. En outre, les élaborateurs devraient décrire la manière dont l'incertitude expérimentale, l'incertitude statistique, l'interférence et le contexte ont été évalués.

3.3.3 Documentation

Cette section décrit les meilleures pratiques de documentation des NAM (y compris les AD et les IATA). Les facteurs énumérés ici peuvent ou non s'appliquer à toutes les NAM. En outre, une agence peut avoir besoin de documents supplémentaires pour évaluer la NAM. Les élaborateurs et/ou promoteurs sont invités à communiquer directement avec l'agence fédérale à laquelle ils ont l'intention de soumettre des méthodes ou des données afin de déterminer les facteurs supplémentaires éventuellement nécessaires à l'examen. Pour certaines agences, un examen indépendant par les pairs de la NAM et des données associées peut être nécessaire avant l'examen par l'agence.

Pour les nouvelles méthodes, la documentation devrait inclure une description de la méthode d'essai proposée et de la manière dont elle pourrait être exploitée à des fins réglementaires ou s'intégrer dans un CU particulier. Les informations pertinentes comprendraient entre autres les éventuelles données mécanistiques et la pertinence biologique de la méthode d'essai et de tout CU proposé (p. ex. concernant l'évaluation des risques). Les domaines d'application chez l'homme ou les taxons appropriés devraient être inclus dans la documentation. D'autres aspects importants de la documentation sont détaillés ci-dessous.

Des ressources supplémentaires fournissant des conseils de documentation sont disponibles. Pour les méthodes *in vitro*, les élaborateurs peuvent consulter le document GIVIMP de l'OCDE (OCDE, 2018).

3.3.3.1 Identité et pureté de la substance à tester

Pour toute substance testée dans la NAM (p. ex. témoin, composé de référence, autres substances à tester), au minimum, l'identité de la substance, dans l'idéal un identifiant unique (p. ex. CASRN, SMILES, InchiKey) et les informations sur la pureté de la substance communiquées par le fournisseur devraient être indiquées. Si le laboratoire possède les capacités nécessaires, la mise en œuvre de mesures de contrôle qualité supplémentaires à l'aide de méthodes analytiques pour évaluer la pureté et l'identité de la substance présente un avantage additionnel. L'aptitude à recueillir ces informations peut également dépendre du CU, car certaines substances à tester (p. ex. échantillons environnementaux représentant des mélanges complexes de composition inconnue ou variable) peuvent ne pas être bien caractérisées.

3.3.3.2 Élaboration de méthodes

Une description précise et détaillée de la méthode devrait être rédigée sur la base des données issues de cette méthode. Elle peut prendre la forme d'un protocole, d'un plan d'étude, d'un rapport et/ou d'une SOP. Chaque étape de la méthode devrait être étudiée afin de déterminer dans quelle mesure les variables environnementales, matricielles ou procédurales peuvent affecter la détection et/ou la quantification des analytes.

Au cours de l'élaboration, une attention particulière devrait notamment être accordée aux facteurs suivants :

- Sélection des réactifs (en précisant la pertinence biologique, la spécificité et la stabilité).
- Méthode ou instrumentation de détection (c.-à-d. procédures de détermination de la performance et de l'étalonnage facilement disponibles).
- Compatibilité des produits jetables (c.-à-d. plaques de microtitration et autres plastiques) avec les mesures de l'essai et les substances à tester.
- Méthode d'analyse/méthode statistique.
- Étapes ou procédés susceptibles d'introduire une variabilité dans l'essai.

Des mesures appropriées devraient être prises pour minimiser les effets externes ou matriciels (ou tout au moins caractériser ces effets) tout au long de l'application de la méthode, en particulier si les composés, la matrice ou l'équipement utilisés pendant l'élaboration sont différents de ceux employés pour la caractérisation technique de la méthode.

3.3.3.3 Mesures des critères d'évaluation et des paramètres

Les mesures de chaque critère d'évaluation ou analyte devraient avoir été dûment testées et documentées. L'élaboration d'une méthode pour une nouvelle NAM devrait inclure la démonstration que la méthode peut mesurer efficacement tous les paramètres pertinents. Les métadonnées propres à chaque essai devraient être saisies et enregistrées afin de relier les données quantitatives à des informations qualitatives, telles que les conditions expérimentales ou la substance à tester et la concentration, et pour assurer le suivi des facteurs externes (tels que la date ou le technicien effectuant l'expérience) qui peuvent contribuer à la variation de l'essai ou aux effets de lot. Ces données et métadonnées devraient être exportées et sauvegardées dans un format accessible afin de pouvoir être consultées lors d'examens indépendants de la validation. Ce sujet est également abordé dans les sections de l'Annexe B sur les cartes de contrôle et les feuilles de relevé.

3.3.3.4 Limites d'utilisation

Les points forts et les limitations spécifiques à la méthode d'essai devraient être clairement identifiés et décrits. Toutes les sources potentielles d'interférences devraient être répertoriées et toutes les substances chimiques ou classes de substances chimiques susceptibles d'interférer avec le test être identifiées. La documentation devrait également identifier toutes les limitations connues quant aux matériaux qui peuvent être testés à l'aide de la NAM.

3.3.3.5 Critère d'évaluation clairement défini

Les données générées par la méthode d'essai devraient mesurer ou prédire de manière adéquate le critère d'évaluation choisi, qui doit être clairement défini et assorti d'une explication de la pertinence biologique, comme décrit dans la section 3.2. Il pourrait s'agir, à titre d'exemple, d'une NAM fournissant des informations sur un événement clé précis d'une AOP. Les données devraient également décrire tout lien entre la nouvelle méthode d'essai et une méthode d'essai existante ou entre la nouvelle méthode d'essai et les effets sur l'espèce cible. Les critères d'un

résultat positif, négatif ou non concluant de la NAM devraient être clairement définis et évalués au fil du temps afin de garantir la stabilité du système.

3.3.3.6 Création d'un modèle statistique

Des modèles statistiques peuvent être élaborés à partir des données des tests intralaboratoires. Ces modèles peuvent être basés sur des approches statistiques bayésiennes ou fréquentistes. Des histogrammes peuvent être utilisés pour évaluer la distribution des données obtenues pour les mesures de contrôle en cours de processus et pour établir le type de distribution (p. ex. normale) des données (voir Petersen et al., 2022b pour plus de détails). Il est utile de développer des modèles pour calculer la variabilité cumulée de la NAM à partir des mesures tant des substances à tester que des mesures de contrôle en cours de processus, au lieu de se contenter des données sur la variabilité des substances à tester. Ces informations peuvent servir à élaborer un modèle statistique susceptible de mener à une décision (p. ex. au sujet de la positivité ou de la négativité de la substance à tester) et d'en connaître le niveau de confiance statistique. Une simple comparaison de la valeur moyenne d'une évaluation de la substance à tester à un seuil ne tient pas compte de la variabilité des résultats de test et ne peut pas établir le niveau de confiance statistique pour la décision.

L'une des principales préoccupations lors de l'élaboration de modèles statistiques pour les NAM tient à la distinction des résultats « négatifs » et « faiblement positifs ». Le seuil du modèle statistique peut être guidé par les données *in vivo* éventuellement disponibles (Friedman et al., 2023 ; Karmaus et al., 2022 ; Pham et al., 2020). Des tests répétés des composés « limites » peuvent être nécessaires pour évaluer la reproductibilité de la NAM ; voir par exemple la ligne directrice n° 497 de l'OCDE (OCDE, 2021a). Un modèle statistique évaluant une relation dose-réponse peut ainsi être utilisé pour évaluer le point de départ ou la concentration qui provoque un effet défini (p. ex. valeur de CE_{50}) et les intervalles de confiance associés basés sur les données autour de ces valeurs. Il peut également servir à évaluer la qualité d'une NAM à l'aide d'approches statistiques telles que le test t de Student, le facteur Z ou d'autres critères statistiques appropriés (Zhang et al., 1999 ; Zhang, 2011).

3.3.3.7 Reproductibilité des résultats d'essai

Le cas échéant, la documentation relative à la reproductibilité technique devrait être jointe aux informations soumises. La reproductibilité de la méthode peut être évaluée par des mesures répétées, y compris pour les contrôles qualité et les échantillons. Cette évaluation devrait inclure une justification de la sélection des substances utilisées pour évaluer la reproductibilité (éventuellement pour toute étude intra- et interlaboratoires menée) et indiquer dans quelle mesure elles représentent la gamme des résultats d'essai possibles. Les valeurs aberrantes devraient être identifiées et discutées. Une analyse statistique quantitative de l'étendue de toute variabilité intra- et/ou interlaboratoires ou une analyse du coefficient de variation devrait être incluse. Les mesures de la tendance centrale et de la variation devraient être résumées pour les données de contrôle historiques (témoin négatif, témoin positif et excipient témoin, le cas échéant). Lorsqu'un ou plusieurs composés sont testés plusieurs fois, les comparaisons peuvent être quantitatives (p. ex. valeurs de CE_{50} obtenues) ou qualitatives (p. ex. classification des

dangers). Dans les cas où la méthode d'essai proposée est semblable sur les plans mécanistique et fonctionnel à une méthode d'essai établie associée à des normes de performance existantes (p. ex. issues d'une TG de l'OCDE), la fiabilité des deux méthodes devrait être comparée et l'impact potentiel de toute différence discuté.

3.3.3.8 Procédure d'interprétation des données

La procédure d'interprétation des données, y compris les critères de réponse positive ou négative, devrait être clairement décrite pour chaque NAM. La combinaison des NAM dans des AD nécessite des procédures d'interprétation des données établies qui sont objectives et ne font pas appel à l'avis d'experts, afin de garantir qu'elles aboutiront au même résultat en cas d'application par des groupes différents (OCDE, 2017). L'utilisation d'algorithmes informatiques, tels que des modèles d'apprentissage automatique, et de logiciels (en précisant le numéro de version) devrait être correctement documentée pour garantir la reproductibilité des conclusions.

3.4 Intégrité des données

L'intégrité des données est essentielle pour garantir que les informations dérivées des NAM sont fiables et dignes de confiance. Il est conseillé aux élaborateurs de méthodes de procéder à une évaluation interne des processus utilisés pour l'acquisition, le transfert et le traitement des données brutes avant que celles-ci soient soumises à des tiers externes indépendants à des fins d'évaluation et d'examen par les pairs, afin de garantir l'intégrité des données et la crédibilité des résultats. Dans la mesure du possible, les études doivent si nécessaire être menées conformément aux principes des BPL (21 CFR § 58 ; 40 CFR § 160 ; 40 CFR § 792 ; OCDE, 1998). En outre, des organismes d'évaluation, tels que le National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), peuvent faciliter les évaluations de la qualité et de l'intégrité du processus d'élaboration de la NAM (NIEHS, 2023b). D'autres ressources fournissant des conseils sur la manière de maximiser l'intégrité des données sont disponibles. Pour les méthodes *in vitro*, les élaborateurs peuvent consulter le document GIVIMP de l'OCDE (OCDE, 2018). Pour les outils numériques et la gestion des données numériques, les élaborateurs peuvent suivre le document FAIR Guiding Principles for scientific data management and stewardship (« Principes directeurs FAIR pour la gestion et la gestion des données scientifiques »), publié en 2016 (Wilkinson et al., 2016).

3.5 Transparence de l'information

La transparence favorise la confiance dans l'utilisation des NAM et accélère ainsi le rythme du processus décisionnel réglementaire d'une agence et l'acceptation ou la qualification réglementaire potentielle. La pertinence d'une NAM pour l'espèce, le CU et la caractérisation technique devrait être communiquée de manière transparente aux pairs évaluateurs, à la communauté scientifique et au public. Le cas échéant, les articles évalués par les pairs et les informations décrivant le CU, la pertinence biologique et la caractérisation technique de la NAM devraient être publiés dans des revues en libre accès et/ou résumés dans des documents réglementaires destinés au public. Idéalement, les principes de la NAM, le protocole, les fichiers

de données brutes et les scripts utilisés pour analyser et représenter graphiquement les données ainsi que les normes de déclaration devraient être communiqués publiquement. Pour les NAM incluant une propriété intellectuelle, l'OCDE fournit des outils pour préserver la transparence, notamment des conditions raisonnables et non discriminatoires pour les engagements en matière de licences (OCDE, 2021b). Des techniques ou des équipements propriétaires ou brevetés peuvent éventuellement être utilisés dans une méthode pour répondre aux besoins de tests réglementaires. Pour certaines agences, l'élaborateur de la NAM peut être amené à transmettre des informations propriétaires ou brevetées étayant l'acceptation ou la qualification réglementaire ; des directives spécifiques à une agence peuvent informer les élaborateurs de méthodes d'essai sur ses exigences en matière d'information.

Les partenaires de la Coopération internationale relative aux méthodes de substitution à l'expérimentation animale (ICATM ; NIEHS, 2023c) publient des informations sur l'évaluation et l'examen par les pairs des NAM via le Système de suivi de méthodes d'essais de substitution jusqu'à leur acceptation dans la réglementation (TSAR) (EURL ECVAM, 2021). Le TSAR indique les étapes franchies par les NAM dans l'optique de leur acceptation en tant que normes reconnues à utiliser dans un contexte réglementaire et fournit une description sommaire et le ou les protocoles ou SOP acceptés. Le cas échéant, le TSAR inclut également des dossiers et des documents pertinents associés à une NAM se rapportant aux différentes étapes de l'ensemble du processus : soumission, validation, examen par les pairs, recommandations et acceptation ou qualification réglementaire. La manière d'interpréter les données générées par une NAM et les critères d'acceptation associés devraient être clairement communiqués afin que les utilisateurs finaux comprennent le processus et puissent le mettre en pratique.

3.6 Examen indépendant

Les informations et données étayant le CU, la pertinence biologique et la caractérisation technique de la NAM peuvent être examinées scientifiquement par des tiers indépendants (dont les membres sont exempts de conflits d'intérêts) ; toutefois, le niveau d'examen nécessaire dépendra des règlements et politiques de chaque agence.

