Validation Study Report on Hand1-Luc EST (Embryonic Stem Cell Test) As a Non-animal Test Evaluating Developmental Toxicity of Chemicals

Prepared by
The validation management team of the Hand1-Luc EST
October, 2016

CONTENTS

Contents 2
List of abbreviations 7
Modular approach 9
Summary 10
1. Rational for the proposed test 12
1.1. Introduction 12
1.2. Objectives and intended use 14
1.3. Patent 15
2. Organization of the study 16
2.1. Validation management team 16
2.2. Participating laboratories 16
2.3. Sponsors 16
2.4. Supporters 17
2.5. Quality check monitoring 17
2.6. Meeting held 17
2.7. Chemical Selection 20
2.8. Chemical distribution management 20
2.9. Statistical analysis 21
3. Pre-validation study 22
3.1. Preliminary studies 22
3.1.1. Cell characterization 22
3.1.1.1. Objectives 22
3.1.1.2. Study design 22
3.1.1.3. Results 23
3.1.2. Protocol amendment 24
3.1.2.1. Objectives 24
3.1.2.2. Study design 25
3.1.2.3. Results 26
3.1.3. Conclusion 28
3.2. Phase 0 study 29
3.2.1. Goal 29
3.2.2. Study design 29
3.2.2.1. Protocol 29
3.2.2.2. Data analysis 29

3.2.3. Chemicals....29
3.2.4. Results....30
3.2.4.1. Lead laboratory....30
3.2.4.2. Lab A....30
3.2.4.3. Lab B....31
3.2.4.4. Lab C....31
3.2.5. Conclusion....32
3.3. Phase 1 study....32
3.3.1. Goal....32
3.3.2. Study design....32
3.3.2.1. Protocol....32
3.3.2.2. Data analysis....33
3.3.3. Chemicals....34
3.3.4. Results....34
3.3.4.1. IC_{50} and ID_{50} values variation and final judgment....34
3.3.4.2. Summary of the within-laboratory reproducibility....39
3.3.5. Conclusion....40
4. Validation study....41
4.1. Phase 2a study....41
4.1.1. Goal....41
4.1.2. Study design....41
4.1.2.1. Protocol....42
4.1.2.1.1. Preparation....43
4.1.2.1.2. Hand1-Luc EST assay....44
4.1.2.2. Prediction model....47
4.1.2.3. Positive and negative judgment....48
4.1.2.4. Acceptance criteria....49
4.1.3. Chemicals....50
4.1.4. Results....50
4.1.4.1. IC_{50} and ID_{50} values variation and final judgment....50
4.1.4.2. Summary of the between-laboratory reproducibility....54
4.1.5. Conclusion....54
4.2. Phase 2b/2c studies....54
4.2.1. Goal....54
4.2.2. Study design....55
4.2.3. Chemicals....55

4.2.4. Results....56
4.2.4.1. IC_{50} and ID_{50} values variation and final judgment....56
4.2.4.2. Summary of the between- and within-laboratory reproducibilities....65
4.2.5. Conclusion....66
4.3. Phase 3 study....66
4.3.1. Goal....66
4.3.2. Study design....66
4.3.2.1. Protocol....66
4.3.2.2. Data analysis....67
4.3.3. Chemicals....68
4.3.4. Results....69
4.3.4.1. IC_{50} and ID_{50} values variation and final judgment....69
4.3.4.2. Summary of the between-laboratory reproducibility....79
4.3.5. Conclusion....80
5. Discussion....82
5.1. Test definition....82
5.2. Definition of the performance standards....82
5.3. Predictive capacity....82
5.3.1. Prediction model....82
5.3.2. Predictive capacity....85
5.4. Transferability....85
5.5. Within-laboratory reproducibility....86
5.6. Between-laboratory reproducibility....86
5.7. Applicability domain....87
5.7.1. Limitations due to the material....87
5.7.2. Limitations due to the measured biological parameter....87
5.7.3. Justifications on misclassified chemicals....88
5.7.3.1. False negative chemicals....88
5.7.3.1.1. Boric acid....88
5.7.3.1.2. Dexamethasone....89
5.7.3.1.3. Dimethadione....89
5.7.3.1.4. Methoxyacetic acid....90
5.7.3.1.5. Phenytoin....90
5.7.3.2. False positive chemicals....91
5.7.3.2.1. Diphenhydramine....91
6. CONCLUSIONS....92

7. ACKNOWLEDGMENTS.....93
8. REFERENCES 94
APPENDIX
Appendix 1 Suzuki *et al*, 2011a
Appendix 2 Suzuki *et al*, 2011b
Appendix 3 Suzuki *et al*, 2012
Appendix 4 Le coz *et al*, 2015
Appendix 5 Nagahori *et al*, 2016
Appendix 6 Quality check sheet（Hand1-Luc Ph0 study)
Appendix 7 Quality check sheet（Hand1-Luc Ph1 study)
Appendix 8 Quality check sheet（Hand1-Luc Ph2a study)
Appendix 9 Quality check sheet（Hand1-Luc Ph2b study)
Appendix 10 Quality check sheet（Hand1-Luc Ph2c study)
Appendix 11 Quality check sheet（Hand1-Luc Ph3 study)
Appendix 12 Quality check statement for each validation phases
Appendix 13 Minutes of Hand1-Luc EST 1st VMT meeting
Appendix 14 Minutes of Hand1-Luc EST 2nd VMT meeting
Appendix 15 Minutes of Hand1-Luc EST 3rd VMT meeting
Appendix 16 Minutes of Hand1-Luc EST 4th VMT meeting
Appendix 17 Minutes of november 6th (2013) Hand-1
Appendix 18 Minutes of Hand1-Luc EST 5th VMT meeting
Appendix 19 Minutes of Hand1-Luc EST 6th VMT meeting
Appendix 20 Minutes of Hand1-Luc EST 6th VMT meeting conference call
Appendix 21 Minutes of Hand1-Luc EST 7th VMT meeting
Appendix 22 Minutes of Hand1-Luc EST 8th VMT meeting
Appendix 23 Minutes conference call 20150528
Appendix 24 Minutes of Hand1-Luc EST 9th VMT meeting
Appendix 25 Minutes conference call 20160406
Appendix 26 Chemical selection report
Appendix 27 Statistical analysis report
Appendix 28 Hand1protocol_ver04E
Appendix 29 Hand1protocol_ver05-1E
Appendix 30 from 6 days to 5 days measurement investigation and result
Appendix 31 Hand1-Luc EST Study plan
Appendix 32 Hand1protocol_ver06E

Appendix 33 Hand1_Data Sheet_verE08.3

Appendix 34 Hand1 protocol_ver07E

Appendix 35 Hand1 protocol_ver08E

Appendix 36 Hand1 71 chemicals data

Appendix 37 Hand1protocol_ver09E

Appendix 38 Hand1 related pathway references

List of abbreviations

AFC	Aminofluorocoumarin
AIST	The National Institute of Advanced Industrial Science and Technology
BG	Background
BfR	Bundesinstitut Riskobewertung(German Federal Institute for Risk Assessment)
DF	Dose finding study
DMSO	Dimethyl Sulfoxide
DS	Definitive study
ECVAM	European Centre for the Validation of Alternative Methods
EST	Embryonic Stem cell Test
EURL	European Union Reference Laboratory
FDSC	Food and Drug Safety Center
Hand1	Heart and neural derivatives 1
HRI	Hatano Research Institute
ICATM	International Cooperation on Alternative Test Methods
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
JaCVAM	Japanese Center for the Validation of Alternative Methods
JCIA	Japan Chemical Industry Association
KAO	Kao Corporation
KoCVAM	Korean Center for the Validation of Alternative Methods
LRI	Long Range Research Initiative
MC	Medium control
MD	Maximum dose
METI	Ministry of Economy Trade and Industry
NEDO	New Energy Development Organization
NICEATM	NTP Interagency Center for the Evaluation of Alternative Toxicological Methods
OECD	Organization for Economic Co-Operation and development
PBS	Phosphate Buffer Saline
RKI	Robert Koch Institute
ROC	Receiver Operating Characteristic
SC	Sumitomo Chemical Co., Ltd.
SCAS	Sumika Chemical Analysis Service, Ltd.
VC	Vehicle control

VMT	Validation Management Team
WEC	Whole Embryo Culture

Modular approach

In 2004, ECVAM proposed a modular approach to the validation of alternative methods (Hartung *et al*., 2004), according to which the various information requirements for peer-review and as generated during the validation process are broken down into seven independent modules as mentioned below:

Module 1: Test definition

Module 2: Within-laboratory reproducibility

Module 3: Transferability

Module 4: Between-laboratory reproducibility

Module 5: Predictive capacity

Module 6: Applicability domain

Module 7: Definition of performance standards

All the modules are summarized and discussed in part 5.

Summary

Risk assessment of developmental toxicity is essential for chemicals to be distributed on the market. Long established developmental toxicity assays with experimental animals are still used in laboratories because they are, with the current human knowledge, the most accurate test to predict developmental toxicity of chemicals. Since the end of the 20th century, regulatory authorities promote the development of alternative test methods to reduce the use of animals and the cost in terms of time and money induced by the current available protocols. The development of these methods is supervised by the ICATM, composed by worldwide associations (ICCVAM for the USA, ECVAM for Europe, JaCVAM for Japan, KoCVAM for Korea and Health up Canada).

In this context, the embryonic stem cell test (EST) had been developed to respond to the needs of the 3Rs (replace, refine, reduce) in 1997 by the group of Dr. Horst Spielmann. In this test, embryotoxicity of chemicals was evaluated by measuring cytotoxicity and differentiation toxicity using mouse ES D3 cells. Cytotoxicity was measured in both 3T3 cells and D3 cells with the MTT assay. Differentiation toxicity was analyzed by microscopically counting the beating of embryoid bodies after 10 days of culture. It was validated by ECVAM in 2002 (Genschow *et al*, 2004) but this protocol did not reach the status of an official OECD test guideline due to some limitations in the manipulations, in the applicability domain or in the reliability and relevance of the system (Buesen *et al,* 2004). Thus, Sumitomo Chemical Co., Ltd. has developed the Hand1-Luc EST to improve the original EST. Hand1 (Heart and neural crest derivatives 1) is a key gene for the development of the heart and some other organs (facial bones, limbs). The test measures both cytotoxicity and differentiation toxicity (Hand1 expression alteration by luciferase activity monitoring) with engineered mouse KOB1-ES cells being stable transformants of a vector containing the promoter of Hand1 gene upstream of luciferase reporter gene. The cells are differentiated in the presence of the chemical during 120 hr (5 days) into cardiomyocytes. This protocol is much easier than the original EST, by simply culturing cells in 96 well round bottom plate and measuring cytotoxicity and differentiation toxicity consequently in the same plate.

The validation study began in February 2013, lasted 3 years. It was applied in a step wise manner (pre-validation phase (phases 0 and 1) and validation phase (phases 2 and 3). The phase 0 consisted in the technical transfer to three different participant laboratories (Kao Corporation, SCAS, Ltd and HRI/FDSC). Each laboratory tested three chemicals with one dose finding study and three definitive studies. At this step, chemicals were incubated 6 days with the cells. The analysis of phase 0 led to improve the protocol, establish acceptance criteria and confirm the technical transfer. At this step, the further phase studies to verify within- and between-laboratory reproducibility were planned and the acceptance criterion for those two parameters was decided to be more than or equal to 75%. Between phase 0 and phase 1, the incubation time of chemicals with the cells has been changed from

6 days to 120 hr (5 days) so as to improve the IC_{50}/ID_{50} ratio that sometimes was lower than one. Phase 1 consisted in three coded chemicals tested once, each with one dose finding study and one definitive study to test the within-laboratory reproducibility. With a newly created prediction model based on the modification implemented at the end of phase 0, the result obtained was 88.9% that largely fulfilled the criterion (75%). A more precise analysis of phase 1 led to fit the curve (two parameter curve fitting) for the determination of the IC_{50} and ID_{50} values, to establish the majority judgment for the evaluation of embryotoxicity and the implementation of the solubility of the chemicals in the medium. The validation proceeded to phase 2 that was separated into three subphases 2a, 2b and 2c. Phase 2a was used to check the newly established improvements and to verify the between-laboratory reproducibility. The prediction model was improved according to the changes applied for phase 2 (acceptance criteria, two parameter curve fitting and improved prediction model). Phase 2a (four coded chemicals tested with one dose finding study and two or three definitive studies in only one set) gave very good results with a between-laboratory reproducibility of 100%. Thus, phases 2b and 2c were launched. Those phases consisted in eight coded chemicals tested three times to verify both between- and within-laboratory reproducibility. The results obtained were both within- (phase 2b and 2c results) and between- (phase 2a, 2b and 2c results) laboratory reproducibility of 83.3%. This phase ended with a new recommendation about fitting curves that should be done not with two parameters but with three parameter curve fitting. This change resulted in the improvement of the prediction model and phase 3 was started. This last phase consisted in sixteen chemicals tested only once with a dose finding study and two or three definitive studies. The results gave a between-laboratory reproducibility of 75% fulfilling the acceptance criterion. These results proved that Hand1-Luc EST is robust, transferable between laboratories, reproducible both within- and between-laboratories and applicable to the prediction of chemical embryotoxicity.

1 Rational for the proposed test

1.1 Introduction

In our society, consumers are more and more intensively exposed to chemicals produced by the industry. The different possible toxicities triggered by those compounds need to be accurately tested before putting chemicals on the market. In the context of 3R (Replacement, Reduction, Refinement), there is a high demand for a rapid, reliable and cost-effective alternative test methods. Developmental toxicity, due to complicated biological mechanisms that are at stake, is still taking its first faltering steps in *in vitro* experiments. Presently, only three *in vitro* developmental toxicity tests are validated by ECVAM: the Whole Embryonic Culture assay (WEC), the Micromass embryotoxicity assay and the Embryonic Stem cell Test (EST) but none of them reached the guideline status (Spielmann *et al*, 2006). Among those three tests, the WEC and the Micromass test although being recognized as alternative test methods are still demanding killing animals. Some say 1R (replacement) should be aimed instead of 3R and thus, those methods may not be sufficient. However, the EST, using mouse embryonic stem cells, does not require any animal killing which led it to be a test with high expectancies. The original test developed in 1997 (Spielmann *et al*, 1997) and validated by ECVAM in 2002 (Genschow *et al*, 2004) investigated the inhibition of cardiac differentiation by chemicals from ES cells within 10 days by microscopic observation of beating cardiomyocytes derived from mouse ES cells. The endpoints used for the prediction were IC_{50} of both mouse D3 ES cells and 3T3 cells, and ID_{50} of mouse ES cells. The IC_{50}, representing the concentration of chemicals for the one cell viability is reduced by 50%, was measured with the MTT assay. The ID_{50} was expressed as the concentration of chemicals inhibiting the development of contracting cardiomyocytes by 50% (number of well containing beating cardiomyocytes). The prediction model consisted in three different equations and the chemical was classified as three categories (strong, week and non embryotoxic). The accuracy of the test reached 78% (Genschow *et al*, 2004). The heart being the one of the main organs to be developing during embryogenesis, the EST was still a reliable tool to test embryotoxicity. However, it was not accepted as an OECD test guideline for the following reasons (Marx-Stoelting *et al*, 2009): a lack of clarity in discrimination between non- and weak/moderate embryotoxicants with a necessity of improving the prediction model, the applicability domain that should be widen with a integration of a metabolic system and the test being time consuming and requiring high experimental expertise.

To face those problems, the Hand1-EST was first developed by Sumitomo Chemical, Co. Ltd (SC) since 2011 supported by the NEDO (New Energy and industrial technology Development Organization) project. The first step of the development consisted in determining the genes involved in ES cell differentiation into cardiomyocytes. 120 marker genes were picked up (Suzuki *et al*, 2011a,

see appendix 1) by DNA chip and quantitative PCR. Among those, thirteen early embryogenesis expressed marker genes expression was analyzed after treatment with embryotoxicants and non-embryotoxicants. According to their role in the development process, their expression pattern and their altered expression after chemical exposure, the Hand1 gene was chosen. This gene has a key role for cardiac differentiation process (McFadden. *et al*., 2005; Reamon-Buettner *et al*., 2009). It is also involved in cranial morphogenesis (Firulli *et al*, 2014) and limbs formation (Laurie *et al*., 2016). The next step consisted in developing a novel high-throughput EST (Suzuki *et al*., 2011b, see appendix 2). In that study, the same type of cells was used as in the original EST (D3 and 3T3 cells). D3 cells were stably transfected with Hand1 promoter upstream of luciferase reporter gene, which allowed to see the influence of chemicals on the gene expression translating differentiation toxicity. The conclusion of the study was high predictability and accuracy with reduced test duration and manpower compared to the original EST. Furthermore, in 2012, another study was published (Suzuki *et al*., 2012, see appendix 3) in which Hand1-Luc EST was investigated to explore reproducibility by comparing a set of 6 well-known test chemicals in four different laboratories. Good correspondence was obtained among all four laboratories, proving the high potency of transferability, within-laboratory variability, and between-laboratory variability of the protocols. In 2016, the present Hand1-Luc EST was established (Le Coz *et al*, 2015; Nagahori *et al*, 2016, see appendix 4 and 5 respectively).