L'évaluation d'une NAM peut être réalisée via plusieurs processus différents et implique la documentation des caractéristiques de performance d'une méthode afin de déterminer si elle est adaptée et fiable pour la ou les applications prévues. Souvent, la fiabilité d'une méthode inclut notamment la reproductibilité, la répétabilité et la robustesse. Cependant, des informations supplémentaires ne figurant pas dans ce rapport peuvent éventuellement être nécessaires pour l'examen de certaines méthodes. Outre la performance et l'applicabilité de la NAM, de bonnes pratiques scientifiques, techniques et de qualité garantissent que le processus d'évaluation indépendant est efficient et efficace et accroît la confiance dans la méthode proposée. Les laboratoires devraient conserver toutes les données requises pour l'évaluation d'une NAM, telles que les informations nécessaires au fonctionnement et à la maintenance de l'équipement utilisé pour la réalisation de la NAM (p. ex. manuels de l'équipement et des logiciels, et certificats de conformité de qualité et de sécurité) ainsi que la documentation des fournisseurs relative aux matériaux, aux cellules et aux réactifs. Pour les études censées être conformes aux BPL, un rapport IQ/OQ/PQ peut également être demandé pour chaque instrument utilisé. Pour les études

hors BPL, la documentation attestant la qualité de l'installation et des tests devrait être conservée afin de démontrer que l'équipement fonctionne comme prévu.

Les données brutes tirées de la NAM et les informations décrivant cette dernière devraient être accessibles pour pouvoir être examinées par des tiers indépendants et/ou les décideurs des organismes de réglementation. L'évaluation et l'examen indépendant par les pairs des NAM peuvent être organisés par des organismes de validation, tels que le NICEATM, le Laboratoire de référence de l'Union européenne pour la promotion des méthodes de substitution à l'expérimentation animale (EURL ECVAM) et son comité consultatif scientifique, et le Centre japonais pour la validation des méthodes de substitution (JaCVAM). Les autres organismes ou organisations internationales pouvant examiner les NAM de manière indépendante comprennent le U.S. Federal Insecticide, Fungicide, and Rodenticide Act Scientific Advisory Panel (Groupe consultatif scientifique de la loi fédérale américaine sur les insecticides, les fongicides et les rodenticides), le Comité scientifique de l'Autorité européenne de sécurité des aliments et l'OCDE. L'élaborateur peut également financer un examen indépendant de la méthode (sans en diriger directement la gestion). Les publications évaluées par les pairs sont utiles pour partager des informations sur les essais avec la communauté scientifique et peuvent compléter un examen plus formel effectué par des tiers indépendants pour étayer l'acceptation de la méthode et son utilisation dans un contexte réglementaire.

L'étendue de l'examen indépendant dépendra du CU, du cadre réglementaire et de la méthode objet de l'évaluation. Les informations communiquées par l'élaborateur à des fins d'examen indépendant incluraient notamment les dossiers relatifs aux études intra- ou interlaboratoires, précisant entre autres si la NAM a été comparée à une autre NAM ou à des données humaines ou animales *in vivo*. Ces évaluations pourraient contribuer à la transférabilité interlaboratoires de la NAM.

4.0 Acceptation des NAM par des agences fédérales des États-Unis

4.1 Comprendre les besoins réglementaires et les contextes décisionnels

Les agences fédérales disposent de différentes autorités pour demander, obtenir et utiliser les données toxicologiques. Ces autorités varient selon les lois, les réglementations et les catégories de produits. Une NAM peut être utile et adéquate dans certains contextes légaux ou réglementaires, mais pas dans d'autres. Par conséquent, les élaborateurs et les utilisateurs d'une NAM doivent tenir compte du contexte dans lequel elle sera utilisée. Ce contexte peut relever d'AD, d'IATA ou de méthodes autonomes, notamment.

L'acceptabilité d'une NAM à des fins réglementaires ou autres dépend de différents facteurs : nature de la décision à prendre ; adéquation de la NAM à son utilisation prévue ; et mesure dans laquelle la détermination de la sécurité, de l'efficacité et/ou des risques nécessaire à la demande réglementaire repose sur les résultats de la NAM. Une NAM présentant une sensibilité élevée et une spécificité faible qui identifie un signal préoccupant peut être utile pour les applications dans lesquelles un grand nombre de composés seront passés au crible afin de hiérarchiser les

évaluations ultérieures. L'utilité d'une telle NAM peut être limitée lorsqu'il s'agit de décider quelle dose d'un composé unique est sans danger pour l'exposition. Dans certains cas, les agences pourraient rechercher des signes préoccupants dans des situations de pénurie de données (identification des dangers). Dans d'autres, elles pourraient devoir prendre des décisions plus quantitatives concernant les composés (évaluation des risques) et ne pas être disposées à accepter un niveau élevé d'incertitude dans les résultats des NAM visant à guider ces décisions. La mise en œuvre des caractéristiques de performance devrait refléter la manière dont la NAM sera utilisée et être adaptée à chaque NAM et à son CU.

Pour faciliter ce processus, certaines agences souhaiteront sans doute évaluer les NAM avant leur utilisation dans le cadre d'un processus de prise de décisions réglementaires. L'évaluation se concentrera probablement sur l'existence ou non d'un CU bien défini pour la NAM. L'un des objectifs de l'évaluation est de permettre aux régulateurs d'appliquer les résultats générés par la NAM sans avoir à revoir toutes les données justificatives sous-jacentes à la NAM (voir les documents cités dans le Tableau 1).

4.2 Considérations relatives au contexte d'utilisation

L'objectif de la NAM devrait être clairement communiqué (p. ex. identification des dangers, évaluation de la puissance ou point de départ de l'évaluation quantitative des risques), et la NAM devrait être évaluée en fonction de cet objectif. Il convient de focaliser l'application d'une nouvelle NAM sur un seul CU. Toutefois, des CU supplémentaires peuvent être ajoutés ultérieurement en s'appuyant sur des données justificatives appropriées. Un CU doit généralement être axé sur un besoin réglementaire particulier. Les premiers efforts visant à qualifier une NAM peuvent être plus fructueux si le CU est restreint. Ce CU pourrait ensuite être étendu sur base de données supplémentaires. Les besoins réglementaires varient d'une agence à l'autre, de sorte que le CU d'une NAM donnée peut également différer. Il est essentiel d'établir un CU approprié pour une NAM proposée avant de lancer le processus de qualification complet.

La détermination d'un CU approprié devrait généralement être discutée entre l'élaborateur de la NAM et la ou les agences concernées. Plusieurs itérations d'un CU peuvent être nécessaires avant d'en définir une version acceptable. Un CU peut même changer au cours du recueil des données, à mesure que l'applicabilité et les limitations d'une NAM se précisent.

4.3 Évolution de la confiance en fonction de l'expérience acquise

Pour qu'une NAM puisse être utilisée à des fins réglementaires, les régulateurs et le secteur réglementé doivent lui accorder une confiance suffisante. La validation et la qualification renforcent cette confiance, mais peuvent ne pas suffire à garantir sa mise en œuvre. Une formation et une expérience pertinentes sont généralement nécessaires avant que l'utilisation d'une NAM puisse être largement acceptée pour l'objectif fixé. Le coût, la complexité et la disponibilité de réactifs et de personnel formé à la conduite et à l'interprétation d'une NAM sont quelques-uns des facteurs pouvant limiter l'adoption d'une NAM, même en présence de données de validation et de qualification.

La performance d'une NAM destinée à compléter ou à remplacer une approche existante devra généralement être comparée à celle de ladite approche. Les approches existantes bénéficiant d'une expérience d'utilisation substantielle suscitent souvent une grande confiance. Même lorsqu'elles n'ont pas fait l'objet d'une validation officielle, l'utilisation réussie et répétée de ces approches ainsi que la validité intrinsèque présumée des tests sur les animaux inspirent généralement une forte confiance. Les utilisateurs exploitant les résultats d'une NAM pour prendre des décisions en matière de sécurité doivent avoir la certitude que cette méthode sera aussi bonne, voire meilleure que les approches existantes. Des degrés d'incertitude variables peuvent être acceptables pour différents CU, le niveau de confiance devant croître à mesure que l'on passe de la hiérarchisation et de la sélection à la caractérisation des dangers et à l'évaluation des risques, par exemple.

L'un des mécanismes de renforcement de la confiance dans une NAM consiste pour les utilisateurs à fournir des données à son sujet en parallèle avec les données de la méthode existante que la NAM est censée remplacer ou compléter. Le partage volontaire d'informations sur les NAM au sein et entre les industries peut aider à établir un ensemble de données suffisamment important pour bâtir la confiance dans les méthodes. Avec le temps, les agences et l'industrie pourraient comprendre comment la NAM peut s'intégrer aux paradigmes d'évaluation existants sans compromettre les normes de sécurité.

5.0 Harmonisation aux États-Unis et à l'échelle internationale

La coordination entre les agences fédérales américaines et, plus généralement, avec les organismes de réglementation internationaux contribuera à assurer l'harmonisation des approches visant à valider les NAM et à soutenir leur application et leur mise en œuvre. L'ICCVAM et le NICEATM jouent un rôle important de promotion de la communication et de la collaboration, au niveau tant national que mondial.

5.1 Harmonisation aux États-Unis : rôle de l'ICCVAM et du NICEATM

La loi d'autorisation de l'ICCVAM définit les objectifs suivants du comité (42 U.S.C 285l-3, 2000 ; NIEHS, 2023d) :

- Accroître l'efficience et l'efficacité de l'examen des méthodes d'essai par les agences fédérales américaines.
- Éliminer les efforts redondants et partager l'expérience entre les organismes fédéraux de réglementation des États-Unis.
    - Plusieurs moyens permettent d'y parvenir, notamment des réunions mensuelles, des groupes de travail, des réunions publiques ouvertes (forum public) et des groupes consultatifs scientifiques (SACATM).
- Optimiser l'utilisation de l'expertise scientifique en dehors du gouvernement fédéral américain.

    - Cela passe souvent par l'organisation de réunions et d'ateliers avec les partenaires de l'ICATM ainsi qu'avec de nombreux scientifiques lors de conférences.
- S'assurer que les méthodes d'essai nouvelles et révisées sont validées pour répondre aux besoins des agences fédérales américaines.
    - Les régulateurs et les scientifiques des agences fédérales coopèrent avec les élaborateurs pour s'assurer que les méthodes produites répondront à un besoin réglementaire.
- Réduire, réajuster ou remplacer l'utilisation d'animaux dans les essais dans la mesure du possible.

L'ICCVAM facilite des collaborations interagences et internationales promouvant le développement, l'acceptation ou la qualification réglementaire ainsi que l'utilisation de tests alternatifs encouragant la réduction, le réajustement ou le remplacement des méthodes d'expérimentation animale. L'ICCVAM fournit des conseils aux élaborateurs de méthodes d'essai, évalue les recommandations issues d'examens par les pairs des méthodes d'essai toxicologiques alternatives et formule des recommandations sur l'utilisation des méthodes d'essai examinées aux agences fédérales voulues. L'ICCVAM remplit ses fonctions par le biais de groupes de travail techniques ad hoc, gérés par le NICEATM et chargés d'effectuer des tâches précises importantes pour le développement ou la validation de méthodes de substitution aux tests sur les animaux. Actuellement, l'ICCVAM soutient par exemple la coordination d'une étude de prévalidation interlaboratoires d'une NAM développée par l'EPA sur la base d'un essai *in vitro* sur microtissus thyroïdiens humains pour le criblage chimique (Deisenroth et al., 2020).

Le NICEATM, un bureau de la division de toxicologie translationnelle du National Institute of Environmental Health Sciences (NIEHS), fournit un soutien technique, scientifique et opérationnel aux activités des groupes de travail de l'ICCVAM et de l'ICCVAM, aux comités d'examen par les pairs, aux groupes d'experts, aux ateliers et aux efforts de validation. Outre son soutien à l'ICCVAM, le NICEATM :

- effectue des analyses et évaluations des méthodes d'essai et coordonne des études de validation indépendantes sur des approches de test alternatives nouvelles et hautement prioritaires ;
- fournit des informations aux élaborateurs de méthodes d'essai, aux régulateurs et à l'industrie réglementée via son site Internet, l'Integrated Chemical Environment et des ateliers sur des sujets pertinents ;
- soutient les activités de la division de toxicologie translationnelle du NIEHS, en particulier celles qui contribuent au consortium interagences Toxicology in the 21st Century (Tox21) du gouvernement américain.

Le forum de communication interagences fourni par l'ICCVAM et le soutien apporté par le NICEATM garantissent que les ressources limitées sont exploitées efficacement pour coordonner les efforts des agences fédérales américaines visant à valider et à qualifier les NAM pour une application réglementaire.

5.2 Harmonisation aux États-Unis : autres collaborations fédérales promouvant les principes des 3R

En complément de leur participation à l'ICCVAM, les agences fédérales américaines collaborent sur les NAM de nombreuses manières. Par exemple, Tox21 (Tox21, n.d.) est une collaboration fédérale entre l'EPA, la division de toxicologie translationnelle du NIEHS, le National Center for Advancing Translational Sciences (NCATS) des National Institutes of Health et la FDA. L'objectif de Tox21 est de développer de meilleures méthodes d'évaluation de la toxicité afin de tester efficacement si certains composés chimiques sont susceptibles de perturber les processus biologiques du corps humain et d'avoir des effets négatifs sur la santé. Tox21 a produit un certain nombre de publications et d'analyses fondamentales qui ont été mises en œuvre à des fins réglementaires, par exemple dans le cadre de l'Endocrine Disruptor Screening Program (Programme de dépistage des perturbateurs endocriniens) de l'EPA (EPA, 2023). Le NICEATM et l'EPA ont également travaillé en étroite collaboration à plusieurs études visant à réduire ou à remplacer l'utilisation d'animaux dans les tests réglementaires. Ces études comprennent des analyses rétrospectives visant à éliminer l'utilisation d'animaux pour : 1) la toxicité aiguë par voie cutanée des pesticides et des formulations de pesticides (EPA, 2020) ; 2) déterminer si l'utilisation de données *in vitro* suffit pour calculer le facteur d'absorption dermique pour l'évaluation des risques pour la santé humaine des pesticides (Allen et al., 2021) ; et 3) déterminer si le même niveau de protection des vertébrés aquatiques non cibles peut être atteint en menant des essais de toxicité aiguë *in vivo* sur moins de trois espèces de poissons (Ceger et al., 2023). La FDA collabore également avec le NCATS à la poursuite du développement de technologies utilisant des systèmes microphysiologiques afin de promouvoir leur avancement et d'accélérer leur utilisation translationnelle (FDA, 2023).