The test has been developed, after having analyzed the weak points of the original EST, to be a more accurate test with a wider applicability and with much easier manipulations. The test is using Hand1-ES (KOB1) cells. These cells are transfected with a vector containing the luciferase gene monitored by the Hand1 promoter. Thanks to the validation process, the test has been improved and a summary of the latest version of the protocol is presented below:

Hand1-ES (KOB1) cells are seeded in a 96 well round bottom plate. Chemicals are dissolved in PBS(-) or DMSO and their solubility in the assay medium is checked to determine the maximum dose (MD) of the experiment. Between 2 and 4 hrs after cell seeding, different concentrations of chemical are added to the wells. The plate is then incubated for 120 hr (5 days). After 120 hr (5 days) after the beginning of differentiation, cytotoxicity is measured with CellTiter-Fluor™ Cell Viability Assay and followed by measurement of Hand1 promoter activity through luciferase activity with Steady-Glo® Luciferase Assay System. Both endpoints are measured consecutively in the same plate (Figure1).

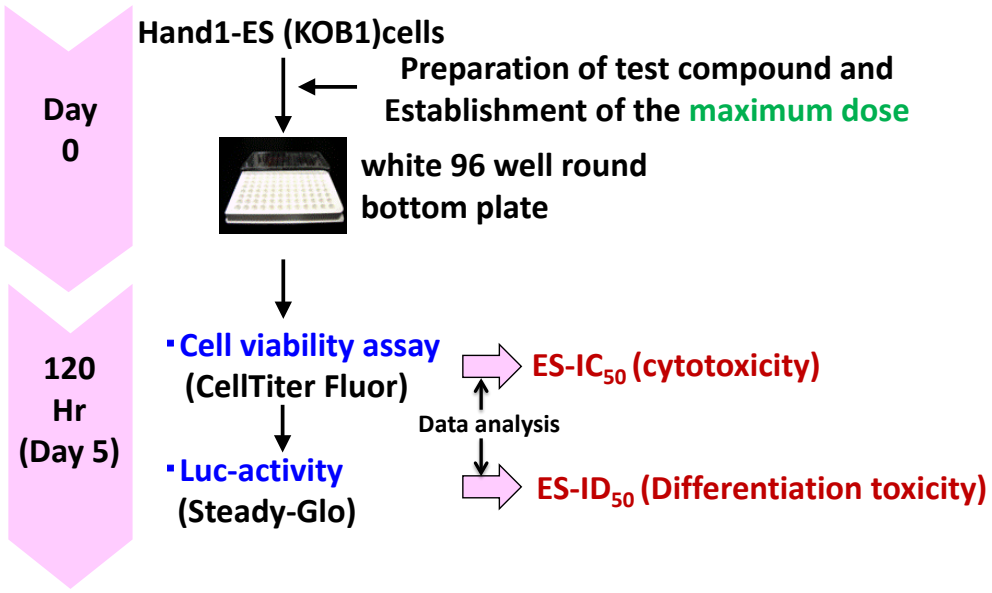


Figure 1 Experiment outline of Hand1-Luc EST

The raw data are then copied in an excel data sheet automatically calculating IC_{50} and ID_{50} results through a three parameter curve fitting. Then, the IC_{50}, the ID_{50} and the MD values are inserted in the prediction model which gives then the predicted positivity or negativity of the chemical.

Compared to the original EST, the Hand1-Luc EST is thus easier, requires low quantity of chemicals, has a wider applicability and considers solubility in the medium that could be considered to be close to the saturated free concentration in fetus that was observed in the reproductive toxicity studies. Finally, through the prediction model, the test allows to predict positive embryotoxicants with high confidence. This test is therefore a high-throughput test for eliminating true embryotoxicants and thus a very powerful tool for screening purposes.

The validation study of the Hand1-Luc EST was started from February 2013. It was applied in a step wise manner (pre-validation phase (phases 0 and 1) and validation phase (phases 2 and 3) in three different laboratories. Twenty eight different coded chemicals were used to verify the reproducibility of the test along the three phases.

The current report, which was prepared by SC, with the support of the VMT members, presents the validation process and the validation outcome of the Hand1-Luc EST.

1.2 Objectives and intended use

The aim of the Hand1-Luc EST validation study was to scientifically validate the Hand1-Luc EST in formal between-laboratory studies in order to assess the within- and the between-laboratory reproducibility of the assay, its transferability and its predictive capacity. Twenty-eight different chemicals with different mode of action were chosen to justify the relatively wide applicability

domain of the assay. The ultimate goal is to demonstrate its reliability for adoption as an OECD test guideline that could be used as a component of a test battery for the prediction of chemical developmental toxicity.

This report will also show the limitation of the test in terms of applicability by providing different biological concrete explanations leading to think that the test should be in the future combined to other tests to predict accurately developmental toxicity of chemicals.

1.3 Patent

The test method and the cells employed in this study are covered by Japanese Patent Number 5428527, and International Unexamined Patent Application Publication Number WO/2009/148177.

2 Organization of the study

2.1 Validation management team

The Validation management was composed of the following members:

Trial coordinator, VMT Chair	Noriho Tanaka (HRI/FDSC/ OTIP, Kanagawa, Japan)
Lead Laboratory	Koichi Saito, Noriyuki Suzuki, Hirohisa Nagahori, Florian Le Coz (SC, Osaka, Japan)
JaCVAM representative	Hajime Kojima (JaCVAM, NIHS, Tokyo, Japan)
Biostatisticians	Takashi Omori, Azusa Mori, Mayumi Kobayashi, Aoi Maruya (Faculty of Medicine of Kobe, Kobe University, Hyogo, Japan)
Consultant	Kazuhiko Matsumoto (Nagoya-City University, Nagoya, Japan)
In vivo teratogenicity expert	Makiko Kuwagata (HRI/FDSC, Kanagawa, Japan)
In vitro teratogenicity expert	Andrea Seiler (RKI, Berlin, Germany)
Quality Check monitoring	Yoshihiro Ohmiya (AIST, Ibaraki, Japan)
ICCVAM representative	David Allen (NICEATM/ICCVAM, North Carolina, USA)
ECVAM Representative	Michael Schaeffer (EURL ECVAM, Ispra, Italy)
KoCVAM Representative	Eui-Bae Jeung (KoCVAM, Chungcheongbuk-do, Korea)

2.2 Participating laboratories

The validation study was conducted in three participating laboratories as follows;

Laboratory 1: HRI/FDSC (Mika Watanabe, Mayu Ikezumi; Kanagawa, Japan)
Laboratory 2: SCAS (Kazunori Yanagi, Takeshi Izukawa; Osaka, Japan)
Laboratory 3: KAO (Toshio Kasamatsu, Naohiro Ikeda, Joshou Ryuu, Yuichi Ito; Tochigi, Japan)

2.3 Sponsors

Ministry of Economy, Trade and Industry (METI), Japan
Japan Chemical Industry Association (JCIA) Long-range Research Initiative (LRI)
Partially supported by National Institute of Advanced Industrial Science and Technology (AIST)

2.4 Supporters

National Institute of Health Sciences (NIHS)
Japanese Center for the Validation of Alternative Methods (JaCVAM)

2.5 Quality Check monitoring

The quality check sheet has been filled in by every laboratory before and during each experiment (see Appendix 6 to 11) to verify that the protocol was correctly respected.
Dr. Ohmiya was responsible for the QC monitoring during the validation study. All the statements submitted for each phases can be found in appendix 12.

2.6 Meetings held

The minutes of the VMT meetings described below can be found in appendix 13 to 25.

Title: 1st VMT meeting (21-22/02/2013, Osaka)
Members: Lori Rinckel (ICCVAM), Andrea Seiler (BfR), Eui-Bae Jeung (KoCVAM), Noriho Tanaka (HRI), Hajime Kojima (NIHS), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Kumiko Kobayashi, Sachiko Kitamoto (SC), Takashi Omori, Maruya Nishio, Aoi Maruya, Mayumi Kobayashi (Doshisha University), Makiko Kuwagata (HRI), Yoshihiro Ohmiya (AIST), Shojiro Yamazaki (HRI) and participating laboratories.
Subject: Discussion and planning of test protocol, judgment criteria and outline of validation study for Hand1-Luc EST
Minutes: Appendix 13

Title: 2nd VMT meeting with conference call (30/04/2013-01/05/2013, Kyoto)
Members: Lori Rinckel (ICCVAM), Andrea Seiler (BfR), Eui-Bae Jeung (KoCVAM), Micheal Schaeffer (ECVAM), Noriho Tanaka (HRI), Hajime Kojima (NIHS), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki (SC), Takashi Omori, Manabu Nishio, Aoi Maruya, Mayumi Kobayashi, Azusa Mori, Fuminari Yoshifuji (Doshisha University), Kazunori Yanagi (SCAS), Shojiro Yamazaki (HRI).
Subject: Justification on points raised in the 1st VMT meeting and planning of phase 0 study. Change of 2D to 3D cell culture, protocol improvement, two plates to one plate change
Minutes: Appendix 14

Title: 3rd VMT meeting with conference call (18-19/07/2013, Kyoto)

Members: Lori Rinckel (ICCVAM), Andrea Seiler (BfR), Hajime Kojima (NIHS), Noriho Tanaka (HRI), Yoshihiro Omiya (AIST), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Florian Le Coz (SC), Takeshi Omori, Aoi Maruya, Manabu Nishio, Mayumi Kobayashi, Azusa Mori, Kamoda, Takayana (Doshisha University), Kazunori Yanagi, Takeshi Izukawa (SCAS), Liu Josho (KAO), Mayu Ikezumi, Shojiro Yamazaki (HRI).

Subject: Improvement of the protocol. Discussion to proceed to phase 0.

Minutes: Appendix 15

Title: Lecture and training for technical transfer (19/07/2013, Osaka)

Members: Noriyuki Suzuki, Florian Le Coz (SC), Ikezumi Mayu (HRI), Liu Josho, Naohiro Ikeda (KAO), Takeshi Izukawa (SCAS)

Subject: Training of the participating laboratories for the protocol

Title: 4th VMT meeting with conference call (22-23/09/2013, Kyoto) followed by another teleconference call (06/11/2013)

Members: Lori Rinckel (ICCVAM), Andrea Seiler (BfR), Eui-Bae Jeung (KoCVAM), Micheal Schaeffer (ECVAM), Hajime Kojima (NIHS), Noriho Tanaka (HRI), Yoshihiro Omiya (AIST), Makiko Kuwagata (HRI), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Florian Le Coz (SC), Takeshi Omori, Aoi Maruya, Manabu Nishio, Mayumi Kobayashi, Azusa Mori, Kamoda, Takayana (Doshisha University), Kazunori Yanagi, Takeshi Izukawa (SCAS), Liu Josho, Ikeda (KAO), Mayu Ikezumi, Mika Watanabe, Shojiro Yamazaki (HRI).

Subject: Analysis of phase 0 data, improvement of the protocol and preparation for phase1.

Minutes: Appendix 16 and 17

Title: 5th VMT meeting with conference call (19-21/02/2014, Kyoto)

Members: Andrea Seiler (BfR), Eui-Bae Jeung (KoCVAM), Michael Schaeffer (ECVAM), Warren Casey (ICCVAM), Noriho Tanaka (HRI), Hajime Kojima (NIHS), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Florian Le Coz (SC), Takashi Omori, Mayumi Kobayashi, Azusa Mori, Manabu Nishio, Aoi Maruya (Doshisha University), Yoshihiro Ohmiya (AIST), Kazuhiko Matsumoto (Nagoya-City University), Makiko Kuwagata (HRI), Liu Shujie, Naohiro Ikeda (KAO), Kazunori Yanagi, Takeshi Izukawa (SCAS), Mayu Ikezumi, Mika Watanabe, Shojiro Yamazaki (HRI)

Subject: Analysis of phase 1 data, protocol improvements and preparation of phase 2

Minutes: Appendix 18

Title: 6th VMT meeting (10/05/2014, Kyoto)

Members: Noriho Tanaka (HRI), Hajime Kojima (NIHS), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Florian Le Coz (SC), Takashi Omori, Mayumi Kobayashi, Azusa Mori, (Doshisha University), Yoshihiro Ohmiya (AIST), Kazuhiko Matsumoto (Nagoya City University), Makiko Kuwagata (HRI), Liu Shujie, Naohiro Ikeda (KAO), Kazunori Yanagi, Takeshi Izukawa (SCAS), Mayu Ikezumi, Mika Watanabe, Shojiro Yamazaki (HRI), Takao Ashikaga

Subject: Analysis of phase 2a results and establishment of phase 2b, 2c plan.

Minutes: Appendix 19

Title: Conference call following the 6th VMT meeting (21/05/2014)

Members: Andrea Seiler (BfR), Eui-Bae Jeung (KoCVAM), Warren Casey, David Allen (ICCVAM), Noriho Tanaka (HRI), Hajime Kojima (NIHS), Makiko Kuwagata (HRI), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Florian Le Coz (SC), Shojiro Yamazaki (HRI)

Subject: Analysis of phase 2a and discussion for the promotion of phase 2b and 2c

Minutes: Appendix 20

Title: 7th VMT meeting (25/09/2014, Tokyo)

Members: Noriho Tanaka (HRI), Hajime Kojima (NIHS), Takashi Omori, Mayumi Kobayashi, Azusa Mori, Aoi Maruya (Doshisha University), Koichi Saito, Noriyuki Suzuki, Le Coz Florian (SC), Joshou Ryuu, Shota Nakagawa (KAO), Kazunori Yanagi, Takeshi Izukawa (SCAS), Mika Watanabe, Mayu Ikezumi, Shojiro Yamazaki (HRI)

Observers: Tomoka Hisaki(Shiseido Co. Ltd.), Tatsuya Mizukoshi(JCIA), Kohji Yamakage(HRI)

Subject: Analysis of phase 2b and discussion for phase 2c

Minutes: Appendix 21

Title: 8th VMT meeting with conference call (20-21/02/2015, Kyoto)

Members: Noriho Tanaka (HRI), Hajime Kojima (NIHS), Warren Casey (ICCVAM), Andrea Seiler (BfR), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Le Coz Florian (SC), Takashi Omori, Mayumi Kobayashi, Azusa Mori, Aoi Maruya (Doshisha University), Yoshihiro Ohmiya (AIST), Makiko Kuwagata (HRI), Kazuhiko Matsumoto (Nagoya-City University), Joshou Ryuu (KAO), Kazunori Yanagi, Takeshi Izukawa (SCAS), Mika Watanabe, Mayu Ikezumi, Shojiro Yamazaki (HRI)

Observers: Tomoka Hisaki (Shiseido Co. Ltd.), Tatsuya Mizukoshi (JCIA), Kohji Yamakage (HRI)

Subject: Analysis of phase 2 and preparation of phase 3 with protocol improvements

Minutes: Appendix 22

Title: Conference call (28/05/2015, Kyoto)

Members: Noriho Tanaka (HRI & OTIP, Chair), Hajime Kojima (JaCVAM,NIHS,Co-chair), David Allen (ICCVAM), Andrea Seiler (ZEBET), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Le Coz Florian (Sumitomo Chemical), Takashi Omori, Azusa Mori (Doshihsa university), Makiko Kuwagata (HRI), Shojiro Yamazaki (OTIP)

Subject: Discussion on 3 parameters curve fitting models for the phase 3 validation study

Minutes: Appendix 23

Title: 9th VMT meeting with conference call (19-20/02/2016, Kobe)

Members: Andrea Seiler (BfR), David Allen (ICCVAM), Noriho Tanaka (HRI), Hajime Kojima (NIHS), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Florian Le Coz (SC), Takashi Omori (Kobe University), Mayumi Kobayashi, Azusa Mori (Doshisha University), Eui-Bae Jeung (KoCVAM), Yoshihiro Ohmiya (AIST), Makiko Kuwagata (HRI), Kazuhiko Matsumoto (Nagoya-City University), Yuichi ito, Joshou Ryuu (KAO), Kazunori Yanagi, Takeshi Izukawa (SCAS), Mika Watanabe, Mayu Ikezumi, Shojiro Yamazaki (HRI)

Advisors: Horst Spielmann (Freie Universität Berlin), Shihori Tanabe (NIHS)

Observers: Akiko Tamura, Tomoka Hisaki (Shiseido Co. Ltd.)