5.3 Harmonisation internationale

Outre les efforts de collaboration internationaux impulsés par l'ICCVAM, les agences américaines collaborent indépendamment à l'échelle internationale à la promotion de l'acceptation des NAM. Les exemples incluent l'ICATM, la collaboration des États-Unis avec le sous-comité d'experts des Nations Unies sur le Système général harmonisé de classification et d'étiquetage des produits chimiques (SGH), la collaboration avec la CIH pour élaborer des directives décrivant l'utilisation des NAM et la participation au Test Guidelines Programme (Programme sur les lignes directrices pour les essais) et au Working Party on Hazard Assessment (Groupe de travail sur l'évaluation des dangers) de l'OCDE.

L'ICATM a été créé en tant que partenariat entre des organisations de validation des États-Unis (ICCVAM), du Japon (JaCVAM), de l'Union européenne (EURL ECVAM) et du Canada (Bureau de la science et de la recherche en santé environnementale de Santé Canada). Parmi les autres organisations participantes figurent le Centre coréen pour la validation des méthodes alternatives, le Centre brésilien pour la validation des méthodes alternatives, l'Agence nationale des produits médicaux de Chine et le Centre de contrôle et de prévention des maladies du Guangdong. Les principaux objectifs de ce groupe sont les suivants :

- Établir une coopération internationale dans les domaines critiques des études de validation, de l'examen indépendant par les pairs et de l'élaboration de recommandations harmonisées afin de favoriser l'acceptation des méthodes/stratégies alternatives dans le monde entier.
- Établir la coopération internationale nécessaire pour garantir que les nouvelles méthodes/stratégies d'essai alternatives adoptées à des fins réglementaires offrent une protection équivalente ou améliorée pour les personnes, les animaux et l'environnement, tout en remplaçant, réduisant ou réajustant (en atténuant la douleur et la détresse) l'utilisation des animaux chaque fois que cela est scientifiquement possible.

Au sein des Nations Unies, le sous-comité du SGH a créé un groupe de travail chargé de mettre à jour divers chapitres du SGH afin d'établir des critères spécifiques à l'utilisation des NAM pour différentes classes de danger (p. ex. irritation/corrosion de la peau, irritation oculaire/lésions oculaires graves, sensibilisation cutanée). Ces efforts ont permis de faire progresser les NAM au niveau international et ont fait des AD des méthodes acceptables de détermination des dangers. Ces efforts internationaux sont importants non seulement pour promouvoir les NAM, mais également pour fournir une expertise aux régions internationales qui peuvent ne pas disposer des ressources suffisantes dans ce domaine.

La valeur de l'harmonisation des approches d'utilisation des NAM est également illustrée par les activités de la CIH. Plusieurs lignes directrices élaborées par la CIH décrivent l'utilisation de méthodes alternatives considérées comme acceptables par une variété d'autorités de réglementation et de groupes industriels du monde entier. Ces méthodes ont été évaluées par des groupes de travail d'experts dans le cadre du processus de la CIH et, le cas échéant, intégrées aux lignes directrices. Les exemples incluent l'utilisation de méthodes *in chemico* et *in vitro* pour l'évaluation de la phototoxicité (ICH, 2013). En outre, les lignes directrices de la CIH sur la toxicité pour la reproduction et le développement incluent certains contextes d'utilisation pour des essais alternatifs ainsi que des recommandations sur l'approche de qualification de ces essais et une liste de composés de référence pour les critères d'évaluation biologiques (FDA, 2021b).

L'OCDE, un forum international visant à harmoniser les directives et les documents d'orientation relatifs aux essais réglementaires, se concentre de plus en plus sur la validation et l'utilisation des NAM. L'ICCVAM joue un rôle important dans la coordination et la contribution à la position des États-Unis vis-à-vis du Health Effects Test Guidelines Programme (Programme sur les lignes directrices pour les essais relatifs aux effets sur la santé) de l'OCDE. Le coordinateur national américain représente les États-Unis à la réunion annuelle du Groupe de travail des coordonnateurs nationaux et dans le cadre d'autres activités d'élaboration de lignes directrices pour les essais. À ce titre, le coordinateur national américain sollicite les contributions des agences pertinentes de l'ICCVAM pour les activités relevant des TG de l'OCDE qui concernent tous les aspects des principes des 3R. Des experts en la matière des agences de l'ICCVAM sont membres de plusieurs groupes d'experts de l'OCDE chargés de fournir des conseils scientifiques sur le développement de produits de l'OCDE, tels que des lignes directrices pour les essais, des AD et des documents d'orientation. Les TG de l'OCDE sont utilisées par les parties prenantes des 38 pays membres de l'OCDE pour évaluer la sécurité

chimique. La clause d'acceptation mutuelle des données de l'OCDE garantit que les données de sécurité générées à l'aide d'une TG de l'OCDE seront acceptées par tous les pays membres, évitant ainsi les tests redondants. Les agences de l'ICCVAM contribuent également au projet d'études de cas d'IATA de l'OCDE, qui permet aux pays de partager et de collaborer sur l'utilisation de nouvelles méthodes des IATA pour évaluer la sécurité chimique dans un contexte réglementaire.

6.0 Communication et formation visant à encourager l'utilisation des NAM

Les communications des agences concernant l'acceptabilité de certaines NAM et la formation à ces dernières peuvent faciliter leur utilisation. Le cas échéant et dans la mesure du possible, les agences pourraient déclarer publiquement quand et comment une NAM est acceptable, en fonction des règles et politiques qui leur sont propres. Par exemple, les NAM peuvent être décrites dans les documents d'orientation ou sur des sites Internet accessibles au public. Les organismes de réglementation peuvent utiliser les programmes de formation existants ou en mettre en œuvre de nouveaux pour le personnel et proposer une formation accessible au public, dans la mesure du possible, pour les nouvelles méthodes. La confiance dans les nouvelles approches peut être promue avant même qu'une NAM soit validée ou qualifiée grâce à l'éducation de la communauté scientifique. Ces efforts d'éducation précoce peuvent se concentrer sur les bases scientifiques des nouvelles approches. Au fur et à mesure que les approches mûrissent et que les données étayant leur validité s'accumulent, les efforts éducatifs peuvent être orientés vers la familiarisation de la communauté avec ces données. Une formation sur l'utilisation et l'interprétation spécifiques des NAM peut avoir lieu une fois que les NAM ont été évaluées pour des CU particuliers. Des entités telles que l'OCDE, l'ICCVAM et d'autres organisations de validation et sociétés scientifiques peuvent également fournir une formation et un accès à des informations sur les NAM afin de renforcer la confiance dans leur utilisation.

L'interaction entre les élaborateurs de NAM, les utilisateurs industriels et les régulateurs peut faciliter le développement et l'adoption des méthodes. De telles interactions peuvent prendre la forme de réunions scientifiques au cours desquelles ces méthodes sont discutées par toutes les parties prenantes et d'interactions réglementaires plus formelles conformes aux processus spécifiques aux agences. Comme indiqué précédemment, l'élaboration de CU et d'ensembles de données de qualification appropriés peut être un processus itératif. Les CU et les domaines d'application peuvent évoluer au cours du développement et de l'exploration d'une NAM, à mesure que les données s'accumulent. Une communication continue entre toutes les parties au cours de ce processus contribue à garantir qu'une voie appropriée vers l'acceptation d'une NAM sera empruntée.

Même si une NAM est disponible pour un critère d'évaluation et acceptée par un organisme de réglementation, le demandeur peut choisir d'utiliser d'autres approches, telles qu'un essai traditionnel sur des animaux. Il est donc possible que la NAM ne soit pas toujours intégrée à une demande réglementaire. Cela peut s'expliquer par plusieurs raisons échappant au contrôle des

autorités de réglementation. Bien qu'un organisme de réglementation puisse suggérer ou recommander de substituer une ou plusieurs NAM à un test traditionnel, le promoteur d'un composé n'est pas toujours obligé d'obtempérer. L'éducation et la familiarisation avec les NAM de toutes les parties prenantes sont nécessaires pour créer une confiance suffisante assurant l'adoption des NAM.

7.0 Conclusion et mise en œuvre

Ce rapport est destiné à aider les élaborateurs de méthodes, les parties prenantes de l'industrie réglementée et les agences fédérales à développer, valider, qualifier et accepter des NAM scientifiquement pertinentes. Nous avons décrit ici les concepts clés qui devraient être pris en compte pour permettre le développement efficace et rapide de NAM adaptées à un objectif, fiables et fournissant des informations pertinentes pour l'espèce considérée. Toutes les informations peuvent s'appliquer ou non à une méthode, une AD ou une IATA particulière. D'autres concepts qui s'appliquent à une méthode, à une AD ou à une IATA peuvent par ailleurs ne pas être abordés dans ce rapport. Il est important que les élaborateurs collaborent étroitement avec les agences fédérales et les utilisateurs finaux en tenant compte de l'utilisation prévue de la NAM, en particulier pour les applications d'évaluation des risques dans le cadre d'un examen réglementaire. L'établissement de la confiance scientifique dans les NAM et la validation ou la qualification des méthodes pour des objectifs et des CU précis devraient être des processus itératifs guidés par une communication multidirectionnelle entre les parties prenantes.

Le domaine des NAM est en pleine évolution, et de nouvelles considérations sur la validation et la qualification des NAM non anticipées au moment de la rédaction du présent rapport pourraient émerger. Par conséquent, les parties prenantes participant au développement, à la validation et à la qualification des NAM devront probablement faire preuve de flexibilité et être prêtes à intégrer des considérations non décrites ici. Ce rapport sera mis à jour régulièrement et selon les besoins.

Références

15 USC §2601, 2016. 15 USC §2601: Findings, policy, and intent [WWW Document]. URL https://uscode.house.gov/view.xhtml?req=(title:15%20section:2601%20edition:prelim) (accessed 11.2.22).

21 CFR § 58, n.d. CFR - Code of Federal Regulations Title 21 [WWW Document]. URL https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=58 (accessed 11.3.22).

40 CFR § 160, n.d. CFR - Code of Federal Regulations Title 40 [WWW Document]. URL https://www.ecfr.gov/current/title-40/chapter-I/subchapter-E/part-160 (accessed 5.9.23).

40 CFR § 792, n.d. CFR - Code of Federal Regulations Title 40 [WWW Document]. URL https://www.ecfr.gov/current/title-40/chapter-I/subchapter-R/part-792 (accessed 5.9.23).

42 U.S.C 285l-3, 2000. ICCVAM Authorization Act of 2000 [WWW Document]. 42 U.S.C. 285l-3 Public Law 106-545. URL https://www.congress.gov/bill/106th-congress/house-bill/4281/text (accessed 5.9.23).

Allen, D.G., Rooney, J., Kleinstreuer, N., Lowit, A., Perron, M., 2021. Retrospective analysis of dermal absorption triple pack data. ALTEX - Alternatives to animal experimentation 38, 463–476. https://doi.org/10.14573/altex.2101121

Browne, P., Delrue, N., Gourmelon, A., 2019. Regulatory use and acceptance of alternative methods for chemical hazard identification. Current Opinion in Toxicology 15, 18–25.

Browne, P., Kleinstreuer, N.C., Ceger, P., Deisenroth, C., Baker, N., Markey, K., Thomas, R.S., Judson, R.J., Casey, W., 2018. Development of a curated Hershberger database. Reproductive Toxicology 81, 259–271. https://doi.org/10.1016/j.reprotox.2018.08.016

Ceger, P., Allen, D., Blankinship, A., Choksi, N., Daniel, A., Eckel, W.P., Hamm, J., Harwood, D.E., Johnson, T., Kleinstreuer, N., Sprankle, C.S., Truax, J., Lowit, M., 2023. Evaluation of the fish acute toxicity test for pesticide registration. Regulatory Toxicology and Pharmacology 139, 105340. https://doi.org/10.1016/j.yrtph.2023.105340

Ceger, P., Garcia-Reyero Vinas, N., Allen, D., Arnold, E., Bloom, R., Brennan, J.C., Clarke, C., Eisenreich, K., Fay, K., Hamm, J., Henry, P.F.P., Horak, K., Hunter, W., Judkins, D., Klein, P., Kleinstreuer, N., Koehrn, K., LaLone, C.A., Laurenson, J.P., Leet, J.K., Lowit, A., Lynn, S.G., Norberg-King, T., Perkins, E.J., Petersen, E.J., Rattner, B.A., Sprankle, C.S., Steeger, T., Warren, J.E., Winfield, S., Odenkirchen, E., 2022. Current ecotoxicity testing needs among selected U.S. federal agencies. Regulatory Toxicology and Pharmacology 133, 105195. https://doi.org/10.1016/j.yrtph.2022.105195

Chang, X., Tan, Y.-M., Allen, D.G., Bell, S., Brown, P.C., Browning, L., Ceger, P., Gearhart, J., Hakkinen, P.J., Kabadi, S.V., Kleinstreuer, N.C., Lumen, A., Matheson, J., Paini, A.,

Pangburn, H.A., Petersen, E.J., Reinke, E.N., Ribeiro, A.J.S., Sipes, N., Sweeney, L.M., Wambaugh, J.F., Wange, R., Wetmore, B.A., Mumtaz, M., 2022. IVIVE: Facilitating the use of in vitro toxicity data in risk assessment and decision making. Toxics 10, 232. https://doi.org/10.3390/toxics10050232

Chiu, W.A., Wright, F.A., Rusyn, I., 2017. A tiered, Bayesian approach to estimating of population variability for regulatory decision-making. ALTEX 34, 377–388. https://doi.org/10.14573/altex.1608251

Choksi, N.Y., Truax, J., Layton, A., Matheson, J., Mattie, D., Varney, T., Tao, J., Yozzo, K., McDougal, A.J., Merrill, J., Lowther, D., Barroso, J., Linke, B., Casey, W., Allen, D., 2019. United States regulatory requirements for skin and eye irritation testing. Cutan Ocul Toxicol 38, 141–155. https://doi.org/10.1080/15569527.2018.1540494

Church, R.J., Gatti, D.M., Urban, T.J., Long, N., Yang, X., Shi, Q., Eaddy, J.S., Mosedale, M., Ballard, S., Churchill, G.A., Navarro, V., Watkins, P.B., Threadgill, D.W., Harrill, A.H., 2015. Sensitivity to hepatotoxicity due to epigallocatechin gallate is affected by genetic background in diversity outbred mice. Food Chem Toxicol 76, 19–26. https://doi.org/10.1016/j.fct.2014.11.008

Clippinger, A.J., Raabe, H.A., Allen, D.G., Choksi, N.Y., van der Zalm, A.J., Kleinstreuer, N.C., Barroso, J., Lowit, A.B., 2021. Human-relevant approaches to assess eye corrosion/irritation potential of agrochemical formulations. Cutan Ocul Toxicol 40, 145–167. https://doi.org/10.1080/15569527.2021.1910291

Corley, R.A., Kuprat, A.P., Suffield, S.R., Kabilan, S., Hinderliter, P.M., Yugulis, K., Ramanarayanan, T.S., 2021. New approach methodology for assessing inhalation risks of a contact respiratory cytotoxicant: computational fluid dynamics-based aerosol dosimetry modeling for cross-species and in vitro comparisons. Toxicological Sciences 182, 243–259. https://doi.org/10.1093/toxsci/kfab062

CPSC, 2020. Proposed Guidance for Industry and Test Method Developers: CPSC Staff Evaluation of Alternative Test Methods and Integrated Testing Approaches and Data Generated from Such Methods to Support FHSA Labeling Requirements [WWW Document]. URL https://www.regulations.gov/document/CPSC-2021-0006-0001 (accessed 5.4.22).

CPSC, 2012. Recommended Procedures Regarding the CPSC's Policy on Animal Testing [WWW Document]. URL https://www.cpsc.gov/Business--Manufacturing/Testing-Certification/Recommended-Procedures-Regarding-the-CPSCs-Policy-on-Animal-Testing (accessed 5.4.22).

Crofton, K.M., Mundy, W.R., 2021. External Scientific Report on the Interpretation of Data from the Developmental Neurotoxicity In Vitro Testing Assays for Use in Integrated Approaches for Testing and Assessment. EFSA Supporting Publications 18, 6924E. https://doi.org/10.2903/sp.efsa.2021.EN-6924

Daniel, A.B., Strickland, J., Allen, D., Casati, S., Zuang, V., Barroso, J., Whelan, M., Régimbald-Krnel, M.J., Kojima, H., Nishikawa, A., Park, H.-K., Lee, J.K., Kim, T.S., Delgado, I., Rios, L., Yang, Y., Wang, G., Kleinstreuer, N., 2018. International regulatory requirements for skin sensitization testing. Regul Toxicol Pharmacol 95, 52–65. https://doi.org/10.1016/j.yrtph.2018.03.003

Deisenroth, C., Soldatow, V.Y., Ford, J., Stewart, W., Brinkman, C., LeCluyse, E.L., MacMillan, D.K., Thomas, R.S., 2020. Development of an *In Vitro* Human Thyroid Microtissue Model for Chemical Screening. Toxicological Sciences 174, 63–78. https://doi.org/10.1093/toxsci/kfz238

Dumont, J., Euwart, D., Mei, B., Estes, S., Kshirsagar, R., 2016. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. Crit Rev Biotechnol 36, 1110–1122. https://doi.org/10.3109/07388551.2015.1084266

Elliott, J.T., Rösslein, M., Song, N.W., Toman, B., Ovaskainen, A.K.-, Maniratanachote, R., Salit, M.L., Petersen, E.J., Sequeira, F., Romsos, E., Kim, S.J., Lee, J., Moos, N.R. von, Rossi, F., Hirsch, C., Krug, H.F., Suchaoin, W., Wick, P., 2017. Toward achieving harmonization in a nanocytotoxicity assay measurement through an interlaboratory comparison study. ALTEX - Alternatives to animal experimentation 34, 201–218. https://doi.org/10.14573/altex.1605021

EPA, 2023. Availability of New Approach Methodologies (NAMs) in the Endocrine Disruptor Screening Program (EDSP) [WWW Document]. URL https://www.regulations.gov/document/EPA-HQ-OPP-2021-0756-0002 (accessed 2.20.23).

EPA, 2021a. EPA New Approach Methods Work Plan [WWW Document]. URL https://www.epa.gov/system/files/documents/2021-11/nams-work-plan_11_15_21_508-tagged.pdf (accessed 5.5.22).

EPA, 2021b. EPA Strategic Plan to Reduce the Use of Vertebrate Animals in Chemical Testing [WWW Document]. URL https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/strategic-plan-reduce-use-vertebrate-animals-chemical (accessed 5.5.22).

EPA, 2021c. Chlorothalonil: Revised Human Health Draft Risk Assessment for Registration Review [WWW Document]. URL https://www.regulations.gov/document/EPA-HQ-OPP-2011-0840-0080 (accessed 3.13.23).

EPA, 2020. Guidance for Waiving Acute Dermal Toxicity Tests for Pesticide Technical Chemicals & Supporting Retrospective Analysis [WWW Document]. URL https://www.regulations.gov/document/EPA-HQ-OPP-2016-0093-0181 (accessed 3.7.23).

EPA, 2018. Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program.

EPA, 2015. Use of an Alternate testing framework for classification of eye irritation [WWW Document]. URL https://www.epa.gov/sites/default/files/2015-05/documents/eye_policy2015update.pdf (accessed 11.13.23).

EURL ECVAM, 2021. EU Reference Laboratory for alternatives to animal testing (EURL ECVAM) [WWW Document]. URL https://joint-research-centre.ec.europa.eu/eu-reference-laboratory-alternatives-animal-testing-eurl-ecvam_en (accessed 11.3.22).

Farmahin, R., Manning, G.E., Crump, D., Wu, D., Mundy, L.J., Jones, S.P., Hahn, M.E., Karchner, S.I., Giesy, J.P., Bursian, S.J., Zwiernik, M.J., Fredricks, T.B., Kennedy, S.W., 2013. Amino Acid Sequence of the Ligand-Binding Domain of the Aryl Hydrocarbon Receptor 1 Predicts Sensitivity of Wild Birds to Effects of Dioxin-Like Compounds. Toxicological Sciences 131, 139–152. https://doi.org/10.1093/toxsci/kfs259

FDA, 2023. MOU 225-23-003 Memorandum of understanding between the National Institutes of Health (NCATS) and the Food and Drug Administration (FDA) for the Microphysiological Systems Program [WWW Document]. FDA. URL https://www.fda.gov/about-fda/domestic-mous/mou-225-23-003 (accessed 5.9.23).

FDA, 2021a. Advancing New Alternative Methodologies at FDA [WWW Document]. URL https://www.fda.gov/media/144891/download (accessed 7.19.23).

FDA, 2021b. S5(R3) Detection of Reproductive and Developmental Toxicity for Human Pharmaceuticals Guidance for Industry [WWW Document]. URL https://www.fda.gov/media/148475/download

FDA, 2020. Qualification Process for Drug Development Tools Guidance for Industry and FDA Staff [WWW Document]. URL https://www.fda.gov/media/133511/download (accessed 7.19.23).

FDA, 2018. Bioanalytical Method Validation Guidance for Industry [WWW Document]. URL https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioanalytical-method-validation-guidance-industry (accessed 7.31.23).

FDA, 2017a. FDA's Predictive Toxicology Roadmap [WWW Document]. URL https://www.fda.gov/media/109634/download (accessed 5.5.22).

FDA, 2017b. Qualification of Medical Device Development Tools: Guidance for Industry, Tool Developers, and Food and Drug Administration Staff [WWW Document]. FDA. URL https://www.fda.gov/medical-devices/science-and-research-medical-devices/medical-device-development-tools-mddt (accessed 5.5.22).

Frick, A., Suzuki, O.T., Benton, C., Parks, B., Fedoriw, Y., Richards, K.L., Thomas, R.S., Wiltshire, T., 2015. Identifying genes that mediate anthracyline toxicity in immune cells. Front Pharmacol 6, 62. https://doi.org/10.3389/fphar.2015.00062

Friedman, K.P., Foster, M.J., Pham, L.L., Feshuk, M., Watford, S.M., Wambaugh, J.F., Judson, R.S., Setzer, R.W., Thomas, R.S., 2023. Reproducibility of organ-level effects in repeat dose animal studies. Comput Toxicol 28, 1–17. https://doi.org/10.1016/j.comtox.2023.100287

Harrill, A.H., 2020. ToxPoint: In the era of precision medicine, diversity should not be neglected in chemical safety assessment. Toxicol Sci 173, 3–4. https://doi.org/10.1093/toxsci/kfz232

Harrill, A.H., McAllister, K.A., 2017. New rodent population models may inform human health risk assessment and identification of genetic susceptibility to environmental exposures. Environ Health Perspect 125, 086002. https://doi.org/10.1289/EHP1274

Hartung, T., 2010. Evidence-based toxicology - the toolbox of validation for the 21st century? ALTEX 27, 253–263. https://doi.org/10.14573/altex.2010.4.253

Hoffmann, S., Kleinstreuer, N., Alépée, N., Allen, D., Api, A.M., Ashikaga, T., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Goebel, C., Kern, P.S., Klaric, M., Kühnl, J., Lalko, J.F., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Parakhia, R., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. Crit Rev Toxicol 48, 344–358. https://doi.org/10.1080/10408444.2018.1429385

Hoffmann, S., Saliner, A.G., Patlewicz, G., Eskes, C., Zuang, V., Worth, A.P., 2008. A feasibility study developing an integrated testing strategy assessing skin irritation potential of chemicals. Toxicol Lett 180, 9–20. https://doi.org/10.1016/j.toxlet.2008.05.004

ICCVAM, 2022. Symposium Webinar: Using New Approach Methodologies to Address Variability and Susceptibility Across Populations [WWW Document]. URL https://ntp.niehs.nih.gov/go/popvar (accessed 3.13.23).

ICCVAM, 2018. A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States [WWW Document]. URL https://ntp.niehs.nih.gov/iccvam/docs/roadmap/iccvam_strategicroadmap_january2018_document_508.pdf (accessed 5.3.22).

ICCVAM, 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods [WWW Document]. URL https://ntp.niehs.nih.gov/iccvam/suppdocs/subguidelines/sd_subg034508.pdf (accessed 5.3.22).

ICCVAM, 1997. Validation and Regulatory Acceptance of Toxicological Test Methods [WWW Document]. URL https://ntp.niehs.nih.gov/iccvam/docs/about_docs/validate.pdf (accessed 3.7.23).

ICH, 2013. ICH Harmonized Tripartite Guideline Photosafety Evaluation of Pharmaceuticals S10 [WWW Document]. URL https://database.ich.org/sites/default/files/S10_Guideline.pdf

Ishikawa, K., 1985. What is Total Quality Control? The Japanese Way. Translated by Lu, David J., 1st ed. Prentice-Hall.

Judson, R.S., Magpantay, F.M., Chickarmane, V., Haskell, C., Tania, N., Taylor, J., Xia, M., Huang, R., Rotroff, D.M., Filer, D.L., Houck, K.A., Martin, M.T., Sipes, N., Richard, A.M., Mansouri, K., Setzer, R.W., Knudsen, T.B., Crofton, K.M., Thomas, R.S., 2015. Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. Toxicol Sci 148, 137–154. https://doi.org/10.1093/toxsci/kfv168

Judson, R.S., Thomas, R.S., Baker, N., Simha, A., Howey, X.M., Marable, C., Kleinstreuer, N.C., Houck, K.A., 2019. Workflow for defining reference chemicals for assessing performance of in vitro assays. ALTEX 36, 261. https://doi.org/10.14573/altex.1809281

Karmaus, A.L., Mansouri, K., To, K.T., Blake, B., Fitzpatrick, J., Strickland, J., Patlewicz, G., Allen, D., Casey, W., Kleinstreuer, N., 2022. Evaluation of variability across rat acute oral systemic toxicity studies. Toxicological Sciences 188, 34–47. https://doi.org/10.1093/toxsci/kfac042

Kleinstreuer, N.C., Ceger, P., Watt, E.D., Martin, M., Houck, K., Browne, P., Thomas, R.S., Casey, W.M., Dix, D.J., Allen, D., Sakamuru, S., Xia, M., Huang, R., Judson, R., 2017. Development and validation of a computational model for androgen receptor activity. Chem. Res. Toxicol. 30, 946–964. https://doi.org/10.1021/acs.chemrestox.6b00347

Kleinstreuer, N.C., Hoffmann, S., Alépée, N., Allen, D., Ashikaga, T., Casey, W., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Göbel, C., Kern, P.S., Klaric, M., Kühnl, J., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Strickland, J., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches *. Crit Rev Toxicol 48, 359–374. https://doi.org/10.1080/10408444.2018.1429386

Kolle, S.N., Van Cott, A., van Ravenzwaay, B., Landsiedel, R., 2017. Lacking applicability of in vitro eye irritation methods to identify seriously eye irritating agrochemical formulations: Results of bovine cornea opacity and permeability assay, isolated chicken eye test and the EpiOcular; ET-50 method to classify according to UN GHS. Regulatory Toxicology and Pharmacology 85, 33–47. https://doi.org/10.1016/j.yrtph.2017.01.013

Krishna, S., Berridge, B., Kleinstreuer, N., 2021. High-throughput screening to identify chemical cardiotoxic potential. Chem Res Toxicol 34, 566–583. https://doi.org/10.1021/acs.chemrestox.0c00382

LaLone, C.A., Villeneuve, D.L., Lyons, D., Helgen, H.W., Robinson, S.L., Swintek, J.A., Saari, T.W., Ankley, G.T., 2016. Editor's Highlight: Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity. Toxicol. Sci. 153, 228–245. https://doi.org/10.1093/toxsci/kfw119

Luechtefeld, T., Maertens, A., Russo, D.P., Rovida, C., Zhu, H., Hartung, T., 2016. Analysis of public oral toxicity data from REACH registrations 2008-2014. ALTEX 33, 111–122. https://doi.org/10.14573/altex.1510054

Madia, F., Pillo, G., Worth, A., Corvi, R., Prieto, P., 2021. Integration of data across toxicity endpoints for improved safety assessment of chemicals: the example of carcinogenicity assessment. Arch Toxicol 95, 1971–1993. https://doi.org/10.1007/s00204-021-03035-x

NIEHS, 2023a. Testing Regulations and Guidelines [WWW Document]. National Toxicology Program. URL https://ntp.niehs.nih.gov/go/837330 (accessed 7.19.23).