Subject: Analysis of phase 3 and final decision on the validation

Minutes: Appendix 24

Title: Conference call (6/04/2016)

Members: Noriho Tanaka (HRI), Hajime Kojima (NIHS), David Allen (ICCVAM), Horst Spielmann (Freie Universität Berlin), Takashi Omori (Kobe University), Koichi Saito, Hirohisa Nagahori, Noriyuki Suzuki, Le Coz Florian (SC), Shojiro Yamazaki (HRI)

Subject: Comparison of prediction models and curve fitting: choice for the best combination and future schedule for the validation.

Minutes: Appendix 25

2.7 Chemical selection

The chemical selection was done under the recommendations of Dr Kuwagata. The details on the criteria can be found in Appendix 26.

2.8 Chemical distribution management

Coding and distribution of chemicals were performed by JaCVAM. The coded chemicals were sent

to the safety officer, who was not involved in the experiments, together with a sealed envelope containing the material safety data sheets (MSDS). Since the chemicals were coded, the laboratories did not know their identity and therefore all chemicals were treated as hazardous chemicals. The disclosure of codes was performed at a VMT meeting, after the data had been finalized per QC confirmation.

2.9 Statistical analysis

The details about the statistical analysis performed for all the phases can be found in Appendix 27.

3 Pre-validation study

This pre-validation study consisted in preliminary, phase 0 (transferability) and phase 1 studies.

3.1 Preliminary studies

Prior to the phase 0 study, the preliminary studies lasted 6 months (From February 2013 to July 2013) according to comments of the VMT members in the 1st and 2nd VMT meetings.

3.1.1 Cell characterization

3.1.1.1 Objectives

Basic concept of the Hand-Luc EST is originated from the NEDO (New energy and industrial technology development organization) EST project to quantify marker genes by luciferase reporter gene assays (Suzuki *et al*, 2011a,b). In the Hand-Luc EST, new engineered cells named Hand1-ES (KOB1) cells are used due to the limited use for mouse ES (D3) cells and international general use in the future. The objective of the study was to obtain background data (cell growth, induction of the luciferase gene and number of vector copy per cells) of the Hand1-ES (KOB1) cells to certify the quality control (QC) of the cells.

3.1.1.2 Study design

Cell growth and induction of the luciferase gene by Hand1 promoter

Three experiments without chemical treatment (medium control) have been done by two different researchers. The study design is illustrated in Figure 2. Materials and methods used in this study were basically described in the previous report (Suzuki *et al*, 2011b).

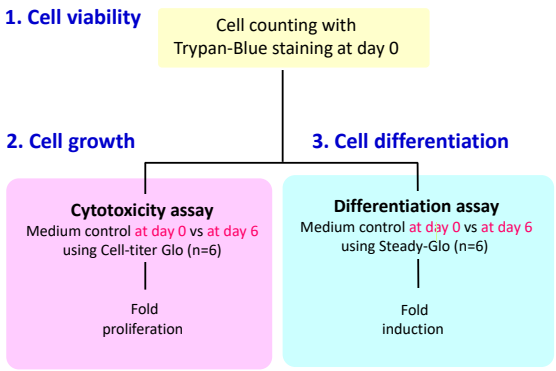


Figure 2 Study design to obtain background data of Hand1-ES (KOB1) cells

Copy number of vector containing luciferase gene governed by the Hand1 promoter

Copy numbers of exogenous DNA in Hand1-ES (KOB1) cells has been determined by Real-time PCR using SYBR® Green. Two samples were used:

- Plasmid (pGL4.17/Hand1-promoter Luc) for calibration curve
- Extracted genome DNA from Hand1-ES (KOB1) cells

3.1.1.3 Results

Cell growth and induction of the luciferase gene by Hand1 promoter

Cell viability at day 0 was more than 93% in all experiments (data not shown). The results of cell growth and luciferase gene inducement of Hand1-ES (KOB1) cells are summarized in Table 1.

In cytotoxicity assays, cell proliferation showed variations from 55 to 479 fold, proving the cells had been correctly growing between day 0 and day 6. The luciferase gene was also correctly induced (fold induction of luciferase activity varying from 1075 to 5726) in the differentiation assays. CV values were below 48% suggesting that the cytotoxicity and differentiation assays using Hand1-ES (KOB1) is stable and robust.

Table 1 Measurement of cell growth and luciferase gene inducement of Hand1-ES (KOB1) cells

Experiment		Cytotoxicity assay			Differentiation assay		
		Count Mean[a] ± SD	CV (%)	Fold	Count Mean[a] ± SD	CV (%)	Fold
Run 1[b]	Day 0	426259 ± 28881	7		235 ± 65	26	
	Day 6	23537940 ± 3910573	17	55	272338 ± 130170	48	1075
Run 2	Day 0	480807 ± 36579	8		163 ± 16	10	
	Day 6	47033807 ± 522981	1	98	269655 ± 61080	23	1653
Run 3	Day 0	516313 ± 68767	13		213 ± 37	18	
	Day 6	53997767 ± 916170	2	105	257849 ± 100480	39	1210
Run 1[c]	Day 0	85987 ± 17792	21		101 ± 8	8	
	Day 6	41197103 ± 406601	1	479	580269 ± 173219	30	5726
Run 2	Day 0	128140 ± 13354	10		122 ± 24	20	
	Day 6	41292740 ± 1444272	3	322	511333 ± 100871	20	4180
Run 3	Day 0	186093 ± 16252	9		110 ± 9	8	
	Day 6	42436440 ± 1651403	4	228	171247 ± 25452	15	1559

[a]Mean values of six wells (counts of luminescence).

[b,c] Data obtained by two independent researchers.

Copy number of vector containing luciferase gene governed by the Hand1 promoter

The ratio copy of plasmid and genome DNA sample gave 20517 copies. Furthermore, the genomic

DNA sample weighs 0.00387 µg and the mouse genome/cell weighs 5.479 x 10^{-12} g. By dividing the genomic DNA sample and the mouse genome per cell, a cell number of 706.3 was obtained. The number of copy found previously divided by the number of cells gave a number of 29 copies per cells.

3.1.2 Protocol amendment

The original protocol (Figure 3) was developed as part of the NEDO (New energy and industrial technology development organization) EST project to quantify marker genes by luciferase reporter gene assays (Suzuki *et al*, 2011a,b). In this protocol, cytotoxicity was measured with CellTiter-Glo® Luminescent Cell Viability Assay and Luciferase activity (differentiation toxicity) with Steady-Glo® Luciferase Assay System in separate plates.

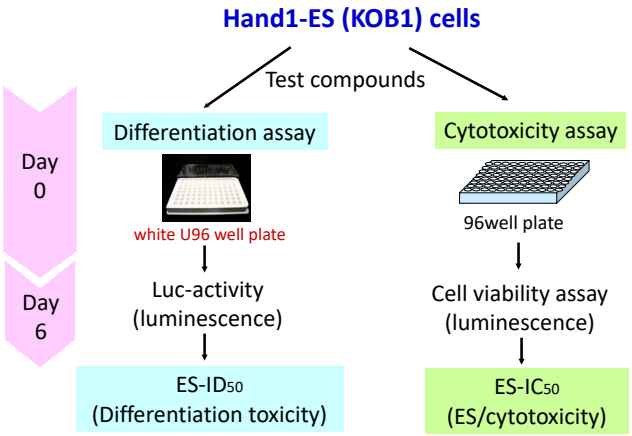


Figure 3 Schematic overview of the original Hand1-Luc EST protocol

The main point raised during the 2nd VMT meeting was to improve the way of culturing cells where it would be better to grow cells the same way (round bottom well plate to form embryoid bodies) for both cytotoxicity and differentiation toxicity. Measurements in the same plate would be ideal for not only scientific standpoints but also shortening of test duration and cost performance. Also, for some chemicals, the ID_{50} value was higher than the IC_{50} value which should not be because luciferase cannot be expressed if the cells are dead. Preliminary studies of SC showed that the improvement with measuring the cytotoxicity and differentiation toxicity in the same plate should solve the problem (data not shown).

Then amendment of the original protocol was proposed. Schematic overview of the new protocol is shown in Figure 4. To measure cytotoxicity, CellTiter-Fluor™ Cell Viability Assay was used since this reagent does not exhibit toxicity to the cells and thus, the next step with the measurement with

Steady-Glo® Luciferase Assay System could be operated.

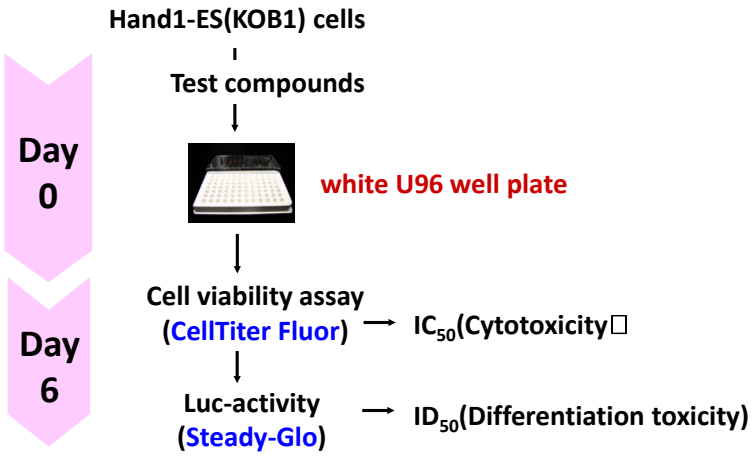


Figure 4 Schematic overview of the new Hand1-Luc EST protocol

3.1.2.1 Objectives

The objectives of the study was not only to prove reproducibility of the protocols but also to fix the protocol that would be used in the validation study by comparing the data from the original (Figure 3) and the new (Figure 4) protocols.

3.1.2.2 Study design

Twenty chemicals were tested three times independently by two different researchers in SC.
Variation of IC_{50} and ID_{50} values were evaluated.

Protocol

Two protocols (Figures 3 and 4) were used. Designs of the test plates in the original and new protocols are illustrated in Figure 5.

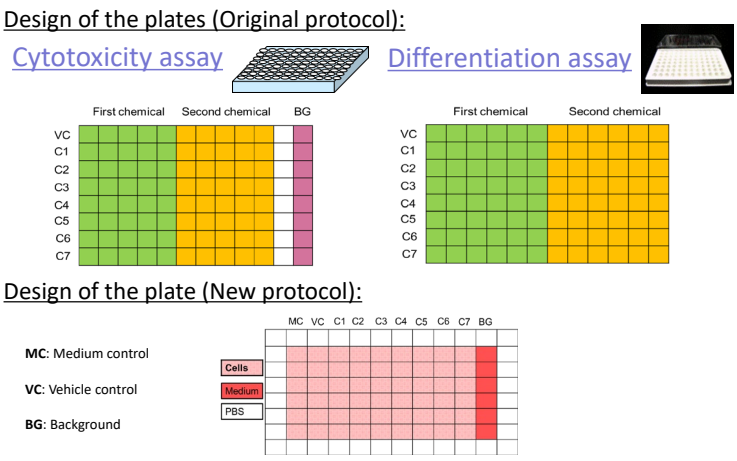


Figure 5 Design of test plates in the original and new protocols

Chemicals

Twenty chemicals were purchased from Sigma-Aldrich (Saint-Louis, MO, USA). All chemicals were reagent grades (purity> 97%). 5-Fluorouracil was used as a standard chemical in this study.

Data analysis

The IC_{50} and ID_{50} values were calculated using two experimental values at two tested concentrations across 50% activity rate in a concentration-response curve. Using the value of higher and lower sides of 50% of concentration and cell viability, a linear equation was created as follows (Kishi *et al.*, 2015).

IC_{50}=10^(log(A/B)×(50-C)/(D-C)+log(B))

A: the concentration of higher side of 50% of cell viability, B: the concentration of lower side of 50% of cell viability, C: cell viability at the concentration of B, D: cell viability at the concentration of A, ^: symbol of power in Excel software.

Similarly, suppression of mESCs differentiation by 50% was also calculated using the same formula described above.

ID_{50}=10^(log(A'/B')×(50-C')/(D'-C')+log(B'))

A': the concentration of higher side of 50% of luciferase activity, B': the concentration of lower side of 50% of luciferase activity, C': luciferase activity at the concentration of B', D': luciferase activity at the concentration of A', ^: symbol of power in Excel software.

3.1.2.3 Results

Data from twenty chemicals tested by the original protocol are summarized in Table 2.

The results show not only that the Hand1-Luc EST is reproducible (CV<54%, IC_{50} value for ascorbic acid) for both the parameters IC_{50} and ID_{50} but also that two different researchers obtain results with a low variability when performing three tests with the same chemical.

The results from the new protocol are shown in Table 3. When comparing the results to Table 2, very similar IC_{50} and ID_{50} values were found. Retinoic acid gave the highest CV (86%). The reason is that the IC_{50} and ID_{50} values are exceptionally low. CV values of the other chemicals are all under 44% showing the protocol is highly reproducible.

The relationship between the IC_{50} and ID_{50} was also improved especially for isoniazid, ascorbic acid, and acrylamide. Although for some cases ID_{50} values higher than the IC_{50} ones especially for cytosine arabinoside where the ratio the ID_{50}/IC_{50} is 1.4, the gap between the IC_{50} and the ID_{50} is considered as reasonable.

Table 2 Results of twenty chemicals tested by the original protocol (two plates)

In vivo category[a]	Test chemicals	IC_{50}		ID_{50}	
		Mean[b] ± SD (μg/ml)	CV(%)	Mean[b] ± SD (μg/ml)	CV(%)
Strong	5-Fluorouracil[c]	0.06 ± 0.01	8	0.06 ± 0.01	13
Strong	5-Fluorouracil[d]	0.06 ± 0.00	5	0.04 ± 0.00	3
Strong	Hydroxyurea[d]	5.69 ± 0.71	12	4.31 ± 0.16	4
Strong	6-Aminonicotinamide[d]	0.93 ± 0.18	19	1.12 ± 0.26	23
Strong	Cytosine arabinoside[d]	0.30 ± 0.00	7	0.06 ± 0.01	20
Strong	Methotrexate[d]	0.02 ± 0.00	20	0.02 ± 0.00	2
Strong	Retinoic acid[d]	0.0006 ± 0.0002	37	0.0003 ± 0.0001	46
Strong	5-Bromo-2'-deoxyuridine[d]	0.29 ± 0.06	21	0.43 ± 0.08	19
Weak	Valproic acid[c]	151.56 ± 34.41	23	134.69 ± 54.47	40
Weak	Dexamethasone[c]	32.22 ± 3.89	12	20.42 ± 1.66	8
Weak	Boric acid[c]	135.39 ± 1.91	1	52.59 ± 6.37	12
Weak	Methoxyacetic acid[c]	618.89 ± 24.89	4	480.34 ± 6.44	1
Weak	Lithium chloride[c]	777.46 ± 32.43	4	384.65 ± 145.29	38
Weak	Dimethadione[c]	>1000	-	>1000	-
Weak	Diphenylhydantoin(Free)[c]	47.75 ± 8.61	18	6.43 ± 0.58	9
Weak	Caffeine[c]	197.05 ± 30.37	15	83.45 ± 13.84	17
Non	Isoniazide[d]	99.86 ± 2.92	3	384.22 ± 44.39	12
Non	Ascorbic acid[d]	40.98 ± 22.07	54	164.04 ± 15.81	10
Non	Diphenhydramine HCl[d]	22.05 ± 3.00	14	15.60 ± 4.66	30
Non	Dimethyl phthalate[c]	>500	-	35.50 ± 2.71	8
Non	Acrylamide[c]	43.02 ± 6.40	15	71.60 ± 10.72	15

[a]Brown *et al*, 2002

[b]Mean values of three studies (µg/ml).

[c,d] Data from two independent researchers.

- Values for the ones the CV could not be determined

Table 3 Results of twenty chemicals tested by the new protocol (one plate)

In vivo category[a]	Test chemicals	IC_{50}		ID_{50}		Ratio ID_{50}/IC_{50}
		Mean[b] ± SD (µg/ml)	CV (%)	Mean[b] ± SD (µg/ml)	CV (%)	
Strong	5-Fluorouracil[c]	0.03 ± 0.002	8	0.04 ± 0.002	5	1.33
Strong	5-Fluorouracil[d]	0.03 ± 0.01	27	0.04 ± 0.01	16	1.33
Strong	Hydroxyurea[d]	5.61 ± 0.65	12	4.50 ± 0.45	10	0.80
Strong	6-Aminonicotinamide[d]	1.13 ± 0.03	3	0.96 ± 0.12	13	0.85
Strong	Cytosine arabinoside[d]	0.05 ± 0.002	3	0.07 ± 0.01	12	1.40
Strong	Methotrexate[d]	0.04 ± 0.01	21	0.03 ± 0.01	31	0.75
Strong	Retinoic acid[d]	0.005	-	0.00006±0.00005	86	0.12
Strong	5-Bromo-2'-deoxyuridine[d]	0.37 ± 0.06	15	0.41 ± 0.03	6	1.11
Weak	Valproic acid[c]	117.05 ± 26.71	23	95.40 ± 35.80	38	0.82
Weak	Dexamethasone[c]	26.15 ± 1.35	5	31.40 ± 5.18	17	1.20
Weak	Boric acid[c]	65.08 ± 6.42	10	63.28 ± 22.10	35	0.97
Weak	Methoxyacetic acid[c]	400.30 ± 12.33	3	475.85 ± 4.42	1	1.19
Weak	Lithium chloride[c]	>1000	-	950.18 ± 41.62	4	<0.95
Weak	Dimethadione[c]	>1000	-	>1000	-	(1.0)
Weak	Diphenylhydantoin[c]	>250	-	11.59 ± 2.43	21	<0.046
Weak	Caffeine[c]	199.56 ± 6.23	3	111.07 ± 25.06	23	0.56
Non	Isoniazide[d]	339.95 ± 59.59	18	410.91 ± 49.53	12	1.21
Non	Ascorbic acid[d]	634.27 ± 148.07	23	804.31 ± 182.36	23	1.23
Non	Diphenhydramine HCl[d]	21.05 ± 5.38	26	14.38 ± 5.67	39	0.68
Non	Dimethyl phthalate[c]	432.62 ± 58.80	14	40.89 ± 17.83	44	0.095
Non	Acrylamide[c]	116.76 ± 3.72	3	86.14 ± 2.57	3	0.74

[a] Brown *et al*, 2002

[b]Mean values of three studies (µg/ml).

[c,d] Data for two independent researchers.

- Cases where the CV could not be determined

3.1.3 Conclusion

Thanks to the analysis of the cells' characteristics and the usefulness of the protocol obtained the preliminary phase, the relevancy of the protocol was verified. Phase 0 was thus launched to verify

the transferability of the test by using the new protocol in three participating laboratories.

3.2 Phase 0 study

This phase began in July 2013 and ended in September 2013.

3.2.1 Goal

The goal of phase 0 was to prove that Hand1-Luc EST is transferable to other laboratories by analyzing the IC_{50} and ID_{50} variability of three non-coded chemicals.

3.2.2 Study design

The technical transfer was done in SC in Osaka with the three participant laboratories.

3.2.2.1 Protocol transfer

All the details of the protocol based on Figure 4 (one plate method) were taught by a SC's researcher. The protocol version used for this phase was 04E (see Appendix 28). For each chemical, three definitive studies (DS) were applied.

3.2.2.2 Data analysis

IC_{50} and ID_{50} values were calculated by the method described in **3.1.2.2 Study design** (Data analysis).

3.2.3 Chemicals

For this phase, the positive control 5-Fluorouracil (5-FU) and three non-coded test chemicals were tested (Table 4).

Table 4 Information about the test chemicals used for phase 0

Test chemicals	CAS No.	Supplier	Physical property	Solvent	*In vivo* classification
Ascorbic acid	50-81-7	Sigma	Solid	PBS	Negative
Boric acid	10043-35-3	Sigma	Solid	PBS	Positive
Hydroxyurea	127-07-1	Sigma	Solid	PBS	Positive

3.2.4 Results

3.2.4.1 Lead laboratory

Table 5 shows the background data obtained in the lab with the chemicals used for phase 0. The chemicals have been tested with three definitive studies and the variance remained very low (CV<35%) showing the test to be highly reproducible. The IC_{50} and ID_{50} values were the highest for ascorbic acid followed by boric acid and then by hydroxyurea .

Table 5 Background data on test chemicals used in phase 0 study (technical transfer confirmation study) in lead laboratory (SC)

Test chemicals	IC_{50}		ID_{50}	
	Mean[b] ± SD (μg/ml)	CV (%)	Mean[b] ± SD (μg/ml)	CV (%)
Ascorbic acid	634.27 ± 148.07	23	804.31 ± 182.36	23
Boric acid	65.08 ± 6.42	10	63.28 ± 22.10	35
Hydroxyurea	5.61 ± 0.65	12	4.50 ± 0.45	10

[a] Lead lab.: SC

[b]Mean values of three definitive studies.

3.2.4.2 Lab A

Table 6 shows the data obtained in phase 0 for lab A for three different test compounds. The IC_{50} and ID_{50} values obtained for the three chemicals tested in three definitive studies revealed a very low variance (CV<38%). IC_{50} and ID_{50} values differ from the lead lab from two to three times for ascorbic acid but for the two other chemicals the results are very close to the ones obtained in the lead lab. The IC_{50} and ID_{50} values were the highest for ascorbic acid followed by boric acid and then by hydroxyurea as observed in the lead lab.

Table 6 Results on Lab A[a] in phase 0 study (technical transfer confirmation study)

Test chemicals	IC_{50}		ID_{50}	
	Mean[b] ± SD (μg/ml)	CV (%)	Mean[b] ± SD (μg/ml)	CV (%)
Ascorbic acid	289.50 ± 76.54	26	251.29 ± 96.58	38
Boric acid	83.73 ± 10.36	12	92.07 ± 6.57	7
Hydroxyurea	4.30 ± 0.27	6	4.91 ± 0.43	9

[a] Lab A: FDSC

[b]Mean values of three definitive studies.

3.2.4.3 Lab B

Table 7 shows the data obtained in phase 0 for lab B for three different test compounds. The IC_{50} and ID_{50} values obtained for the three chemicals tested in three definitive studies revealed a very low variance (CV<28%). All of the data are very close to the ones obtained in the lead lab. The IC_{50} and ID_{50} values were the highest for ascorbic acid followed by boric acid and then by hydroxyurea as observed in lab A and the lead laboratory.

Table 7 Results on Lab B[a] in phase 0 study (technical transfer confirmation study)

Test chemicals	IC_{50}		ID_{50}	
	Mean[b] ± SD (μg/ml)	CV (%)	Mean[b] ± SD (μg/ml)	CV (%)
Ascorbic acid	552.29 ± 153.85	28	932.86 ± 116.26	12
Boric acid	58.80 ± 4.10	7	87.61 ± 6.51	7
Hydroxyurea	4.44 ± 0.33	7	3.47 ± 0.86	25

[a] Lab B: KAO

[b]Mean values of three definitive studies.

3.2.4.4 Lab C

Table 8 shows the data obtained in phase 0 for lab B for three different test compounds. The IC_{50} and ID_{50} values obtained for the boric acid and hydroxyurea tested in three definitive studies revealed a very low variance (CV<27%). However, the CV values obtained for ascorbic acid were much higher (73% for IC_{50} and 87% for ID_{50}) and the IC_{50} and ID_{50} values obtained were three to four times lower than the one obtained in the lead lab. All of the other data are very consistent and the IC_{50} and ID_{50} values are close to the ones obtained in the lead lab. The IC_{50} and ID_{50} values were the highest for ascorbic acid followed by boric acid and then by hydroxyurea as observed in lab A, B and the lead lab.

Table 8 Results on LabC[a] in phase 0 study (technical transfer confirmation study)

Test chemicals	IC_{50}		ID_{50}	
	Mean[b] ± SD (μg/ml)	CV (%)	Mean[b] ± SD (μg/ml)	CV (%)
Ascorbic acid	212.79 ± 155.62	73	178.78 ± 154.73	87
Boric acid	88.22 ± 3.33	4	109.35 ± 6.79	6
Hydroxyurea	5.32 ± 1.43	27	5.54 ± 0.45	8

[a] Lab C: SCAS

[b]Mean values of three definitive studies.

3.2.5 Conclusion

Except for ascorbic acid, the data of the two chemicals were close with a very low variance. Lab C, with ascorbic acid, obtained data with a bit higher variance and IC_{50} and ID_{50} values were a bit lower than the ones obtained in lab A and B. Given the fact that the historical data of the lead lab (see Table 2) for this chemical show a high variance, then it may be possible this variance is due to the intrinsic properties of the chemical.
Phase 0 showed a successful transferability of the protocol in the three participant labs with highly reproducible results obtained with the test and phase 1 was started.

3.3 Phase 1 study

Phase 1 study began in November 2013 and ended in March 2014.

3.3.1 Goal

The goal of phase 1 was to verify the within-laboratory reproducibility.

3.3.2 Study design

All chemicals were tested in three sets, a set consisting in one dose finding study (DF) and one definitive study (DS) to test the within-laboratory reproducibility of Hand1-Luc EST. The DF consists in testing the chemical with a 10 dilution ratio to find the range where both IC_{50} and ID_{50} are situated. The DS is then implemented with a lower dilution ratio within the range defined by the DF to improve accuracy of the results.

3.3.2.1 Protocol

The protocol version used for this phase was version 05-1E (see Appendix 29). The incubation time of the cells for differentiation was changed from 6 days to 120 hr (5 days). The reason for this change is explained in detailed in Appendix 30. The schematic overview is presented in Figure 6. After having obtained the IC_{50} and ID_{50} values, the positivity or the negativity of chemicals was calculated thanks to the prediction model detailed in part 3.3.2.2. The condition for the success of phase 1 is a between-laboratory reproducibility higher than 75% as fixed in the study plan (see Appendix 31).

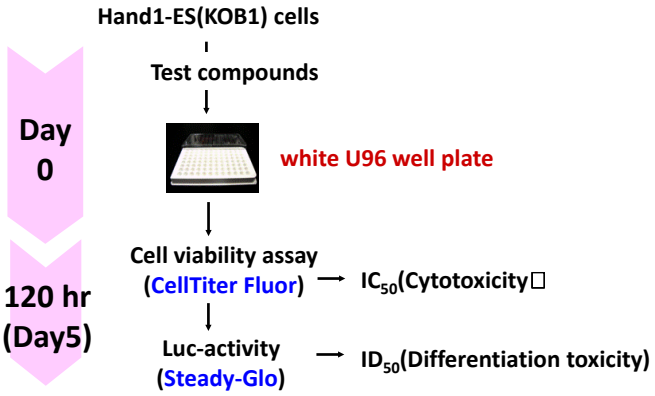


Figure 6 Schematic overview of the modified new protocol

3.3.2.2 Data analysis

IC_{50} and ID_{50} values were calculated by the method described in **3.1.2.2 Study design** (Data analysis).

The prediction model for phase 1 was constructed based on the data of twenty-one chemicals obtained in the lead lab (each chemical tested three times independently) (boric acid, carbamazepine, cefotaxime, chlorthalidone, cytosine arabinoside, dimethadione, diphenylhydantoin, glyburide, hydrochlorothiazide, indomethacin, isoniazid, lithium chloride, methotrexate, penicillin G, saccharin sodium salt, Sant-1, sulfasalazine, tomatidine hydrochloride, trazadone hydrochloride, valproic acid, warfarin) (Figure 7).

The general concept of the prediction consisted in using both IC_{50} and ID_{50} along with water solubility. The first two parameters (IC_{50} and ID_{50}) were determined by the Hand1-Luc EST and the water solubility by a free software called EPIsuite. Since the participating labs were not able to get the water solubility values (software not provided), those values were given by the lead lab to the biostatisticians so the calculation for prediction can be made. The model was based on two notions:

- The larger is the gap between the IC_{50} and the ID_{50}, the higher is the potential for the chemical to be embryotoxic (while the chemical does not kill the cells, there is an effect on Hand1 gene that would lead to embryotoxicity).

- The larger is the gap between the ID_{50} and the water solubility, the higher is the potential to be embryotoxic (if the chemical is highly soluble then it will have more chances to reach the fetus cells and have an effect).

Discriminant equation is as follows:

$$Score = 1.279 \times \log (IC_{50}/ID_{50}) + 0.33 \times \log (W_{Sol}/ID_{50}) - 0.67$$

Probability was calculated as an inverse logit transform of the score. When the probability shows more than 0.5, the chemical is evaluated as “Positive”. When the probability shows less than 0.5, the chemical is evaluated as “Negative”. The fifty percent line was decided because this was the best cut-off found to separate the *in vivo* positive and *in vivo* negative chemicals (Figure 7).

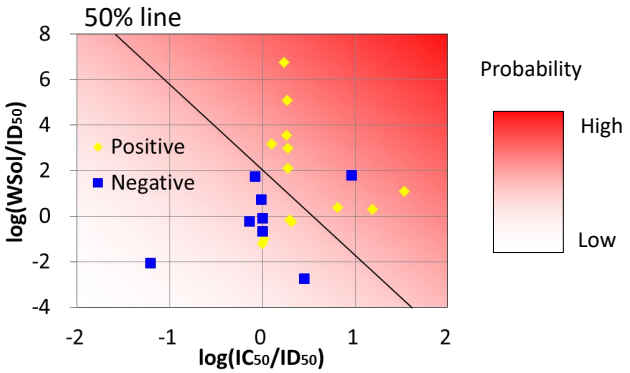


Figure 7　Development of equation for prediction model for phase 1

3.3.3　Chemicals

The chemicals used for phase 1 are presented in Table 9.

Table 9　Coded test chemicals used for phase 1 study

Set	Code No.			Test Chemicals	CAS No.	Supplier	Physical property	Solvent	*In vivo* classification
	SCAS	KAO	FDSC						
1	HA01	HB02	HC03	6-Aminonicotinamide	329-89-5	TCI	Solid	DMSO	Positive
	HA02	HB03	HC01	Isoniazid	54-85-3	Sigma(Fluka)	Solid	PBS	Negative
	HA03	HB01	HC02	5,5-Diphenylhydantoin	57-41-0	Wako	Solid	DMSO	Positive
2	HA06	HB04	HC05	6-Aminonicotinamide	329-89-5	TCI	Solid	DMSO	Positive
	HA04	HB05	HC06	Isoniazid	54-85-3	Sigma(Fluka)	Solid	PBS	Negative
	HA05	HB06	HC04	5,5-Diphenylhydantoin	57-41-0	Wako	Solid	DMSO	Positive
3	HA08	HB09	HC07	6-Aminonicotinamide	329-89-5	TCI	Solid	DMSO	Positive
	HA09	HB07	HC08	Isoniazid	54-85-3	Sigma(Fluka)	Solid	PBS	Negative
	HA07	HB08	HC09	5,5-Diphenylhydantoin	57-41-0	Wako	Solid	DMSO	Positive

3.3.4　Results

3.3.4.1　IC_{50} and ID_{50} values variation and final judgment

For each chemical, one dose finding study (DF) and one definitive study (DS) was applied.

6-Aminonicotinamide:

Table 10 shows the results of IC_{50} and ID_{50} values for 6-aminonicotinamide in phase 1. For each studies (dose finding and definitive), the values were very close to each other between the three sets. This is supported by figure 8 where no dispersion was observed. Furthermore, the judgment was the same for all the sets in every lab verifying again the within-laboratory reproducibility.