NIEHS, 2023b. Funding Opportunities for Test Method Developers [WWW Document]. Funding Opportunities for Test Method Developers. URL https://ntp.niehs.nih.gov/go/alt-funding (accessed 5.9.23).

NIEHS, 2023c. International Cooperation on Alternative Test Methods [WWW Document]. International Cooperation on Alternative Test Methods. URL https://ntp.niehs.nih.gov/go/icatm (accessed 5.9.23).

NIEHS, 2023d. About ICCVAM [WWW Document]. About ICCVAM. URL https://ntp.niehs.nih.gov/go/iccvam (accessed 5.9.23).

OECD, 2023. No. 377: Initial Recommendations on Evaluation of Data from the Developmental Neurotoxicity (DNT) In-Vitro Testing Battery, OECD Series on Testing and Assessment. OECD Publishing, Paris.

OECD, 2022a. No. 364: Case study on the use of Integrated Approaches for Testing and Assessment for DNT to prioritize a class of Organophosphorus flame retardants, Series on Testing and Assessment. OECD Publishing, Paris.

OECD, 2022b. No. 367: Case Study on the use of an Integrated Approach for Testing and Assessment (IATA) for New Approach Methodology (NAM) for Refining Inhalation Risk Assessment from Point of Contact Toxicity of the Pesticide, Chlorothalonil, OECD Series on Testing and Assessment. OECD Publishing, Paris.

OECD, 2021a. Guideline No. 497: Defined Approaches on Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris.

OECD, 2021b. No. 298: Guiding Principles on Good Practices for the Availability/Distribution of Protected Elements in OECD Test Guidelines, Series on Testing and Assessment. OECD Publishing.

OECD, 2018. No. 286: Guidance Document on Good In Vitro Method Practices (GIVIMP), OECD Series on Testing and Assessment. OECD Publishing. https://doi.org/10.1787/9789264304796-en

OECD, 2017. No. 255: Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment, OECD Series on Testing and Assessment. OECD Publishing, Paris. https://doi.org/10.1787/9789264274822-en

OECD, 2014. No. 203: New Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Skin Corrosion and Irritation, OECD Series on Testing and Assessment. OECD Publishing, Paris.

OECD, 2007. No. 69: Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment. OECD Publishing, Paris. https://doi.org/10.1787/9789264085442-en.

OECD, 2005. No. 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, OECD Environment, Health and Safety Publications Series on Testing and Assessment. OECD Publishing, Paris.

OECD, 1998. No. 1: OECD Principles of Good Laboratory Practice, Series on Principles of Good Laboratory Practice and Compliance Monitoring. OECD Publishing, Paris.

Pamies, D., Leist, M., Coecke, S., Bowe, G., Allen, D.G., Gstraunthaler, G., Bal-Price, A., Pistollato, F., Vries, R.B.M. de, Hogberg, H.T., Hartung, T., Stacey, G., 2022. Guidance document on Good Cell and Tissue Culture Practice 2.0 (GCCP 2.0). ALTEX - Alternatives to animal experimentation 39, 30–70. https://doi.org/10.14573/altex.2111011

Parish, S.T., Aschner, M., Casey, W., Corvaro, M., Embry, M.R., Fitzpatrick, S., Kidd, D., Kleinstreuer, N.C., Lima, B.S., Settivari, R.S., Wolf, D.C., Yamazaki, D., Boobis, A., 2020. An evaluation framework for new approach methodologies (NAMs) for human health safety assessment. Regulatory Toxicology and Pharmacology 112, 104592. https://doi.org/10.1016/j.yrtph.2020.104592

Petersen, E., 2021. Characteristics to consider when selecting a positive control material for an in vitro assay. ALTEX. https://doi.org/10.14573/altex.2102111

Petersen, E.J., Ceger, P., Allen, D.G., Coyle, J., Derk, R., Garcia-Reyero, N., Gordon, J., Kleinstreuer, N.C., Matheson, J., McShan, D., Nelson, B.C., Patri, A.K., Rice, P., Rojanasakul, L., Sasidharan, A., Scarano, L., Chang, X., 2022a. U.S. federal agency interests and key considerations for new approach methodologies for nanomaterials.

ALTEX - Alternatives to animal experimentation 39, 183–206. https://doi.org/10.14573/altex.2105041

Petersen, E.J., Elliott, J.T., Gordon, J., Kleinstreuer, N.C., Reinke, E., Roesslein, M., Toman, B., 2022b. Technical framework for enabling high-quality measurements in new approach methodologies (NAMs). ALTEX - Alternatives to animal experimentation. https://doi.org/10.14573/altex.2205081

Petersen, E.J., Uhl, R., Toman, B., Elliott, J.T., Strickland, J., Truax, J., Gordon, J., 2022c. Development of a 96-well electrophilic allergen screening assay for skin sensitization using a measurement science approach. Toxics 10, 257. https://doi.org/10.3390/toxics10050257

Pham, L.L., Watford, S., Pradeep, P., Martin, M.T., Thomas, R., Judson, R., Setzer, R.W., Paul Friedman, K., 2020. Variability in in vivo studies: Defining the upper limit of performance for predictions of systemic effect levels. Comput Toxicol 15, 1–100126. https://doi.org/10.1016/j.comtox.2020.100126

Piersma, A.H., van Benthem, J., Ezendam, J., Kienhuis, A.S., 2018. Validation redefined. Toxicol In Vitro 46, 163–165. https://doi.org/10.1016/j.tiv.2017.10.013

Prior, H., Casey, W., Kimber, I., Whelan, M., Sewell, F., 2019. Reflections on the progress towards non-animal methods for acute toxicity testing of chemicals. Regul Toxicol Pharmacol 102, 30–33. https://doi.org/10.1016/j.yrtph.2018.12.008

Rooney, J.P., Choksi, N.Y., Ceger, P., Daniel, A.B., Truax, J., Allen, D., Kleinstreuer, N., 2021. Analysis of variability in the rabbit skin irritation assay. Regulatory Toxicology and Pharmacology 122, 104920. https://doi.org/10.1016/j.yrtph.2021.104920

Rösslein, M., Elliott, J.T., Salit, M., Petersen, E.J., Hirsch, C., Krug, H.F., Wick, P., 2015. Use of Cause-and-Effect Analysis to Design a High-Quality Nanocytotoxicology Assay. Chem. Res. Toxicol. 28, 21–30. https://doi.org/10.1021/tx500327y

Rusyn, I., Chiu, W.A., Wright, F.A., 2022. Model systems and organisms for addressing inter- and intra-species variability in risk assessment. Regul Toxicol Pharmacol 132, 105197. https://doi.org/10.1016/j.yrtph.2022.105197

Sewell, F., Doe, J., Gellatly, N., Ragan, I., Burden, N., 2017. Steps towards the international regulatory acceptance of non-animal methodology in safety assessment. Regul Toxicol Pharmacol 89, 50–56. https://doi.org/10.1016/j.yrtph.2017.07.001

Shaffer, R.M., 2021. Environmental Health Risk Assessment in the Federal Government: A Visual Overview and a Renewed Call for Coordination. Environ. Sci. Technol. 55, 10923–10927. https://doi.org/10.1021/acs.est.1c01955

Slezák, P., Waczulíková, I., 2011. Letter to the Editor: Reproducibility and Repeatability. Physiol Res 60, 203–205.

Smirnova, L., Hogberg, H.T., Leist, M., Hartung, T., 2014. Developmental neurotoxicity - challenges in the 21st century and in vitro opportunities. ALTEX 31, 129–156. https://doi.org/10.14573/altex.1403271

Strickland, J., Clippinger, A.J., Brown, J., Allen, D., Jacobs, A., Matheson, J., Lowit, A., Reinke, E.N., Johnson, M.S., Quinn, M.J., Mattie, D., Fitzpatrick, S.C., Ahir, S., Kleinstreuer, N., Casey, W., 2018. Status of acute systemic toxicity testing requirements and data uses by U.S. regulatory agencies. Regul Toxicol Pharmacol 94, 183–196. https://doi.org/10.1016/j.yrtph.2018.01.022

Strickland, J., Daniel, A.B., Allen, D., Aguila, C., Ahir, S., Bancos, S., Craig, E., Germolec, D., Ghosh, C., Hudson, N.L., Jacobs, A., Lehmann, D.M., Matheson, J., Reinke, E.N., Sadrieh, N., Vukmanovic, S., Kleinstreuer, N., 2019. Skin sensitization testing needs and data uses by US regulatory and research agencies. Arch Toxicol 93, 273–291. https://doi.org/10.1007/s00204-018-2341-6

Thomas, R.S., Bahadori, T., Buckley, T.J., Cowden, J., Deisenroth, C., Dionisio, K.L., Frithsen, J.B., Grulke, C.M., Gwinn, M.R., Harrill, J.A., Higuchi, M., Houck, K.A., Hughes, M.F., Hunter, E.S., Isaacs, K.K., Judson, R.S., Knudsen, T.B., Lambert, J.C., Linnenbrink, M., Martin, T.M., Newton, S.R., Padilla, S., Patlewicz, G., Paul-Friedman, K., Phillips, K.A., Richard, A.M., Sams, R., Shafer, T.J., Setzer, R.W., Shah, I., Simmons, J.E., Simmons, S.O., Singh, A., Sobus, J.R., Strynar, M., Swank, A., Tornero-Valez, R., Ulrich, E.M., Villeneuve, D.L., Wambaugh, J.F., Wetmore, B.A., Williams, A.J., 2019. The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. Toxicol Sci 169, 317–332. https://doi.org/10.1093/toxsci/kfz058

Tox21, n.d. Tox21 - Toxicology in the 21st Century [WWW Document]. Tox21. URL https://tox21.gov/ (accessed 5.9.23).

Tsuji, R., Crofton, K.M., 2012. Developmental neurotoxicity guideline study: issues with methodology, evaluation and regulation. Congenit Anom (Kyoto) 52, 122–128. https://doi.org/10.1111/j.1741-4520.2012.00374.x

van der Zalm, A.J., Barroso, J., Browne, P., Casey, W., Gordon, J., Henry, T.R., Kleinstreuer, N.C., Lowit, A.B., Perron, M., Clippinger, A.J., 2022. A framework for establishing scientific confidence in new approach methodologies. Arch Toxicol 96, 2865–2879. https://doi.org/10.1007/s00204-022-03365-4

Wikoff, D., Lewis, R.J., Erraguntla, N., Franzen, A., Foreman, J., 2020. Facilitation of risk assessment with evidence-based methods – A framework for use of systematic mapping and systematic reviews in determining hazard, developing toxicity values, and characterizing uncertainty. Regulatory Toxicology and Pharmacology 118, 104790. https://doi.org/10.1016/j.yrtph.2020.104790

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wolffe, T.A.M., Whaley, P., Halsall, C., Rooney, A.A., Walker, V.R., 2019. Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. Environ Int 130, 104871. https://doi.org/10.1016/j.envint.2019.05.065

Zhang, J.H., Chung, T.D., Oldenburg, K.R., 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4, 67–73. https://doi.org/10.1177/108705719900400206

Zhang, X.D., 2011. Illustration of SSMD, z score, SSMD*, z* score, and t statistic for hit selection in RNAi high-throughput screens. J Biomol Screen 16, 775–785. https://doi.org/10.1177/1087057111405851

ANNEXE A : GLOSSAIRE

AVERTISSEMENT : ce glossaire est destiné à faciliter la compréhension de la terminologie employée dans ce rapport. La signification ou l'utilisation d'un terme particulier peut varier d'un organisme de réglementation à l'autre.

Adapté à l'objectif : correspondance du type et du degré de certitude des informations fournies par une NAM (ou un ensemble de NAM) avec le type et le degré de certitude des informations nécessaires pour une décision donnée (EPA, 2021a).

Approche définie (AD) : obtention d'un résultat, à l'aide de données d'entrée générées à partir d'un ensemble défini de sources d'informations et d'une procédure d'interprétation des données fixe, qui peut être utilisé seul ou conjointement avec d'autres sources d'informations au sein d'une IATA pour répondre à un besoin réglementaire précis. Une approche définie des essais et de l'évaluation peut être utilisée pour faciliter l'identification des dangers, la caractérisation des dangers et/ou l'évaluation de la sécurité des substances chimiques (OCDE, 2017).

Approche intégrée en matière d'essai et d'évaluation (IATA) : approche basée sur de multiples sources d'information utilisées pour l'identification des dangers, la caractérisation des dangers et/ou l'évaluation de la sécurité des substances chimiques (OCDE, 2017).

Base de données/liste préparée : ensemble structuré d'informations fiables et bien caractérisées, soigneusement compilées et mises à jour.

Bonnes pratiques de culture cellulaire (BPCC) : ensemble de principes développés pour une utilisation pratique en laboratoire afin de garantir la reproductibilité des analyses *in vitro* (basées sur des cellules) et d'améliorer la qualité des données scientifiques (Pamies et al., 2022).