Table 10 Results of IC_{50} and ID_{50} values and final judgment obtained in each sets for 6-aminonicotinamide in phase 1

Lab	Set No.	IC_{50} values (μg/ml)		ID_{50} values (μg/ml)		Final Judgment
		Dose finding (DF)	Definitive study (DS)	Dose finding (DF)	Definitive study (DS)	
A	1	1.13	1.71	1.11	0.70	P
	2	1.91	1.28	1.87	1.01	P
	3	1.93	1.47	1.88	1.10	P
B	1	1.60	1.01	1.45	0.83	P
	2	1.27	1.24	0.80	0.91	P
	3	1.90	1.14	0.47	0.95	P
C	1	0.83	1.45	0.52	1.02	P
	2	1.07	0.89	0.42	0.70	P
	3	1.37	1.40	0.62	0.67	P

[a]Lab A: SCAS, Lab B: KAO, Lab C: FDSC

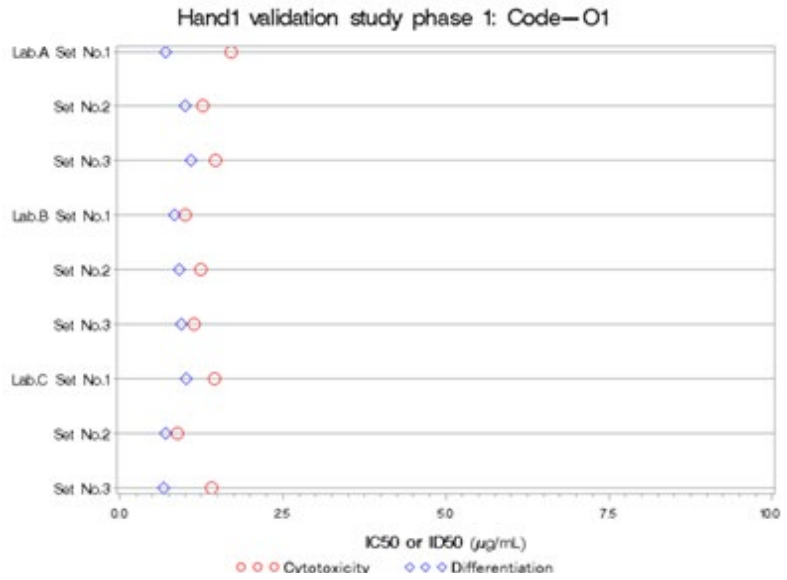


Figure 8 IC_{50} (circle) and ID_{50} (diamond) values obtained in the definitive study for 6-aminonicotinamide (Lab A: SCAS, Lab B: KAO, Lab C: FDSC)

Isoniazid:

Table 11 shows the results of IC_{50} and ID_{50} values for isoniazid in phase 1. The ID_{50} values of lab A were much lower compared to the values of labs B and C (Figure 9) which led to get a positive final judgment. However, the within-laboratory reproducibility was verified with each lab having the same results for each set.

Table 11 Results of IC_{50} and ID_{50} values and final judgment obtained in each sets for isoniazid in phase 1

Lab	Set No.	IC_{50} values (μg/ml)		ID_{50} values (μg/ml)		Final judgment
		Dose finding (DF)	Definitive study (DS)	Dose finding (DF)	Definitive study (DS)	
A	1	>500	291.36	28.90	37.99	P
	2	314.83	311.72	53.13	30.83	P
	3	365.15	564.78	135.56	<15.63	P
B	1	108.54	353.40	149.59	418.08	N
	2	207.74	354.36	190.42	374.32	N
	3	186.28	325.12	161.81	390.01	N
C	1	282.90	344.03	324.28	402.20	N
	2	254.14	315.63	414.69	394.64	N
	3	266.80	324.21	484.44	402.14	N

[a]Lab A: SCAS, Lab B: KAO, Lab C: FDSC

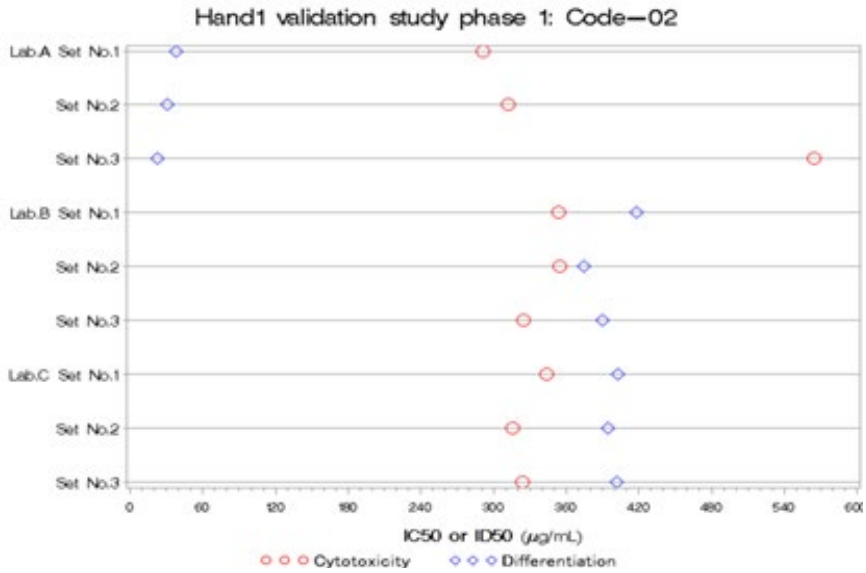


Figure 9 IC_{50} (circle) and ID_{50} (diamond) values obtained in the definitive study for isoniazid (Lab A: SCAS, Lab B: KAO, Lab C: FDSC).

As shown in Figure 9, the IC_{50} and the ID_{50} values did not vary between the sets in each laboratory verifying the within-laboratory reproducibility. In the case of lab A, the differentiation toxicity is different from the other labs, but the same type of curve was observed in each laboratory (Figure 10) supporting the success of the within-laboratory reproducibility. The reason why the values of lab A are much lower than the other labs could not be explained.

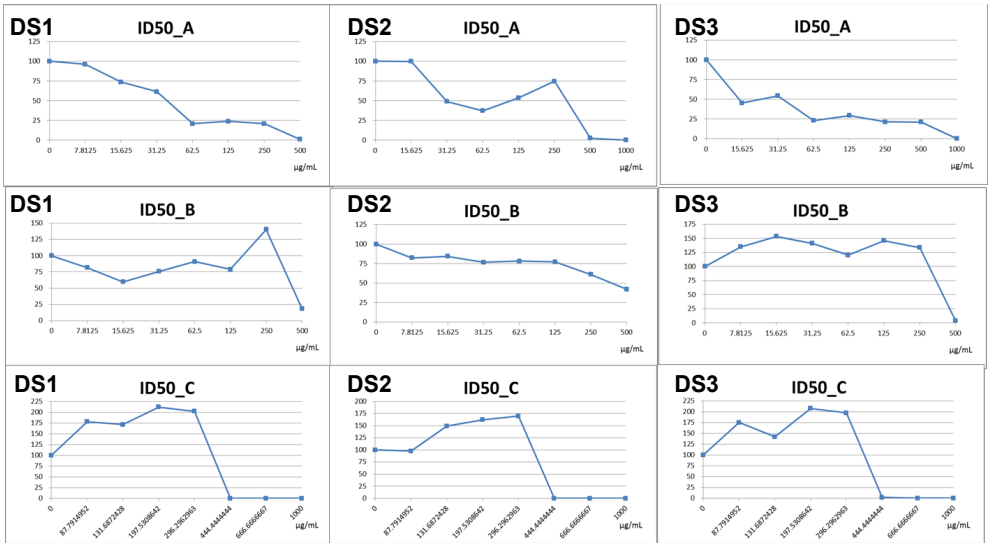


Figure 10 Differentiation toxicity curves obtained in each lab for three sets for isoniazid

5, 5-Diphenylhydantoin:

Table 12 shows the results of IC_{50} and ID_{50} values for 5, 5-diphenylhydantoin in phase 1. All the data within each lab are close to each other and do not vary. However, there are two cases where the IC_{50} value was very low compared to the other runs (see also Figure 11). This was observed for the 3rd set of Lab B (10.96 µg/ml) and the 1st set of lab C (34.95 µg/ml). These values are explained in figure 12. Indeed, the cytotoxicity curves, with a V shape, crossed two times the 50% line and thus, a very low IC_{50} value was triggered. When entering the data in the prediction model, a negative final judgment for the 3rd set of lab B was obtained because the gap between the IC_{50} and the ID_{50} was reduced. A positive result was still triggered for the 1st set of lab C because of a slight larger gap between the IC_{50} and ID_{50} values compared to the 3rd set of lab B. Although the curves generated very low IC_{50}, when looking at the shape of the curve, they are the same in each laboratory (Figure 12). This supports again the success of the within-laboratory reproducibility although the judgment between set 3 and sets 1/2 for Lab B was different.

Table 12　Results of IC_{50} and ID_{50} values and final judgment obtained in each sets for 5, 5-Diphenylhydantoin in phase 1

Lab	Set No.	IC_{50} values (μg/ml)		ID_{50} values (μg/ml)		Final judgment
		Dose finding (DF)	Definitive study (DS)	Dose finding (DF)	Definitive study (DS)	
A	1	>250	>250	6.79	2.46	P
	2	>250	>250	1.38	3.63	P
	3	>250	>250	8.25	2.37	P
B	1	>500	>500	27.03	32.72	P
	2	>500	>500	9.33	8.67	P
	3	>500	10.96	12.87	9.69	N
C	1	>500	34.85	5.48	8.59	P
	2	>500	>500	4.73	8.74	P
	3	>500	>500	8.42	4.98	P

[a]Lab A: SCAS, Lab B: KAO, Lab C: FDSC

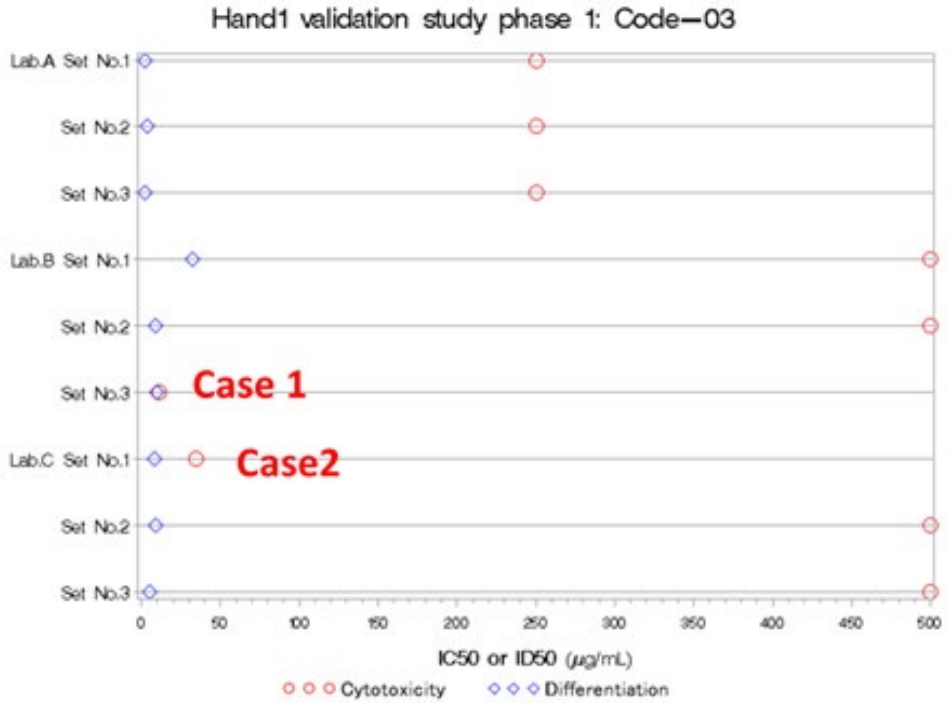


Figure 11　IC_{50} (circle) and ID_{50} (diamond) values obtained in the definitive study for 5, 5-diphenylhydantoin (Lab A: SCAS, Lab B: KAO, Lab C: FDSC).

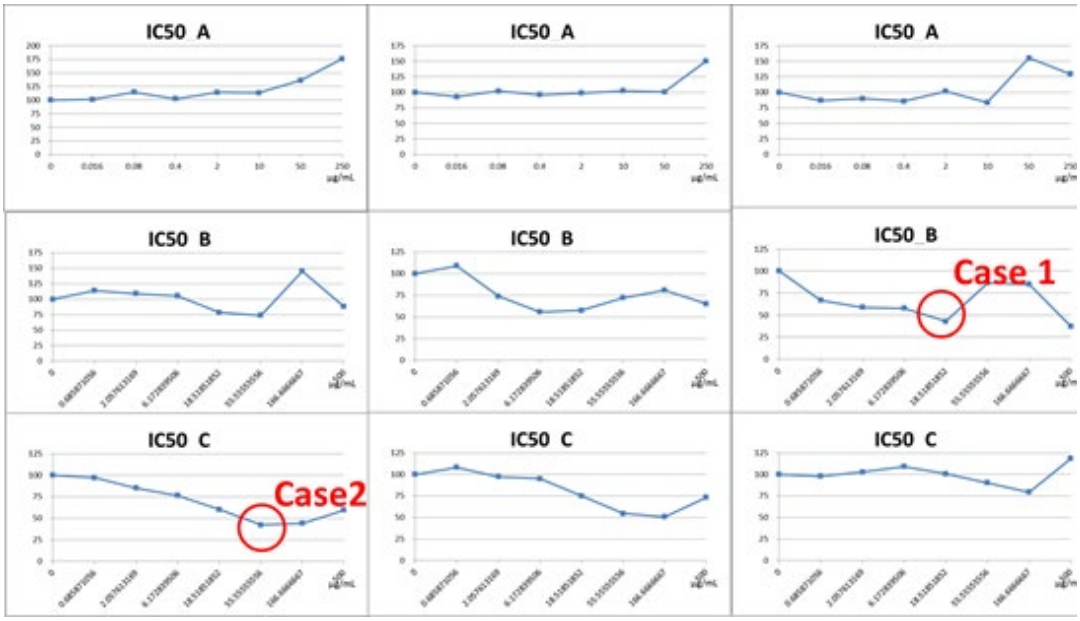


Figure 12　Cytotoxicity curves obtained in each lab for 3 sets with diphenylhydantoin (Lab A: SCAS, Lab B: KAO, Lab C: FDSC)

3.3.4.2　Summary of within-laboratory reproducibility

The summary of the final judgment is shown in Table 13.

Table 13　Summary of the judgment obtained for each chemical in each laboratory results of phase 1 study

Laboratory[a]	Chemicals	Set No.			Final judgment	*In vivo* classification
		1	2	3		
A	6-Amininononicotinamide	P	P	P	P	P
	Isoniazid	P	P	P	P	N
	5, 5-Diphenylhydantoin	P	P	P	P	P
B	6-Amininononicotinamide	P	P	P	P	P
	Isoniazid	N	N	N	N	N
	5, 5-Diphenylhydantoin	P	P	N	P	P
C	6-Amininononicotinamide	P	P	P	P	P
	Isoniazid	N	N	N	N	N
	5, 5-Diphenylhydantoin	P	P	P	P	P

[a]Lab A: SCAS, Lab B: KAO, Lab C: FDSC

The obtained within-laboratory reproducibility was 89% (8/9 chemicals).

The within-laboratory reproducibility of phase 1 satisfied then the criterion of 75% fixed in the study

plan (see Appendix 31). For two chemicals, Isoniazid and 6-aminonicotinamide all the labs obtained the same results for all the three sets. However, for 5,5 diphenylhydantoin, lab B did not satisfy the reproducibility (reason explained before).

3.3.5 Conclusion

All the results obtained in phase 1 show highly reproducible results. The cases where the results differ are due to curves with a V shape (Figure 12). Furthermore, there were cases where precipitation of the chemical at the highest concentration led to curves with shapes as shown in Figure 13 (next part). Those problems were discussed and solutions were provided for the next phase. The within-laboratory reproducibility (89%) of phase 1 satisfied the criterion of 75% fixed in the study plan (see Appendix 31). Given those good results, the validation study with phase 2 was set in motion.

4 Validation study

The validation study consisted in phases 2 and 3.
Phase 2 was separated into three subphases, phase 2a, 2b and 2c. Phase 2a began in March 2014 and ended in May 2014.

4.1 Phase 2a study

4.1.1 Goal

Phase 2a was established to verify the between-laboratory reproducibility with four coded chemicals.

4.1.2 Study design

For this phase, each chemical was tested in only one set composed of one dose finding study and two definitive studies. The dose finding study was taken into account for the judgment.
In phase 1, some curves had a V shape leading to truncated IC_{50} or ID_{50} value (see figure 12). Thus, the curves were fitted with a two parameter curve fitting to avoid this case. This fitting would also erase the problem of curve rising from the concentration 0 and then decreasing along with the increasing concentration of chemicals (see Figure 10, lab C). The prediction model was revised according to this improvement (see part 4.1.2.2).
Finally, during phase 1, there were some cases of chemical precipitation that truncated the curves (Figure 13).

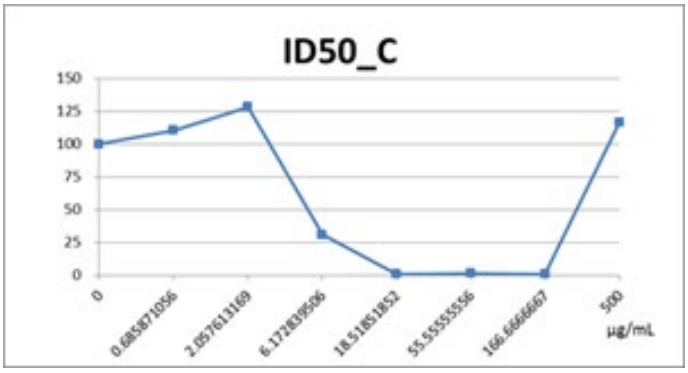


Figure 13 Differentiation toxicity curve obtained for 5, 5-diphenylhydantoin.

In this case, there are chances that the chemical precipitated on the tips or the chemical was not

correctly pipetted, leading to the concentration to be lower than expected triggering a non-toxic result at this concentration. To avoid precipitation, the protocol was improved by implementing the precipitation evaluation and the setting of the maximum dose (MD).

4.1.2.1 Protocol

Thanks to the pre-validation process, the protocol set until phase 1 has been improved. The protocol version used for this phase was the 06E version. The detailed version can be found in Appendix 32.
Hand1-Luc EST consists in using mouse ES cells stably transformed with a vector containing the luciferase gene governed by the Hand1 gene promoter and make them differentiate into cardiomyocytes for 120 hr (5 days). The chemical is added to the cells from the beginning of the differentiation (day 0). After 120 hr (5 days) of incubation, cytotoxicity (measured with CellTiter-Fluor™ Cell Viability Assay, IC_{50}) and differentiation toxicity (Hand1 gene expression inhibition measured with Steady-Glo® Luciferase Assay System through luciferase activity, ID_{50}) are measured. The highest concentration of chemical tested is the one for the chemical dissolves in the assay medium and is defined as the maximum dose (MD). IC_{50}, ID_{50} and MD are used for the prediction model to determine the positivity or the negativity of the chemical (see Figure 14).

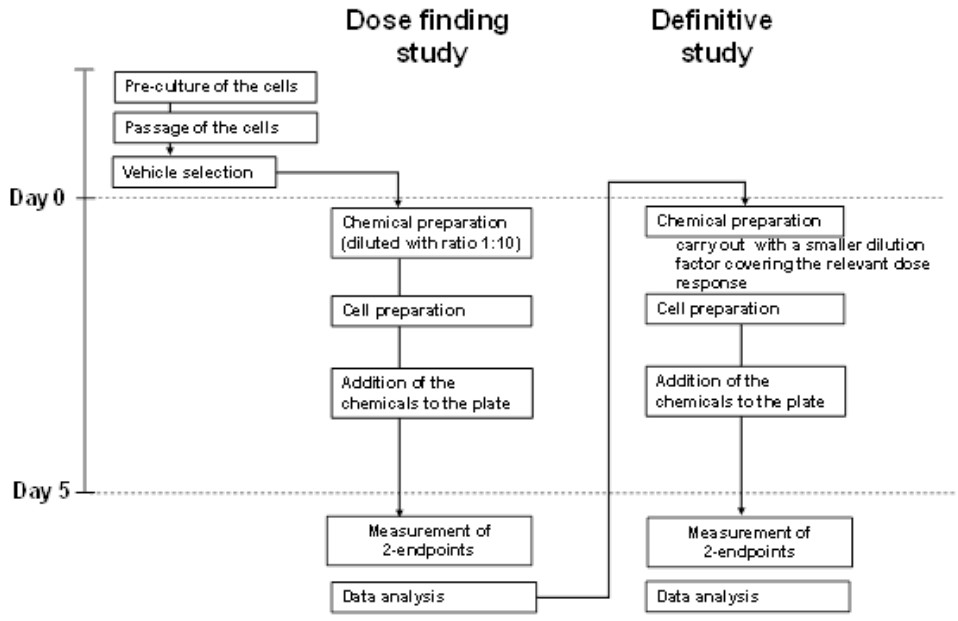


Figure 14 Outline of the Hand1-Luc EST

4.1.2.1.1 Preparation

Cells

Transgenic mouse embryonic stem (ES) cell, named as Hand1-ES (KOB1), containing firefly luciferase gene under the control of Hand1 (heart and neural crest derivatives expressed transcript 1) promoter were established by SC as a previous method (Suzuki *et al*., 2011b) with minor modifications. An embryonic stem cell line (named as KOB1-ES) from C57/B6 mouse established by SC was used for development of the Hand1-ES (KOB1) cells. For the validation, the cells used were provided by the lead lab from frozen tubes.

Materials

Two different media are used for the Hand1-Luc EST. The first one is the maintenance medium used for cell passage (Table 14) where the cells are maintained in an undifferentiated state. The second one is the assay medium (Table 15) used for the assay when the cells are differentiated into cardiomyocytes.

Table 14 Composition of the maintenance medium

Content	Volume
StemMedium	10 mL
100mM 2-ME solution	10 mL
ESGRO (10^7unit/mL)	1 mL
50 mg/mL G418 solution	20 mL

Table 15 Composition of the assay medium

Content	Volume
DMEM	82 mL
FBS (Heat inactivated)	15 mL
GlutaMaxI solution (100× conc.)	1 mL
100mM NEAA solution	1 mL
Penicillin/Streptomycin solution	1 mL
100mM 2-ME solution	100 mL

All of the necessary ingredients for the assay were distributed to the participant laboratories by the lead lab before each phase.

Materials are detailed in the protocol that can be found in Appendix 32.

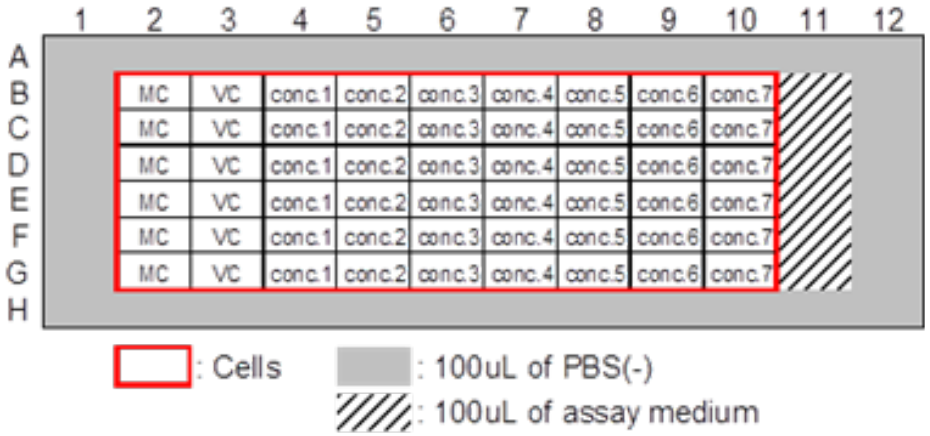
Maintenance of cells

Hand1-ES (KOB1) cells are thawed and seeded in 6 cm diameter gelatin coated plates and are passaged every 2 or 3 days. The day before each assay, cells are passaged one more time and must reach confluence the day of the experiment (80 to 90%). As a rule, cells must be used within two passages.

4.1.2.1.2 Hand1-Luc EST assay

Cell preparation

The assay is done in a 96 round bottom well white plate to allow cells to aggregate and form embryoid bodies autonomously. The design of the plate is shown in Figure 15.



MC: medium Control

VC: Vehicle Control

Column 11: BG: Background

Figure 15 Hand1-Luc EST's plate design

The cells (750 per well) are seeded. The cell viability must be higher or equal to 90%. MC consists in medium and cells only, VC consists in cells, assay medium and solvent (0.1% for DMSO and 1% for PBS) and BG consists in medium only. All those three columns are used for the final calculation to obtain IC_{50} and ID_{50} values but also to set acceptance criteria to evaluate the quality of the experiment. From column 10 to 4, a gradient of chemical is prepared according to the dilution ratio decided. Column 10 is the column with the highest chemical concentration and column 4 with the lowest.

Preparation of test chemicals and maximum dose (MD) setting

Chemicals are first dissolved in PBS or DMSO according to the step described in Figure 16.

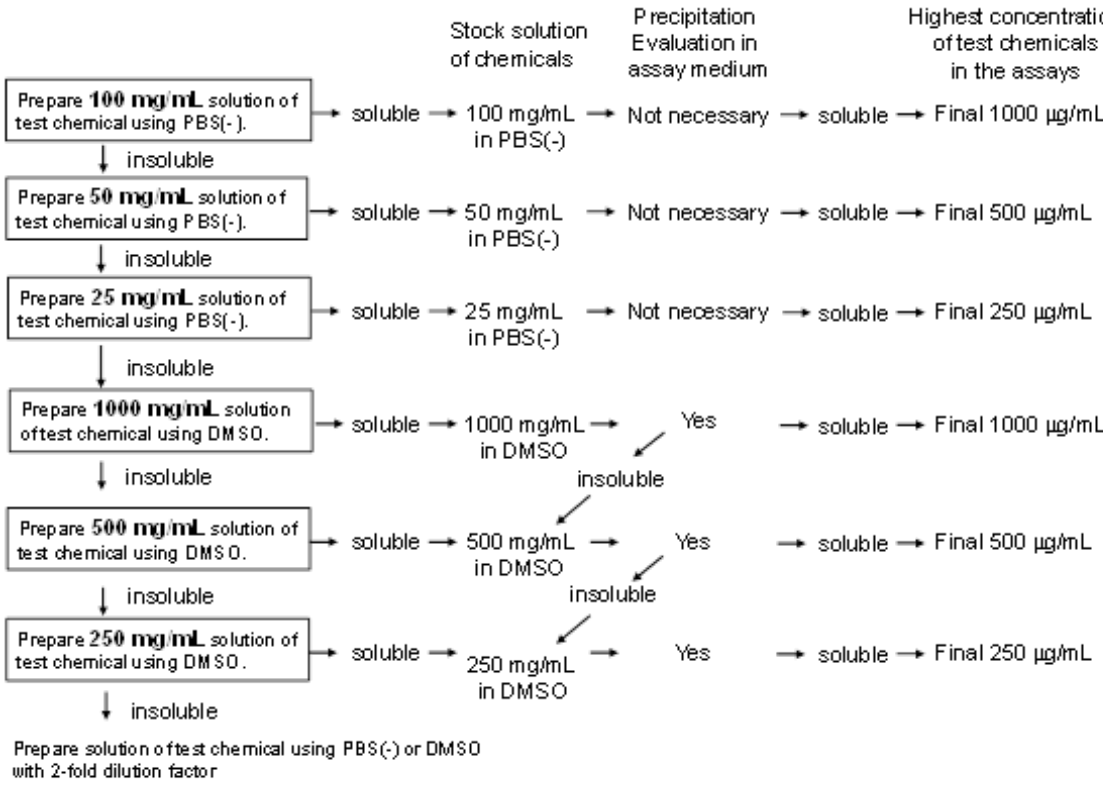


Figure 16 Vehicle selection and preparation of test chemicals

If the chemical does not dissolve in PBS(-) until a concentration of 25 mg/ml then DMSO is used to dissolve the chemical until complete dissolution. Once the chemical is dissolved, the precipitation test is done. It consists in adding 1 µl of chemical in 999 µl of assay medium and evaluate if precipitation occurs. If not, the concentration is determined as the maximum one tested for the assay. If it precipitates, then the concentration is divided by 2 in the solvent and the precipitation evaluation in assay medium is done again. This is repeated until the chemical dissolves. All the steps are described in Figure 16. The maximum dose (MD) is therefore defined as the highest dissolvable concentration.

Chemical dilution for dose finding and definitive studies

Chemicals, after being dissolved in the appropriate solvent are diluted in a 96 well storage round bottom plate from the maximum dose determined as described previously. The chemicals are added in the assay medium (20 µl in 980 µl of assay medium if the chemical is dissolved in PBS (-) (1% v/v) and 2 µl in 998 µl of assay medium if the chemical is dissolved in DMSO (0.1%)). For the dose finding study, used to determine the concentration range for the one both IC_{50} and ID_{50} are

observable, a 10 dilution ratio is systematically applied.
After obtaining the results, the dilution ratio is lowered to get more accurate values of the IC_{50} and ID_{50}.

Measurement of cytotoxicity and differentiation toxicity

Chemicals can be added to the cells between 2 hrs and 4 hrs after seeding (50 µl of medium containing chemicals added to the already seeded 50 µl of cells. The final volume in each well is 100 µl/well). The cells are incubated for 120 hr (5 days) at 37ºC, 5% CO_2.
After 120 hr (5 days) of incubation, cytotoxicity is measured followed by differentiation toxicity.
Cytotoxicity is expressed as the concentration of chemicals that reduced the viability of cells to 50% of the control level against the vehicle control. The reagent used to determine the value is CellTiter-Fluor® Cell Viability Assay. The principle of this product is the following: The cell-permeant substrate enters the cell, where it is cleaved by the live-cell protease activity to produce the fluorescent AFC. The live-cell protease is labile in membrane-compromised cells and cannot cleave the substrate. After having determined cytotoxicity, differentiation toxicity is measured in the same plate by adding 100 µl of Steady-Glo® Luciferase Assay System. This product contains beetle luciferin which is cleaved by the firefly luciferase express by the cells which produces oxyluciferin being luminescent. The inhibition of differentiation was expressed as the concentration of the test chemical that reduces the luminescence by 50%.

Estimation of IC_{50} and ID_{50} values

The algorithm for the estimation of IC_{50} and ID_{50} values was implemented in a developed Microsoft Excel data sheet through all phases. An example of the data sheet used in this validation study is shown in Appendix 33. The data sheet files in all phases remain separated into three parts: the face cover sheet, the calculation sheets and the summary sheet. Once a user enters data into the face cover sheet, the IC_{50} and ID_{50} values are estimated with the concentration-response plots and the acceptance criteria are also calculated in the calculation sheets. Finally, the summarized results appear in the summary sheet. IC_{50} and ID_{50} values were estimated by fitting the curve to these data (two parameter fitting)

Two parameter curve fitting:

Phase 1 showed some cases where the IC_{50} was lower than the ID_{50}. Thus, to correct all those problems, a curve fitting method was proposed by VMT members.
Two parameter curve fitting was implemented in the calculation sheet by Dr. Omori, biostatistician. Figure 17 shows that the new IC_{50} and ID_{50} obtained after curve fitting is not significantly different from the non-fitted data. Furthermore, thanks to this fitting, the IC_{50} and ID_{50} relationship was

improved and the cases where the curve crossed the 50 % line (V curve) were no longer possible.

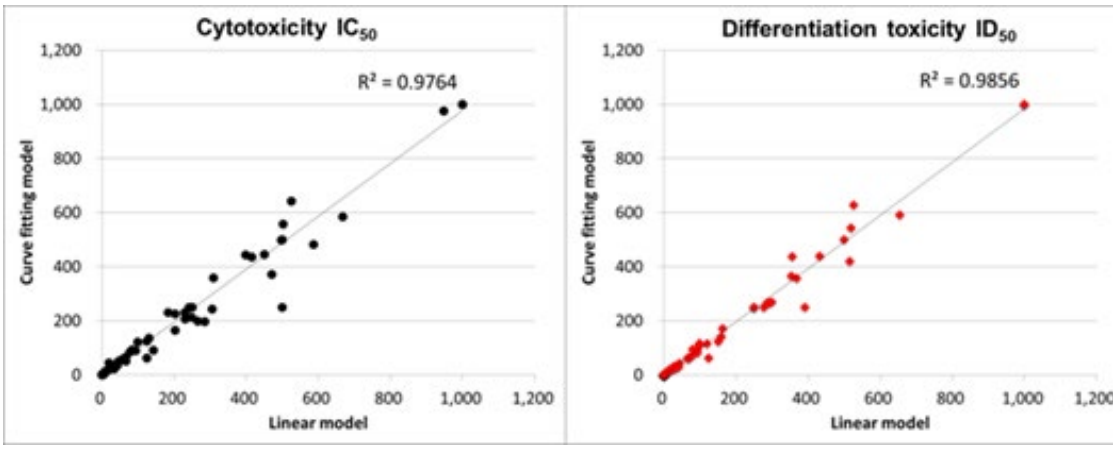


Figure 17 Plotting of the IC_{50} and ID_{50} value obtained in both way of calculating (with or without curve fitting)

4.1.2.2 Prediction model

In addition to the implementation of the curve fitting, the evaluation of precipitation was included in the protocol. Those two significant changes led to the revision of the prediction model because the IC_{50} and ID_{50} values are slightly different and the MD (maximum dose) is used for the model. The Figure 18 shows the new prediction model which was established by the analysis of the previously used twenty-one chemicals data revised with their respective MD, IC_{50} and ID_{50} values with the new two parameters curve fitting. The equation is as follows:

$$\text{Score} = 1.252 \times \log(IC_{50}/ID_{50}) + 0.548 \times \log(MD/ID_{50}) - 0.805$$

And the probability is obtained by applying the inverse logit to the score.