Bonnes pratiques de laboratoire (BPL) : règlements promulgués par des autorités telles que l'EPA des États-Unis, la FDA des États-Unis et l'OCDE qui décrivent les procédures de tenue des dossiers et d'assurance qualité pour les dossiers de laboratoire qui constitueront la base des soumissions de données aux organismes nationaux de réglementation.

Bonnes pratiques de laboratoire appliquées aux études in vitro : cadre complet décrit dans le document d'orientation de l'OCDE intitulé Guidance Document on Good In Vitro Method Practices (« Document d'orientation sur les bonnes pratiques relatives aux méthodes in vitro ») (OCDE, 2018), qui fournit des recommandations pour le développement, la validation, l'acceptation ou la qualification réglementaire et l'utilisation de méthodes *in vitro*.

Classification des dangers : classement d'un danger chimique ou d'un produit dans une catégorie sur la base des résultats d'une méthode d'essai standard pour un critère d'évaluation de toxicité spécifique ; le plus souvent utilisée à des fins d'étiquetage.

Composés de référence : substances chimiques sélectionnées pour être utilisées lors de la recherche, du développement ou de l'évaluation d'une méthode d'essai proposée, parce que leur réponse lors de la méthode d'essai de référence ou chez l'espèce concernée est connue (voir « méthode d'essai de référence »).

Concordance : degré d'accord ou de cohérence entre deux variables. La concordance peut être utilisée pour décrire qualitativement la pertinence biologique d'une NAM par rapport à la méthode d'essai sur l'animal de référence et/ou aux données de référence chez l'homme éventuellement disponibles. La concordance peut également servir à décrire quantitativement la proportion de toutes les substances chimiques testées qui sont correctement classées comme positives ou négatives. Ce terme est souvent utilisé de manière interchangeable avec celui de « taux de reconnaissance global » dans ce contexte.

Contexte d'utilisation (CU) : description clairement articulée du mode et de l'objectif de l'utilisation d'une méthode, d'une approche ou d'une application particulière.

Contrôle qualité : ensemble d'activités ou d'échantillons utilisés pour vérifier le maintien de la qualité attendue du produit ou de la méthode.

Courbe standard : méthode quantitative permettant une représentation graphique des données d'essai pour déterminer la concentration d'une substance dans un échantillon inconnu en comparant celui-ci à un échantillon standard de concentration connue (en utilisant souvent le témoin positif).

Critère d'évaluation : processus, réponse ou effet biologique ou chimique évalué par une méthode d'essai.

Danger : possibilité de survenue d'un effet indésirable ou nocif pour la santé ou l'environnement.

Discordance : proportion de l'ensemble des substances chimiques testées qui sont erronément classées comme positives ou négatives.

Domaine d'application : types de substances chimiques qui peuvent être testées à l'aide d'une méthode, ou types de substances chimiques pour lesquels les résultats produits par cette méthode sont considérés comme acceptables.

Élaborateur de la méthode d'essai : organisation ou personne qui conçoit initialement une méthode d'essai et veille à sa reproductibilité et son adéquation à l'utilisation prévue.

Espèce cible : espèce chez laquelle des informations relatives à la toxicité potentielle d'une substance chimique sont recherchées.

Essai : système expérimental utilisé. Terme souvent utilisé de manière interchangeable avec ceux de « test » et de « méthode d'essai ».

Évaluation des risques : évaluation des effets nocifs potentiels sur la santé et l'environnement d'une espèce cible résultant de l'exposition à des agents exogènes.

Faux négatif : substance identifiée à tort comme négative par une méthode d'essai à la lumière des données de référence spécifiées.

Faux positif : substance identifiée à tort comme positive par une méthode d'essai à la lumière des données de référence spécifiées.

Fiabilité : mesure de la capacité d'une méthode d'essai à être exécutée de manière reproductible au sein d'un laboratoire et entre laboratoires au fil du temps. Elle est évaluée en calculant la reproductibilité intra- et interlaboratoires et la répétabilité intralaboratoires.

Identification des dangers : partie de l'évaluation des risques qui permet de déterminer si l'exposition à une substance particulière est ou pourrait être associée à des effets indésirables sur la santé ou l'environnement.

Méthode d'essai de référence : méthode d'essai acceptée utilisée à des fins réglementaires pour évaluer la possibilité qu'une substance à tester soit dangereuse pour l'espèce considérée.

Méthode d'essai validée : méthode d'essai acceptée pour laquelle des études de validation ont été réalisées afin de déterminer le taux de reconnaissance global, la fiabilité et la pertinence de cette méthode pour une utilisation proposée particulière.

Méthode d'essai : procédé ou procédure utilisé pour caractériser une substance ou un agent. Les méthodes d'essai toxicologiques fournissent des informations concernant la capacité d'une substance ou d'un agent à produire un effet biologique spécifié dans des conditions établies. Ce terme est employé de manière interchangeable avec ceux de « test » et d'« essai ». Voir également « méthode d'essai validée » et « méthode d'essai de référence ».

Normes de performance : normes fondées sur une méthode d'essai validée qui fournissent une base pour évaluer la comparabilité d'une méthode d'essai proposée qui est similaire du point de vue mécanistique et fonctionnel. Sont inclus (1) les composants essentiels de la méthode d'essai ; (2) une liste de substances chimiques de référence sélectionnées parmi les substances chimiques utilisées pour démontrer la performance acceptable de la méthode d'essai validée ; et (3) les niveaux comparables de taux de reconnaissance global et de fiabilité, sur la base des résultats de la méthode d'essai validée, que la méthode d'essai proposée devrait présenter lorsqu'elle est évaluée à l'aide de la liste minimale de substances chimiques de référence.

Nouvelles méthodes d'approche (NAM) : références descriptives générales à toute technologie, méthodologie, approche ou combinaison de celles-ci qui peut être utilisée pour fournir des informations sur l'évaluation des dangers et des risques des substances chimiques et qui promeut le remplacement, la réduction ou le réajustement de l'utilisation des animaux (principes des 3R).

Organe cible : organe sur lequel des informations relatives à la toxicité potentielle d'une substance chimique sont recherchées.

Performance : caractéristiques de taux de reconnaissance global et de fiabilité d'une méthode d'essai (voir « taux de reconnaissance global » et « fiabilité »).

Pertinence biologique : mesure du caractère approprié d'une méthode censée évaluer les effets d'une substance chimique au sein des taxons considérés.

Pertinence mécanistique : mesure du caractère approprié pour l'évaluation du processus ou de la voie biochimique par lequel une substance chimique peut exercer un effet.

Précision balancée : paramètre statistique utilisé pour corriger un ensemble de données déséquilibré dans lequel une « classe » est beaucoup plus fréquente que l'autre (p. ex. plus grand nombre de données négatives que positives). La précision balancée correspond à la moyenne arithmétique de la sensibilité et de la spécificité.

Précision : proximité des différentes mesures d'un analyte après de multiples analyses, souvent sur un seul échantillon. La précision est souvent exprimée par le coefficient de variation.

Prédictivité négative : proportion de réponses négatives correctes à la lumière des données de référence définies parmi les substances testées négatives par une méthode d'essai. Il s'agit d'un indicateur du taux de reconnaissance global de la méthode d'essai. La prédictivité négative est fonction de la spécificité de la méthode d'essai et de la prévalence de négatifs parmi les substances testées.

Prédictivité positive : proportion de réponses positives correctes à la lumière des données de référence définies parmi les substances testées positives par une méthode d'essai. Il s'agit d'un indicateur du taux de reconnaissance global de la méthode d'essai. La prédictivité positive est fonction de la sensibilité de la méthode d'essai et de la prévalence de positifs parmi les substances testées.

Principes des 3R : principes d'une technique expérimentale sans cruauté, supposant en particulier la réduction (utiliser aussi peu d'animaux que possible), le remplacement (utiliser une espèce d'ordre inférieur ou des essais ne portant pas sur des animaux) ou le réajustement (atténuer au maximum la douleur ou la détresse que les animaux de laboratoire peuvent endurer en raison d'une technique particulière) de l'utilisation d'animaux dans la recherche et les tests de sécurité chimique.

Procédures opérationnelles standard (SOP) : procédures écrites formelles qui décrivent la manière dont chaque opération de laboratoire doit être effectuée. Elles sont requises par les directives des BPL.

Promoteur de la méthode d'essai : organisation ou personne à l'origine de la soumission d'une méthode d'essai ; il peut s'agir de l'organisation ou de la personne jouant le rôle d'élaborateur de la méthode d'essai.

Protocole : description précise étape par étape d'une méthode d'essai, y compris la liste de tous les réactifs et de tous les critères et procédures nécessaires pour générer et évaluer les données de test.

Puissance : mesure de l'activité biologique ou chimique relative d'une substance. La puissance d'une substance unique peut varier selon les effets biologiques ou biochimiques considérés.

Qualification : conclusion que les résultats d'une évaluation basée sur un modèle ou un essai validé sont fiables en vue d'une interprétation et d'une application précises dans le cadre du développement de produits et de la prise de décisions réglementaires.

Répétabilité : cohérence des résultats de test obtenus lorsque la procédure est effectuée sur la même substance dans des conditions identiques au cours d'une période donnée ; « … degré d'accord entre les résultats indépendants obtenus avec la même méthode sur le ou les mêmes sujets (ou objets ou matériaux de test), dans les *mêmes conditions* » (Slezák et Waczulíková, 2011).

Reproductibilité interlaboratoires : mesure permettant de déterminer si différents laboratoires qualifiés utilisant le même protocole et les mêmes substances chimiques à tester peuvent produire des résultats qualitativement et quantitativement similaires. La reproductibilité interlaboratoires est établie au cours des processus de prévalidation et de validation et indique dans quelle mesure une méthode d'essai peut être transférée avec succès entre laboratoires.

Reproductibilité : cohérence des résultats de test individuels obtenus en appliquant le même protocole de test aux mêmes échantillons ; « … degré d'accord entre les résultats indépendants obtenus avec la même méthode sur le ou les mêmes sujets (ou objets ou matériel d'essai), mais *dans des conditions différentes* (différents observateurs, laboratoires, etc.) » (Slezák et Waczulíková, 2011).

Robustesse : capacité d'une méthode à être reproduite dans différentes conditions ou circonstances sans que les résultats obtenus montrent des différences inattendues.

Sensibilité : proportion de l'ensemble des substances chimiques positives qui sont correctement classées comme telles dans une méthode d'essai. La sensibilité peut également être définie dans le contexte des limites de détection comme la plus faible concentration d'analyte pouvant être mesurée avec un taux de reconnaissance global et/ou une précision acceptables.

Soumission de la méthode d'essai : ensemble de la documentation à l'appui d'une méthode d'essai proposée pour une application réglementaire ou une autre application précise. Une soumission de la méthode d'essai comprend généralement des dossiers des études de validation réalisées pour caractériser l'utilité et les limitations de la méthode d'essai pour une exigence ou une application d'essai réglementaire particulière ainsi que d'autres documents adéquats sur la validité scientifique préparés conformément aux directives de soumission de la méthode d'essai.

Spécificité : proportion de l'ensemble des substances chimiques négatives correctement classées comme telles dans une méthode d'essai. Il s'agit d'une mesure du taux de reconnaissance global

de la méthode d'essai. Ce terme est également utilisé pour décrire la capacité d'une méthode analytique à détecter un analyte particulier.

Stabilité : capacité d'un matériau de test (p. ex. substance à tester, appareil d'essai, réactif ou analyte) à produire des résultats similaires et acceptables sur une période donnée dans un environnement donné.

Taux de reconnaissance global : degré de concordance entre le résultat d'une méthode d'essai et une valeur de référence acceptée.

Taux de recouvrement : méthode quantitative permettant de vérifier l'efficacité et la reproductibilité d'une méthode d'extraction en comparant les résultats des échantillons extraits avec ceux d'échantillons enrichis d'une matrice similaire et/ou de blancs enrichis.

Test : système expérimental utilisé ; terme employé de manière interchangeable avec ceux de « méthode d'essai » et d'« essai ».

Transférabilité : capacité d'une méthode ou d'une procédure d'essai à être exécutée de manière exacte et fiable dans différents laboratoires compétents. Voir également « reproductibilité interlaboratoires ».

Validation : processus par lequel le taux de reconnaissance global, la fiabilité et la pertinence d'une procédure sont établies pour un objectif précis. La validation pour un objectif donné n'implique pas une validation pour d'autres objectifs. Une qualification supplémentaire peut être nécessaire pour un contexte d'utilisation particulier.

Voie conduisant à des effets indésirables (AOP) : représentation structurée d'événements séquentiels qui se produisent à différents niveaux de l'organisation biologique et entraînent un effet indésirable lorsqu'un organisme est exposé à une substance.

ANNEXE B : OUTILS DE QUALITÉ

1.0 Organigrammes

Dans un organigramme, chaque étape d'un protocole est schématisée (voir exemple de la Figure S1). En garantissant la couverture et le suivi de chaque étape, dans la mesure du possible, il peut être utile à la conception de mesures de contrôle pour la méthode d'essai. Il est possible que plusieurs mesures de contrôle couvrent une seule étape, et inversement qu'une mesure de contrôle couvre plusieurs étapes. En outre, la comparaison des mesures de contrôle de différentes NAM peut révéler qu'elles comportent des étapes similaires. Pour ces étapes, les sources de variabilité seront probablement similaires.

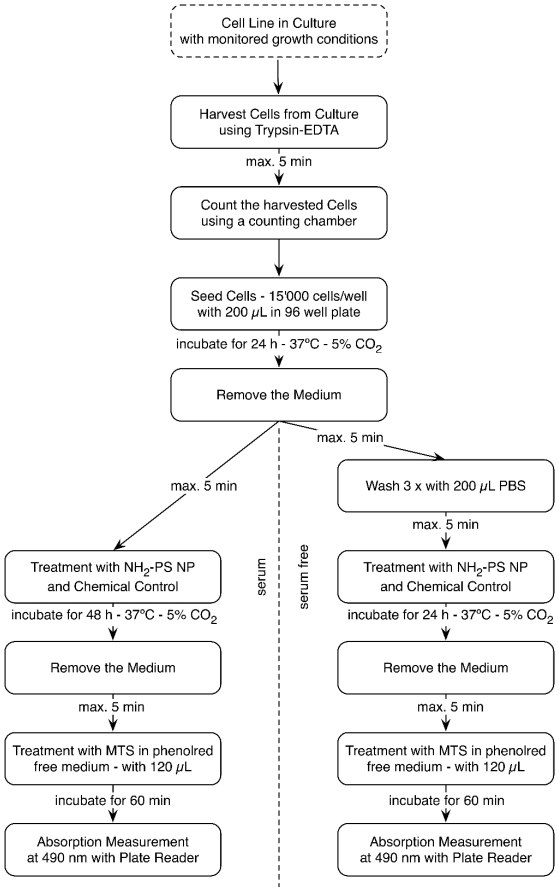


Figure supplémentaire 1. Organigramme décrivant le protocole modifié relatif au MTS (3-(4,5-diméthylthiazol-2-yl)-5-(3-carboxyméthoxyphényl)-2-(4-sulfophényl)-2H-tétrazolium) pour une étude interlaboratoires. Cette figure est réimprimée avec à l'autorisation d'Elliott et al. (2017).

2.0 Analyse des causes et des effets

L'analyse des causes et des effets (C&E) est un outil conceptuel qui peut être utilisé pour identifier d'éventuelles sources importantes de variabilité et les présenter sous forme de diagrammes C&E, parfois appelés diagrammes en arêtes de poisson (voir exemple de la Figure S2). Le processus d'élaboration de ces diagrammes peut inclure une réflexion collective et une revue de la littérature sur un essai. Chaque branche du diagramme C&E indique une source importante de variabilité attendue. Les diagrammes C&E peuvent également faciliter le développement de nouvelles NAM, car les branches communes des diagrammes C&E (p. ex. réalisation de mesures à l'aide du même type d'essai de cytotoxicité) pourront être effectuées plus rapidement, nécessiteront probablement des mesures de contrôle semblables et comporteront sans doute des stratégies d'atténuation de la variabilité similaires. L'analyse des diagrammes C&E aide à identifier les aspects d'une méthode dont la standardisation peut être complexe (p. ex. instrument difficile à étalonner ou réactifs de test instables). Les diagrammes C&E peuvent guider les tests de robustesse et la sélection des mesures de contrôle afin d'analyser, dans l'idéal, les sources de variabilité dans chaque branche et sous-branche du diagramme C&E.

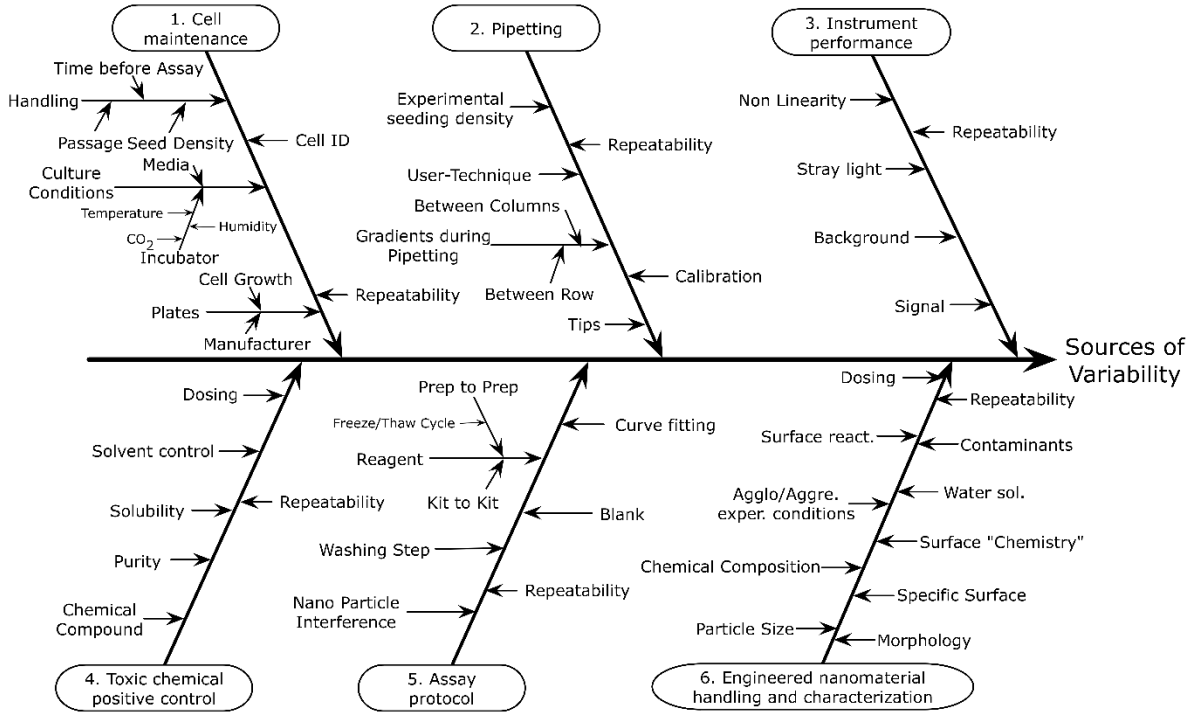


Figure supplémentaire 2. Diagramme des causes et des effets d'un test de cytotoxicité du MTS conçu pour être utilisé avec des nanomatériaux modifiés. Reproduit avec l'autorisation de Rösslein et al. (2015).

3.0 Cartes de contrôle

Les mesures de contrôle peuvent prendre la forme d'expériences préliminaires ponctuelles (p. ex. recherche de biais potentiels), de mesures périodiques effectuées à une fréquence prédéterminée (p. ex. évaluation de la performance et de l'étalonnage de l'instrument) et de mesures de contrôle en cours de processus effectuées chaque fois que l'essai est effectué. Pour certains essais qui permettent d'analyser simultanément un nombre limité d'échantillons (p. ex. essai par inhalation avec exposition à des substances chimiques en aérosol à l'aide d'un système d'exposition continue), des mesures périodiques peuvent être utiles. Des essais préliminaires ponctuels peuvent être utiles pour évaluer si une substance à tester peut provoquer un biais, par exemple si une nanoparticule peut adsorber un réactif d'essai important. Les mesures de contrôle en cours de processus peuvent être utilisées pour mesurer les principales sources de variabilité chaque fois que l'essai est effectué. Le témoin positif constitue par exemple une mesure de contrôle en cours de processus courante. Cette mesure de contrôle peut révéler si une réponse maximale (effet de 100 %) est atteinte et être utilisée pour démontrer la sensibilité de la réponse de l'essai à des concentrations chimiques qui produisent des réponses plus modérées. Les principales considérations pour la sélection d'un témoin positif ont été décrites par Petersen et al. (2021). Les autres mesures de contrôle en cours de processus courantes sont un témoin sans cellules et sans réactif d'essai supplémentaire et un témoin contenant des cellules ainsi que des réactifs d'essai ajoutés, mais sans substances à tester. Toutes les mesures de contrôle en cours de processus potentielles ne peuvent être incluses, et des compromis sont parfois nécessaires en la matière.

Les principales sources de variabilité peuvent être surveillées dans le temps et entre tests à l'aide de cartes de contrôle qui affichent les valeurs moyennes et la variabilité des mesures de contrôle en cours de processus (voir exemple de la Figure S3). Elles peuvent aider à détecter des changements systématiques des valeurs moyennes ou de variabilité au fil du temps, qui suggèrent l'altération d'un composant de l'essai (p. ex. instabilité d'un réactif). L'examen des feuilles de relevé peut être utile pour identifier les raisons des changements.

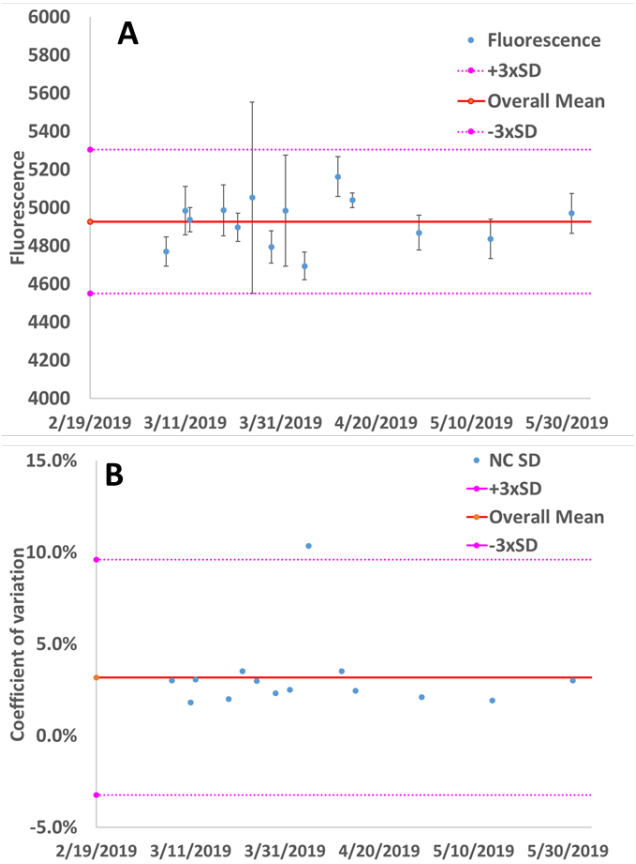


Figure supplémentaire 3. Données des cartes de contrôle pour la méthode de fluorescence de l'essai de dépistage des allergènes électrophiles pour le témoin négatif d'un incubateur : (A) moyenne et (B) coefficient de variation pour tous les tests en fonction de la date de réalisation. Dans le graphique B, une valeur est aberrante pour le coefficient de variation et ne correspond pas aux spécifications de cette étude (moyenne globale ± 3 fois la valeur moyenne de l'écart type). De plus, ni la moyenne ni le coefficient de variation ne montrent de tendance systématique en fonction du temps. Cette figure a été modifiée et réimprimée avec l'autorisation de Petersen et al. (2022c), tandis que la légende de la figure est modifiée et réimprimée d'après Petersen et al. (2022b).

4.0 Feuilles de relevé

Des feuilles de relevé peuvent être utilisées pour enregistrer les métadonnées clés et répondre aux exigences des BPL. Si les BPL ne sont pas utilisées lors du développement d'une méthode, les principes énoncés dans les BPL, tels que l'enregistrement des principales données à l'aide de feuilles de relevé, seront importants si la NAM est destinée à des soumissions réglementaires, lorsque la conformité aux BPL peut être requise. Les ressources alternatives pertinentes lors du développement d'essais sont le document d'orientation sur le GCCP 2.0 (Pamies et al., 2022) et le GIVIMP (OCDE, 2018), qui sont spécifiques aux méthodes cellulaires avec des listes de gestion de la qualité supplémentaires pour garantir la reproductibilité et des données scientifiques de haute qualité. Les feuilles de relevé peuvent également être utiles pour surveiller les mesures de contrôle en cours de processus et faciliter la résolution des éventuels problèmes. Par exemple, elles peuvent être utilisées pour documenter les modifications des numéros de lot des réactifs et des consommables (p. ex. pipettes et plaques à micropuits). Cela peut faciliter la résolution des problèmes en cas de non-respect particulièrement fréquent d'une spécification donnée. Les feuilles de relevé peuvent également inclure des calculateurs de données pour évaluer si les données d'un essai répondent à toutes les spécifications et effectuer des évaluations statistiques des résultats de l'essai. Des blocs-notes électroniques et des systèmes de gestion des informations de laboratoire peuvent être substitués aux feuilles de relevé papier pour plus de commodité (stockage aisé, transférabilité entre laboratoires et recherche simplifiée).

5.0 Diagrammes de dispersion

Les diagrammes de dispersion peuvent être utilisés pour évaluer s'il existe une interaction entre différentes mesures de contrôle en cours de processus ou entre ces mesures de contrôle et les résultats de la substance à tester (voir exemple de la Figure S4). Les résultats de la substance à tester devraient être indépendants des résultats des mesures de contrôle en cours de processus dans la plage définie par les spécifications.

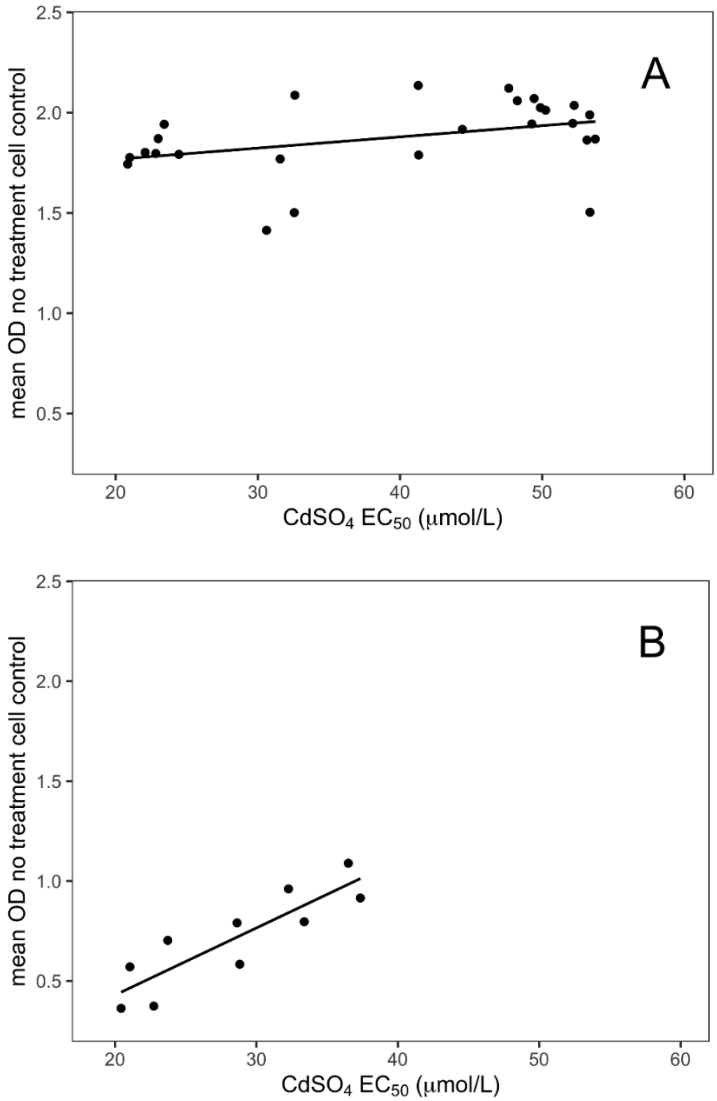


Figure supplémentaire 4. Corrélation entre les valeurs de la CE_{50} du sulfate de cadmium ($CdSO_4$) déterminées à l'aide de l'essai du MTS et les valeurs moyennes de densité optique (DO) de cellules A549 non exposées à une substance à tester ou au témoin positif. Ces données montrent soit une absence d'interaction entre les valeurs de la CE_{50} (partie A), soit une interaction (partie B), en fonction de la plage des valeurs moyennes de DO. Les lignes continues sont des ajustements de régression linéaire. La pente de la partie B est statistiquement différente de 0, ce qui indique que la valeur de la CE_{50} est corrélée aux valeurs de DO. Modifié et réimprimé avec l'autorisation de Petersen et al. (2022b) ; la légende de la figure est réimprimée avec l'autorisation de Petersen et al. (2022b).