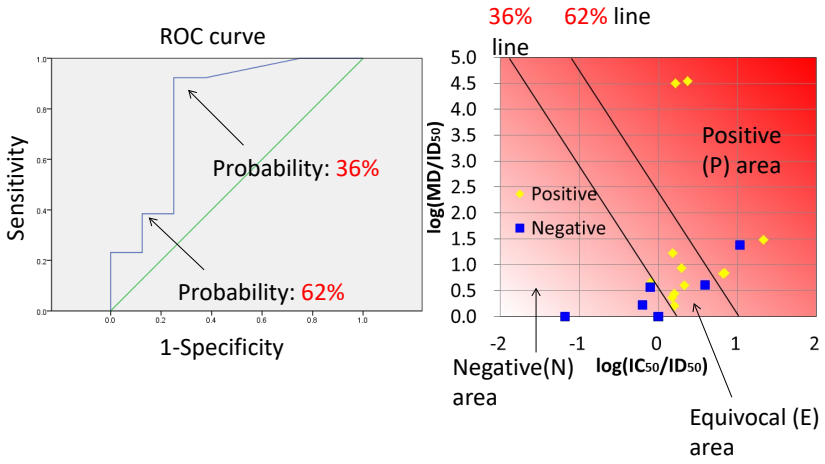


Figure 18 Development of equation for prediction model for phase 2a

The ROC curve showed cut-off criteria at 36% and 62% and the chemical with a probability lower than 0.36 are considered negative. Chemicals with a probability situated between 0.36 and 0.62 are considered as equivocal and the chemicals with a probability higher than 0.62 are considered positive. After discussion about the prediction model, the chemicals situated in the equivocal area were determined as positive because this led to a higher predictive power for positive chemicals and a better accuracy when comparing to the *in vivo* results.

4.1.2.3 Positive and negative judgment

The positivity or the negativity of the chemical was decided with one dose finding study (DF) and 2 definitive studies (DS). The dose finding study result was used for the final judgment as shown in the table below (Table 16).

Table 16 Majority judgment and number of studies needed according to the different possibilities for phase 1

	Dose finding study	Definitive study 1	Definitive study 2	Final judgment
Case 1	N	N	N	N
Case 2	N	N	P	N
Case 3	N	P	N	N
Case 4	P	N	N	N
Case 5	P	P	P	P
Case 6	P	P	N	P
Case 7	P	N	P	P
Case 8	N	P	P	P

4.1.2.4 Acceptance criteria

Acceptance criteria were established for i) quality control of cell condition, ii) quality control of differentiation and cytotoxicity assays, iii) performance standard of the assay and iv) quality control for the effect of the vehicle (PBS(-) or DMSO). The criteria described in the protocol are the followings:

Quality control

i) Quality control of cell condition

Check viability of the cells by staining an aliquot of the cell suspension with Trypan blue. A viability of 90% or greater is acceptable.

ii) Quality control of differentiation and cytotoxicity assays

To verify the cell growth and cell differentiation, the comparison of the medium control (MC) and background (BG) is used.

- The lower limit of the 95% confidence intervals of the ratio of MC / BG should be above 1 for the cytotoxicity assay
- The lower limit of the 95% confidence intervals of the ratio of MC / BG should be above 10 for the differentiation assay

iii) Performance standard of the assay

The quality of the assay must be controlled using 5-FU as a positive-reference chemical.

- The range of ID_{50} for 5-FU should be within 0.003 and 0.067 μg/mL (according to phase 0 study).
- The range of IC_{50} for 5-FU should be within 0.003 and 0.065 μg/mL (according to phase 0 study).

iv) Quality control for effect of vehicle

To verify the effect of vehicles, the comparison of the medium control (MC) and vehicle control (VC) is used.

- The lower limit of the 95% confidence intervals of the ratio of VC / MC should be above 0.2 for the cytotoxicity and differentiation assays.
- In addition to the above criterion
  CV (SD/mean x 100%) of VC should be below 100%.

4.1.3. Chemicals

Information about the coded chemicals used for phase 2a is shown in Table 17.

Table 17 Coded chemicals used for phase 2a

No	Code No.			Chemical name	CAS No.	Supplier	Physical property	Solvent	*In vivo* classification
	FDSC	SCAS	KAO						
1	HA11	HB19	HC27	Cytosine arabinoside	69-74-9	Sigma	Solid	PBS	Positive
2	HA12	HB20	HC28	Lithium Chloride	7447-41-8	Wako	Solid	PBS	Positive
3	HA13	HB21	HC29	Cefotaxime sodium salt	64485-93-4	TCI	Solid	PBS	Negative
4	HA14	HB22	HC30	Penicillin G	69-57-8	TCI	Solid	PBS	Negative

4.1.4 Results

4.1.4.1 IC_{50} and ID_{50} values variation and final judgment

The results are presented for each chemical.

Cytosine arabinoside

Table 18 shows the results obtained with cytosine arabinoside during phase 2a.

Table 18 Results obtained with cytosine arabinoside during phase 2a showing the IC_{50} and ID_{50} values with the final judgment obtained for the DF and the DS studies

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Cytosine arabinoside	A	DF	0.02	0.03	P	P
		DS1	0.03	0.04	P	
		DS2	0.05	0.05	P	
	B	DF	0.06	0.03	P	P
		DS1	0.44	0.71	P	
		DS2	0.41	0.43	P	
	C	DF	0.13	0.03	P	P
		DS1	0.03	0.03	P	
		DS2	0.04	0.04	P	

The IC_{50} and ID_{50} values are close to each other in each laboratory. However, higher values for lab B were found, around 10 times higher than the other labs. Each lab obtained a positive result for the chemical showing the between-laboratory reproducibility to be verified.

Lithium Chloride:

Table 19 shows the results obtained with lithium chloride during phase 2a.

Table 19 Results obtained with lithium chloride during phase 2a showing the IC_{50} and ID_{50} values with the final judgment obtained for the DF and the DS studies

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Lithium chloride	A	DF	106.12	246.71	N	P
		DS1	569.71	454.90	P(E)[a]	
		DS2	522.23	455.71	P(E)[a]	
	B	DF	214.87	252.91	P(E)[a]	P
		DS1	640.11	295.99	P(E)[a]	
		DS2	450.04	276.00	P(E)[a]	
	C	DF	459.26	228.60	P(E)[a]	P
		DS1	493.04	318.69	P(E)[a]	
		DS2	471.03	310.24	P(E)[a]	

[a]P(E) is used to show the chemical that is in the equivocal area (see Figure 19) but is considered as positive.

As shown in Table 19, all the labs for each run obtained the same positive judgment except for DF of Lab A where the IC_{50} value (106.12 µg/ml) is lower than the values obtained in the other labs for the DF studies. The difference is due to the fact that those dose finding studies are done with a 10

dilution ratio (to find the range of the IC_{50} and ID_{50}). This triggers less accurate results than the DS studies and thus tends to show variation. The curves obtained have the same shape for all the labs and thus the between-laboratory reproducibility has been verified (see Figure 19).

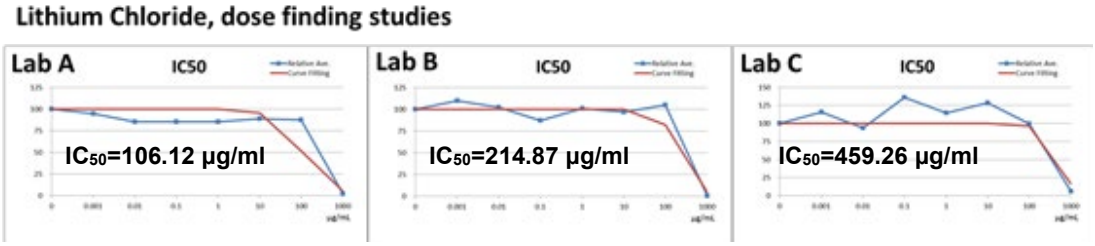


Figure 19 Dose finding studies in the three participant laboratories for lithium chloride

Cefotaxime sodium salt

Table 20 shows the results obtained with cefotaxime sodium salt during phase 2a.

Table 20 Results obtained with cefotaxime sodium salt during phase 2a showing the IC_{50} and ID_{50} values with the final judgment obtained for the DF and the DS studies

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Cefotaxime sodium salt	A	DF	55.64	228.56	N	N
		DS1	256.18	332.07	N	
		DS2	280.23	320.37	N	
	B	DF	51.77	358.95	N	N
		DS1	353.57	570.78	N	
		DS2	342.07	621.18	N	
	C	DF	155.03	294.18	N	N
		DS1	221.23	521.54	N	
		DS2	49.95	370.29	N	

The definitive studies for cefotaxime sodium salt reveal high reproducibility between the labs. However, the DS2 of lab C the IC_{50} value was much lower. This is explained in Figure 21.

At the concentration 2 of chemical, the counts in the very lower well (743.419) were much lower than the other well (~1500) of the same concentration of chemical (Figure 20A). This made the cytotoxicity curve decreasing at the lowest concentrations and thus triggering a very low IC_{50} value of 49.95 µg/ml (Figure 20B).

All the laboratories for each run obtained the same judgment (N) confirming the between-laboratory reproducibility.

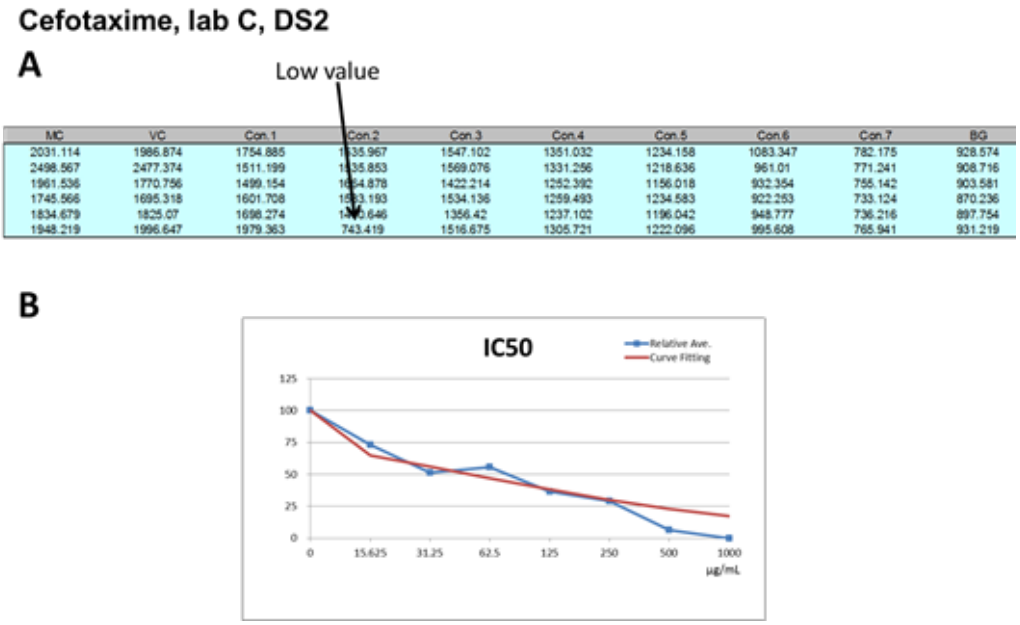


Figure 20 Data obtained with Cefotaxime sodium salt in lab C during DS2

Penicillin G

Table 21 shows the results obtained with penicillin G during phase 2a.

Table 21 Results obtained with penicillin G during phase 2a showing the IC_{50} and ID_{50} values with the final judgment obtained for the DF and the DS studies

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Penicillin G	A	DF	>1000	>1000	N	N
		DS1	942.75	>1000	N	
		DS2	>1000	>1000	N	
	B	DF	>1000	>1000	N	N
		DS1	>1000	>1000	N	
		DS2	>1000	>1000	N	
	C	DF	>1000	>1000	N	N
		DS1	902.64	>1000	N	
		DS2	>1000	>1000	N	

Penicillin G showed highly reproducible results (IC_{50} and ID_{50} values and the judgment) for both DF and DS studies in the three participating lab confirming the between-laboratory reproducibility. Furthermore, the judgment was the same for all the laboratories supporting the between-laboratory reproducibility.

4.1.4.2 Summary of the between-laboratory reproducibility

The result of the between-laboratory reproducibility is shown in Table 22.

Chemical name	Laboratory		
	A	B	C
Cytosine arabinoside	P	P	P
Lithium chloride	P	P	P
Cefotaxime sodium salt	N	N	N
Penicillin G	N	N	N

Table 22 Probability values obtained for each chemical of phase 2a and their classification

As shown in table 22, the between-laboratory reproducibility was verified for all the chemicals where all the labs managed to have the same classification (4/4 = 100%).

4.1.5 Conclusion

The new prediction model and the improvement of the protocol along with the acceptance criteria showed high between-laboratory reproducibility (100%) with a low variability between the labs concerning the IC_{50} and ID_{50} values along with a correct classification of the chemicals.
The next phase was then launched given the good results.

4.2 Phase 2b/2c studies

The second part of phase 2 was again divided into two: phases 2b and 2c. Phase 2b was implemented between May and September 2014 followed by phase 2c from September 2014 to February 2015. The data presented below show the mix of both phases 2b and 2c (2b/2c).
The protocol version used for this phase was the 07E version (see Appendix 34).

4.2.1 Goal

The goal of phase 2b/2c was to verify the within- and the between-laboratory reproducibility of the protocol with more chemicals than the previous phase (eight chemicals).

4.2.2 Study design

The prediction model used for phases 2b/2c is the same as the one used for phase 2a. The curve

fitting used to determine IC_{50} and ID_{50} value is the two parameter curve fitting.
For each chemical, one dose finding study and two or three definitive studies were applied. However, from this phase the dose finding study is not taken into account for the final judgment. The majority judgment and its possibilities are shown in Table 23.

Table 23 Majority judgment applied from phase 2b, 2c detailing the possible cases obtainable

	Dose finding study	Definitive study 1	Definitive study 2	Definitive study 3	Final judgment
Case 1	-	P	P	-	P
Case 2	-	N	P	P	P
Case 3	-	N	P	N	N
Case 4	-	N	N	-	N
Case 5	-	P	N	P	P
Case 6	-	P	N	N	N

To verify the within-laboratory reproducibility, each set (DF+ two or three DS) was repeated three times.

4.2.3 Chemicals

The chemicals used in phases 2b/2c and their information is shown in Table 24.

Table 24 Information of chemicals used in phases 2b/2c

Set	Code No.			Test chemicals	CAS No.	Supplier	Physical property	Solvent	*In vivo* classification
	FDSC	SCAS	KAO						
5	HA101	HB106	HC208	Caffeine	59-08-2	Sigma	Solid	PBS	P
	HA110	HB206	HC105						
	HA201	HB109	HC108						
6	HA102	HB101	HC211	Glibenclamide	10238-21-8	Sigma	Solid	DMSO	N
	HA111	HB209	HC106						
	HA204	HB110	HC109						
7	HA103	HB212	HC101	Acetazolamide	59-66-5	Sigma	Solid	DMSO	P
	HA207	HB102	HC110						
	HA112	HB111	HC203						
8	HA104	HB201	HC102	Valproic acid	1069-66-5	Sigma	Solid	PBS	P
	HA210	HB103	HC111						
	HA107	HB112	HC206						
9	HA202	HB104	HC103	Saccharin sodium salt	82385-42-0	Sigma	Solid	PBS	N
	HA105	HB107	HC209						
	HA108	HB204	HC112						
10	HA205	HB105	HC104	(+)-Camphor	464-49-3	TCI	Solid	DMSO	N
	HA106	HB108	HC212						
	HA109	HB207	HC107						
11	HA208	HB210	HC201	Ascorbic acid	50-81-7	Sigma	Solid	PBS	N
	HA211	HB202	HC204						
	HA203	HB205	HC207						
12	HA206	HB208	HC210	BrdU	59-14-3	TCI	Solid	DMSO	P
	HA209	HB211	HC202						
	HA212	HB203	HC205						

4.2.4 Results

4.2.4.1 IC_{50} and ID_{50} values variation and final judgment

The results are presented for each chemical.

Caffeine

Table 25 shows the results obtained with caffeine during phase 2b/2c in the three participating labs.