ANNEXE C : ÉVALUATION DE LA MÉTHODE ANALYTIQUE

Les informations de la présente annexe proviennent de plusieurs références, notamment les publications Validation and Regulatory Acceptance of Toxicological Test Methods (« Validation et acceptation réglementaire des méthodes d'essai toxicologiques ») (ICCVAM, 1997) et Bioanalytical Method Validation Guidance for Industry (« Guide de validation des méthodes bioanalytiques pour l'industrie ») (FDA, 2018), le GD 34 de l'OCDE (OCDE, 2005) et le document GIVIMP de l'OCDE (OCDE, 2018). Ces informations visent à fournir un cadre général pour l'évaluation des méthodes. Certains aspects peuvent ne pas être applicables à toutes les méthodes.

1.0 Limites de détection et de quantification

Les limites de détection et de quantification définissent l'intervalle d'une méthode analytique. La méthode analytique est souvent testée soit à la limite inférieure de quantification, soit à la limite de détection, en fonction des caractéristiques du système. Ce processus détermine la limite inférieure de l'intervalle d'une méthode analytique en établissant la quantité ou la concentration la plus faible d'un analyte qui peut être détectée (limite de détection) ou quantifiée (limite inférieure de quantification) de manière fiable au-dessus du blanc de réactif ou dans la courbe standard. La concentration la plus élevée sur la courbe standard correspond à la limite supérieure de quantification.

2.0 Identification des interférences

Les élaborateurs de méthodes devraient documenter les substances provoquant des interférences, qui peuvent provenir de composants critiques et non critiques de la méthode, y compris toute interférence avec le signal détecté (p. ex. fluorescence/absorbance, luciférase, enzymatique) de la méthode. Les interférences peuvent également provenir des consommables, tels que certains plastiques utilisés dans les méthodes d'essai des perturbateurs endocriniens ainsi que d'autres réactifs de la méthode. Les substances susceptibles d'induire des interférences comprennent entre autres les composants de la matrice endogène, les métabolites, les produits de décomposition et d'autres xénobiotiques. Si la méthode vise à quantifier plus d'un analyte, tous les analytes prévus devraient être testés pour s'assurer qu'il n'y a pas d'interférence. Des échantillons témoins sont souvent utilisés pour tester les interférences.

3.0 Évaluation de la précision analytique

Il est utile d'inclure une description de la précision de la méthode analytique utilisée et de tout autre test de précision, comme ceux qui évaluent la variabilité de la performance lorsque différents membres du personnel utilisent la méthode proposée ou lorsque différents instruments sont utilisés pour la méthode ; dans la mesure du possible, la participation à des études de comparaison interlaboratoires est également recommandée.

4.0 Stabilité des matériaux

La stabilité des matériaux utilisés dans la NAM (p. ex. substances à tester, appareil d'essai, réactifs et analytes) fait référence à la capacité d'un matériau à produire des résultats similaires et acceptables sur une période donnée dans un environnement donné. La stabilité de la NAM elle-même, évaluée au moyen de cartes de contrôle en fonction du temps, est abordée plus haut (Annexe B, section 3.0). Les effets du prélèvement, de la manipulation et des conditions de conservation des échantillons devraient être évalués. Les matériaux susceptibles d'être affectés par des problèmes de stabilité devraient être assortis d'une date de péremption appropriée. La stabilité de la substance à tester, des réactifs et de l'appareil d'essai (p. ex. microplaque en plastique) devrait être garantie pour éviter les interférences causées par les produits de dégradation et les modifications de la dose réelle appliquée. Les études de stabilité réalisées sur tout composant devraient être incluses dans le rapport. Tout mélange chimique ou échantillon préparé avant le jour de l'étude devrait faire l'objet d'une évaluation de la stabilité avant d'être utilisé dans des études de développement, de validation ou de qualification.

La stabilité chimique d'une matrice ou d'un mélange donné dans des conditions précises sur des intervalles de temps donnés est évaluée de plusieurs manières. Les évaluations de stabilité préalables à l'étude devraient couvrir les conditions de manipulation et de conservation attendues des échantillons pendant la conduite de l'étude, y compris les conditions sur le site d'essai, pendant le transport et sur tous les sites secondaires. La stabilité d'un analyte dans un mélange, une matrice et un système de récipients particuliers est spécifique à ce mélange, cette matrice et ce système de récipients et ne peut être généralisée. Les tests de stabilité devraient évaluer la stabilité des analytes pour la conservation à long terme (échantillons congelés à la température de conservation prévue) et à court terme (plan de travail, température ambiante), et après les cycles de congélation et décongélation et le processus analytique. Les conditions des tests de stabilité devraient refléter les situations susceptibles d'être rencontrées lors de la manipulation et de l'analyse réelles des échantillons. Si, au cours de l'analyse d'échantillons pour une étude, les conditions de conservation ont changé et/ou ont enfreint les conditions de conservation des échantillons évaluées lors de la caractérisation technique de la méthode, la stabilité devrait être établie dans ces nouvelles conditions. Des études de stabilité indépendantes devraient être menées et couvrir toutes les conditions auxquelles un réactif critique peut être soumis. Les tests de stabilité d'un réactif ou d'un matériau devraient être envisagés dans divers contextes, notamment les cycles de congélation-décongélation et l'utilisation sur le plan de travail, dans des solutions mères, dans des échantillons traités et dans le cadre d'études à long terme.

5.0 Test de robustesse

La robustesse est la capacité d'une méthode à être reproduite dans différentes conditions ou circonstances sans que les résultats obtenus montrent des différences inattendues. L'un des aspects de la robustesse d'une NAM est la cohérence des résultats des cartes de contrôle dans le temps, comme décrit à l'Annexe B, section 3.0. Les tests de robustesse sont souvent utilisés pour détecter des changements des résultats dus à des variations involontaires des réactifs ou des protocoles expérimentaux. Ils sont recommandés pour tous les aspects des méthodes d'essai, et

des intervalles pour tous les paramètres et mesures devraient être établis chaque fois que cela est possible.

Par exemple, si un temps d'incubation de 5 minutes a été jugé optimal pour une étude, mais qu'après un test de robustesse, les données ont satisfait à toutes les exigences de contrôle qualité pour un temps de 5 minutes ± 30 secondes, le temps d'incubation acceptable testé pour la robustesse est de 5 minutes ± 30 secondes.

Voici quelques-uns des paramètres de l'étude pour lesquels un intervalle d'acceptation devrait être établi :

- Temps d'incubation
- Températures d'incubation
- pH
- Sources de réactifs
- Densités cellulaires (le cas échéant)
- Conditions expérimentales
- Logiciel d'analyse

La conception du modèle déterminera quelles conditions expérimentales nécessiteront des intervalles d'acceptation. Par exemple, les modèles comportant un milieu fluide auront besoin d'intervalles d'acceptation pour les paramètres liés au débit. Les modèles plus complexes sont plus susceptibles d'avoir des paramètres d'étude nécessitant un contrôle.

Le cas échéant, les résultats des tests de robustesse des réactifs critiques et non critiques devraient également être communiqués. Différents fournisseurs (dans la mesure du possible) devraient être testés pour déterminer s'il convient de se procurer un réactif auprès d'un seul fournisseur ou si plusieurs fournisseurs peuvent être utilisés. Cela peut également être utile pour se prémunir contre des perturbations de la chaîne d'approvisionnement ou l'arrêt de la fabrication d'un produit par un fabricant particulier.

Des tests de robustesse des instruments devraient également être effectués chaque fois que cela est possible. La validation interlaboratoires porte souvent sur différentes marques d'instruments ayant des paramètres de performance et des capacités différents, ce qui peut accroître la variabilité des données ou modifier les paramètres de la méthode utilisée.

6.0 Analyse du taux de recouvrement

Lors d'essais d'extraction à partir d'une matrice donnée, des études du taux de recouvrement devraient être réalisées afin de vérifier l'efficacité et la reproductibilité de la méthode d'extraction. Les études du taux de recouvrement sont souvent réalisées en comparant les

résultats des échantillons extraits avec ceux d'échantillons enrichis d'une matrice similaire et/ou de blancs enrichis.

7.0 Analyse technique du domaine d'application

Des données adéquates sur la méthode d'essai devraient être disponibles pour les substances chimiques et/ou les produits représentatifs de ceux requis pour le CU correspondant à l'essai proposé. La documentation pour la NAM, l'AD ou l'IATA devrait décrire clairement les propriétés physicochimiques du domaine d'application. Cela comprend toute limitation de la méthode, par exemple pour les substances chimiques présentant un intervalle de poids moléculaire, une volatilité, une solubilité ou une stabilité spécifique.

Le domaine d'application devrait être évalué analytiquement lors du développement initial d'un essai, mais l'utilisation plus large de la méthode pourrait générer des données additionnelles. Ces données pourraient révéler des interférences susceptibles de nécessiter des mesures techniques supplémentaires (Petersen et al., 2022a). Si un essai n'est pas évalué pour des substances chimiques présentant certaines propriétés physicochimiques, la performance de l'essai sera plus incertaine pour lesdites substances, pour des raisons tant analytiques (p. ex. risque de biais ayant une incidence sur la performance de l'essai) que de concordance potentielle (p. ex. capacité de l'essai à donner des résultats semblables à ceux d'un essai *in vivo*). Par conséquent, il peut également être utile d'évaluer la concordance de l'essai pour diverses substances chimiques ayant des propriétés physicochimiques différentes afin d'accroître la confiance dans l'applicabilité de l'essai à un plus large éventail de substances chimiques.

8.0 Témoin positif

Le développement de l'essai devrait comprendre au moins un témoin positif. Ce type de témoin peut être utilisé conjointement avec des courbes standard ou en tant que concentration unique pour le témoin positif. Les témoins positifs devraient être adaptés au critère d'évaluation de détection de l'essai. La concentration du témoin positif devrait se situer dans la fenêtre de détection de l'essai. Souvent, la concentration du témoin positif est choisie de manière à obtenir une réponse comprise entre 50 % et 80 % de l'intervalle de détection maximal si une seule concentration est testée et que des données sur la résolution de la dose maximale sont disponibles. Une courbe standard est souvent produite en utilisant le témoin positif du système ; elle correspond à la relation entre la réponse à l'essai et des concentrations connues de l'analyte. La courbe standard devrait être reproductible dans le temps. Les courbes standard devraient être préparées avec le même excipient ou solvant que les échantillons à tester durant l'étude. Lorsqu'il est impossible d'obtenir le même mélange ou la même matrice, des mélanges ou matrices de substitution peuvent être utilisés, à condition de fournir une documentation et une justification appropriées.

9.0 Nomes de référence pour l'étalonnage des instruments

Le cas échéant, les instruments devraient être étalonnés à l'aide de normes d'étalonnage et/ou d'échantillons de contrôle qualité. La source de l'échantillon standard ou de contrôle qualité

devrait être documentée, en précisant la pureté, la stabilité, la source, le numéro de lot, le certificat d'analyse et la date de péremption. Il existe plusieurs types de normes, telles que les normes officinales de la Pharmacopée américaine et des matériaux fournis dans le commerce obtenus auprès d'une source commerciale réputée et dont la pureté est documentée.

Pour certains instruments (p. ex. spectromètres de masse), la norme de contrôle devrait idéalement être identique à l'analyte objet de l'essai. Souvent, cela n'est pas possible et une forme chimique établie de pureté connue peut être utilisée.

10.0 Établissement de spécifications

Des spécifications peuvent être définies pour les mesures de contrôle en cours de processus sur la base des résultats de tests intralaboratoires ou interlaboratoires. Plusieurs approches peuvent être utilisées. Les méthodes couramment utilisées appliquent ainsi un intervalle de confiance de 95 % aux mesures de contrôle en cours de processus ou utilisent la moyenne ± 2 ou 3 fois l'écart type, mais d'autres approches statistiques sont souvent utilisées. L'établissement de spécifications implique de trouver un équilibre entre différents objectifs. Il est important que les spécifications soient suffisamment strictes pour exclure les données indiquant que l'essai ne fonctionne pas comme prévu ou que les données d'une mesure de contrôle en cours de processus se situent dans un intervalle susceptible de fausser le résultat de la substance à tester. Il est également important que les spécifications ne soient pas strictes au point qu'un nombre trop élevé de séries d'essais donne des résultats hors spécifications alors que les données de test se situent dans un intervalle considéré comme normal.