Table 25. IC_{50} and ID_{50} values and the judgment obtained during phase 2b/2c with caffeine (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Test chemical	Lab	Set	Run	IC_{50} (µg/ml)	ID_{50} (µg/ml)	Set Judgment	[a]Final judgment
Caffeine	A	1	DS1	> 125	124.48	N	P
			DS2	> 125	> 125		
		2	DS1	175.09	89.74	P	
			DS2	158.83	80.00		
		3	DS1	194.3	100.63	P	
			DS2	210.92	105.44		
	B	1	DS1	> 125	> 125	N	N
			DS2	> 125	> 125		
		2	DS1	> 250	163.32	N	
			DS2	> 250	> 250		
			DS3	> 250	> 250		
		3	DS1	> 125	> 125	N	
			DS2	> 125	92.17		
			DS3	> 125	112.27		
	C	1	DS1	> 250	127.88	P	
			DS2	> 250	144.08		
		2	DS1	> 7.8	> 7.8	N	
			DS2	> 7.8	> 7.8		
		3	DS1	> 62.5	> 62.5	N	
			DS2	> 62.5	> 62.5		

[a]. Judgment decided with the majority judgment of the 3 sets

Within-laboratory reproducibility (see Table 25, set judgment)

The within-laboratory reproducibility has been verified only for lab B where all the IC_{50} and ID_{50} values are close to each other in each set. However, the first set of lab A showed a negative result while the two other sets give a positive result. This is due to a different MD chosen. Indeed set 2 and 3 have an MD set at 250 µg/ml but the MD of the first set was determined at 125 µg/ml. In this case only, set 2 and set 3 managed to reach the cytotoxic dose and to trigger an IC_{50}. When entering the values in the prediction model, set 2 and set 3 give a positive prediction because the gap between the IC_{50} and the ID_{50} is higher than in set 1 which triggered a negative prediction.

For the same reason (MD determination), the first set of lab C triggered a positive prediction while the two other set triggered a negative prediction for caffeine.

Between-laboratory reproducibility (see Table 25, final judgment)

For the same reason as mentioned above (MD determination), the between-laboratory reproducibility was not satisfied for caffeine due to the positive prediction obtained for Lab A while the two other

laboratories got a negative result.

Glibenclamide

Table 26 shows the results obtained with glibenclamide during phase 2b/2c in the three participating labs.

Table 26. IC_{50} and ID_{50} values and the judgment obtained during phase 2b/2c with glibenclamide (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Test chemical	Lab	Set	Run	IC_{50} (µg/ml)	ID_{50} (µg/ml)	Set Judgment	[a]Final judgment
Glibenclamide	A	1	DS1	86.78	31.69	P	P
			DS2	91.77	18.35		
		2	DS1	57.31	17.98	P	
			DS2	36.16	25.13		
		3	DS1	> 62.5	23.69	P	
			DS2	> 62.5	24.47		
	B	1	DS1	91.32	18.81	P	P
			DS2	100.04	31.93		
		2	DS1	> 125	10.62	P	
			DS2	> 125	31.98		
		3	DS1	> 250	17.09	P	
			DS2	23.45	2.58		
	C	1	DS1	> 125	12.55	P	P
			DS2	64.34	32.26		
		2	DS1	> 62.5	33.01	P	
			DS2	> 62.5	22.46		
		3	DS1	87.81	46.79	P	
			DS2	89.80	38.32		

[a]. Judgment decided with the majority judgment of the 3 sets

Within-laboratory reproducibility (see Table 26, set judgment)

Each lab obtained the same prediction (positive) in each set confirming the reproducibility of the results within the labs. The ID_{50} values did not vary between the studies. However, the IC_{50} values showed a higher variance between the studies which is due to different MD chosen. Furthermore, in the case of the third set of lab B DS2, the IC50 value is much lower than the other ones (23.45 µg/ml). This is due to a dilution ratio of 10 that was used during the experiment greatly lowering the accuracy of the results. The reason why no cytotoxicity was observed for the third set of lab B DS1 could not be explained.

Between-laboratory reproducibility (see Table 26, final judgment)

Each lab obtained the same prediction (positive) confirming the reproducibility of the results between the labs.

Acetazolamide

Table 27 shows the results obtained with acetazolamide during phase 2b/2c in the three participating labs.

Table 27. IC_{50} and ID_{50} values and the judgment obtained during phase 2b/2c with acetazolamide (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Test chemical	Lab	Set	Run	IC_{50} (µg/ml)	ID_{50} (µg/ml)	Set Judgment	[a]Final judgment
Acetazolamide	A	1	DS1	322.52	27.77	P	P
			DS2	> 500	17.99		
		2	DS1	124.39	24.40	P	
			DS2	88.97	11.95		
		3	DS1	> 250	24.72	P	
			DS2	> 250	20.22		
	B	1	DS1	> 25	>25	P	P
			DS2	> 25	0.23		
			DS3	>25	4.49		
		2	DS1	> 250	1.37	P	
			DS2	> 250	1.09		
		3	DS1	> 250	2.01	P	
			DS2	> 250	1.46		
	C	1	DS1	> 250	8.02	P	P
			DS2	139.58	17.72		
		2	DS1	> 125	31.73	P	
			DS2	> 125	14.3		
		3	DS1	> 250	10.4	P	
			DS2	> 250	2.45		

[a]. Judgment decided with the majority judgment of the 3 sets

Within-laboratory reproducibility (see Table 27, set judgment)

Each lab obtained the same prediction (positive) in each set confirming the reproducibility of the results within the labs. The ID_{50} values did not highly vary between the studies. However, for the first step of Lab B, the IC_{50} is set at > 25 µg/ml. Although it did not affect the final judgment, this case is worth noticing. The reason is that the laboratory set the highest concentration tested at 25 µg/ml, maybe to try to reduce the concentration range to get a better accuracy of the results (due to the very low ID_{50} values). However, since no cytotoxicity was triggered at 250 µg/ml (DF, data not

shown), the highest concentration should not have been lowered.

Between-laboratory reproducibility (see Table 27, final judgment)

Each lab obtained the same prediction (positive) confirming the reproducibility of the results between the labs.

Valproic acid

Table 28 shows the results obtained with valproic acid during phase 2b/2c in the three participating labs.

Table 28. IC_{50} and ID_{50} values and the judgment obtained during phase 2b/2c with valproic acid (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Test chemical	Lab	Set	Run	IC_{50} (μg/ml)	ID_{50} (μg/ml)	Set Judgment	[a]Final judgment
Valproic acid	A	1	DS1	146.15	127.1	P	P
			DS2	130.34	122.84		
		2	DS1	140.94	110.5	P	
			DS2	149.78	134.23		
		3	DS1	102.39	236.78	P	
			DS2	100.3	104.77		
			DS3	116.86	123.44		
	B	1	DS1	156.55	37.91	P	P
			DS2	120.96	65.89		
		2	DS1	161.85	27.54	P	
			DS2	126.18	67.7		
		3	DS1	122.64	92.18	P	
			DS2	131.48	86.1		
	C	1	DS1	119.89	65.36	P	P
			DS2	149.01	39.58		
		2	DS1	156.87	66.43	P	
			DS2	103.91	60.27		
		3	DS1	128.93	115.31	P	
			DS2	126.42	46.3		

[a]. Judgment decided with the majority judgment of the 3 sets

Within-laboratory reproducibility (see Table 28, set judgment)

Each lab obtained the same prediction (positive) in each set confirming the reproducibility of the results within the labs. The IC_{50} and ID_{50} values did not vary between the studies.

Between-laboratory reproducibility (see Table 28, final judgment)

Each lab obtained the same prediction (positive) confirming the reproducibility of the results between the labs.

Saccharin sodium salt

Table 29 shows the results obtained with saccharin sodium salt during phase 2b/2c in the three participating labs.

Table 29. IC_{50} and ID_{50} values and the judgment obtained during phase 2b/2c with saccharin sodium salt (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Test chemical	Lab	Set	Run	IC_{50} (µg/ml)	ID_{50} (µg/ml)	Set Judgment	[a]Final judgment
Saccharin sodium salt	A	1	DS1	> 1000	> 1000	N	N
			DS2	> 1000	> 1000		
		2	DS1	> 1000	> 1000	N	
			DS2	> 1000	> 1000		
		3	DS1	> 1000	> 1000	N	
			DS2	> 1000	> 1000		
	B	1	DS1	> 1000	> 1000	N	N
			DS2	> 1000	> 1000		
		2	DS1	> 1000	> 1000	N	
			DS2	> 1000	> 1000		
		3	DS1	> 1000	> 1000	N	
			DS2	> 1000	> 1000		
	C	1	DS1	> 1000	> 1000	N	N
			DS2	> 1000	> 1000		
		2	DS1	> 1000	> 1000	N	
			DS2	> 1000	> 1000		
		3	DS1	> 1000	> 1000	N	
			DS2	> 1000	> 1000		

[a]. Judgment decided with the majority judgment of the 3 sets

Within-laboratory reproducibility (see Table 29, set judgment)

Each lab obtained the same prediction (negative) in each set confirming the reproducibility of the results within the labs. The IC_{50} and ID_{50} values did not vary between the studies.

Between-laboratory reproducibility (see Table 29, final judgment)

Each lab obtained the same prediction (negative) confirming the reproducibility of the results between the labs.

(+)-Camphor

Table 30 shows the results obtained with (+)-camphor during phase 2b/2c in the three participating labs.

Table 30. IC_{50} and ID_{50} values and the judgment obtained during phase 2b/2c with (+)-camphor (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Test chemical	Lab	Set	Run	IC_{50} (µg/ml)	ID_{50} (µg/ml)	Set Judgment	[a]Final judgment
(+)-Camphor	A	1	DS1	> 250	> 250	N	N
			DS2	> 250	> 250		
		2	DS1	> 250	> 250	N	
			DS2	> 250	> 250		
		3	DS1	> 250	> 250	N	
			DS2	> 250	> 250		
	B	1	DS1	> 250	> 250	N	N
			DS2	> 250	> 250		
		2	DS1	> 250	127.02	N	
			DS2	> 250	> 250		
			DS3	> 250	> 250		
		3	DS1	> 250	> 250	N	
			DS2	> 250	> 250		
	C	1	DS1	> 250	> 250	N	N
			DS2	> 250	> 250		
		2	DS1	> 125	> 125	N	
			DS2	> 125	> 125		
		3	DS1	> 125	> 125	N	
			DS2	> 125	> 125		

[a]. Judgment decided with the majority judgment of the 3 sets

Within-laboratory reproducibility (see Table 30, set judgment)

Each lab obtained the same prediction (negative) in each set confirming the reproducibility of the results within the labs. The IC_{50} and ID_{50} values did not vary between the studies. The maximum dose chosen for lab C set 2 and 3 varies from the other studies and labs but this did not affect the final judgment.

Between-laboratory reproducibility (see Table 30, final judgment)

Each lab obtained the same prediction (negative) confirming the reproducibility of the results between the labs.

Ascorbic acid

Table 31 shows the results obtained with ascorbic acid during phase 2b, 2c in the three participating labs.

Table 31. IC_{50} and ID_{50} values and the judgment obtained during phase 2b/2c with ascorbic acid (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Test chemical	Lab	Set	Run	IC50 (µg/ml)	ID50 (µg/ml)	Set Judgment	[a]Final judgment
Ascorbic acid	A	1	DS1	552.44	216.15	P	N
			DS2	538.77	470.80		
		2	DS1	462.28	548.47	N	
			DS2	225.34	630.78		
		3	DS1	545.64	681.89	N	
			DS2	324.62	597.13		
	B	1	DS1	62.71	195.45	N	N
			DS2	119.77	293.70		
		2	DS1	74.55	60.26	P	
			DS2	13.52	18.39		
		3	DS1	630.97	749.75	N	
			DS2	848.43	389.97		
			DS3	736.1	621.97		
	C	1	DS1	668.39	445.86	P	P
			DS2	964.70	255.66		
		2	DS1	> 1000	9.13	P	
			DS2	359.14	222.02		
		3	DS1	> 1000	41.95	P	
			DS2	770.06	5.14		

[a]. Judgment decided with the majority judgment of the 3 sets

Within-laboratory reproducibility (See table 31, set judgment)

The reproducibility has been verified in lab C where all the sets triggered the same judgment (positive).

However, for lab A and B, the results were different. For lab A, the first set gave a positive result while the two other ones gave a negative result and in the case of lab B, the second set gave a positive result while the two other ones gave a negative result. Even when looking at the cytotoxicity and differentiation toxicity curves, no peculiar reason could be found. However, when referring to phase 0 and the lead lab results, it may be possible that this chemical has a tendency to give results with a high variation. This may be due to the intrinsic properties of the chemical.

Between-laboratory reproducibility (See Table 31, final judgment)

The between-laboratory reproducibility could not be verified for ascorbic acid where lab C had a different result compared to the two other labs. The reason is the same as the one proposed in the within-laboratory reproducibility part.

BrdU

Table 32 shows the results obtained with BrdU during phase 2b/2c in the three participating labs.

Table 32. IC_{50} and ID_{50} values and the judgment obtained during phase 2b/2c with BrdU (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Test chemical	Lab	Set	Run	IC_{50} (µg/ml)	ID_{50} (µg/ml)	Set Judgment	[a]Final judgment
BrdU	A	1	DS1	0.41	0.28	P	P
			DS2	0.38	0.3		
		2	DS1	0.35	0.34	P	
			DS2	> 0.5	0.21		
			DS3	0.35	0.24		
		3	DS1	0.44	0.44	P	
			DS2	0.25	0.27		
	B	1	DS1	0.22	0.53	P	P
			DS2	0.19	0.11		
		2	DS1	0.33	0.3	P	
			DS2	0.49	0.53		
		3	DS1	0.26	0.34	P	
			DS2	0.29	0.87		
	C	1	DS1	0.04	0.37	P	P
			DS2	0.32	0.14		
		2	DS1	0.51	0.13	P	
			DS2	0.69	0.37		
		3	DS1	1.09	0.25	P	
			DS2	0.54	0.41		

[a]. Judgment decided with the majority judgment of the 3 sets

Within-laboratory reproducibility (see Table 32, set judgment)

Each lab obtained the same prediction (positive) in each set confirming the reproducibility of the results within the labs. The positivity result is due to the very low values of both IC_{50} and ID_{50} values. The IC_{50} and ID_{50} values did not vary between the studies.

Between-laboratory reproducibility (see Table 32, final judgment)

Each lab obtained the same prediction (positive) confirming the reproducibility of the results between the labs.

4.2.4.2 Summary of the between- and within-laboratory reproducibilities

Phase 2b/ 2c allowed to verify the within- and the between-laboratory reproducibility of the protocol. The summary of the results is presented in Table 33.

Table 33 Judgment results in each lab along the sets for the eight chemicals tested in phases 2b/2c

Test chemical	Laboratory	Set			Concordance between sets	Mean
		1	2	3		
Caffeine	A	N	P	P	2/3	P
	B	N	N	N	3/3	N
	C	P	N	N	2/3	N
Glibenclamide	A	P	P	P	3/3	P
	B	P	P	P	3/3	P
	C	P	P	P	3/3	P
Acetazolamide	A	P	P	P	3/3	P
	B	P	P	P	3/3	P
	C	P	P	P	3/3	P
Valproic acid	A	P	P	P	3/3	P
	B	P	P	P	3/3	P
	C	P	P	P	3/3	P
Saccharin sodium salt	A	N	N	N	3/3	N
	B	N	N	N	3/3	N
	C	N	N	N	3/3	N
(+)-camphor	A	N	N	N	3/3	N
	B	N	N	N	3/3	N
	C	N	N	N	3/3	N
Ascorbic acid	A	P	N	N	2/3	N
	B	N	P	N	2/3	N
	C	P	P	P	3/3	P
BrdU	A	P	P	P	3/3	P
	B	P	P	P	3/3	P
	C	P	P	P	3/3	P

Concerning the within-laboratory reproducibility, twenty coded chemicals out of twenty-four satisfied the reproducibility (same judgment obtained three times for the same chemical) which

raises the within-laboratory reproducibility to 83.3%, satisfying the criterion fixed in the study plan (75%).

As for the between-laboratory reproducibility, it has been verified by compiling the results obtained in phases 2a, 2b and 2c. During phase 2a, four chemicals out of four satisfied the between-laboratory reproducibility (see part **4.1.4.2**) and phase 2b/2c six out of eight which leads to a general between-laboratory reproducibility of 10/12= 83.3% satisfying again the study plan's criterion of 75%.

4.2.5 Conclusion

The revised prediction model and the improved the protocol showed high between- and within-laboratory reproducibilities (83.3%). There were still some concerns about the two parameters curve fitting applied (where sometimes the IC_{50} value was lower than the ID_{50} one) and a three parameter curve fitting was advised at the end of phase 2c. The prediction model was revised (see **4.3.2.2**) and phase 3 was then launched with sixteen other coded chemicals thanks to the good results obtained during phase 2.

4.3 Phase 3 study

Phase 3 study started in February 2015 and lasted one year.

4.3.1 Goal

The goal of phase 3 was to verify the between-laboratory reproducibility with the new prediction model and the three parameter curve fitting analysis with sixteen coded chemicals (Table 34).

4.3.2 Study design

4.3.2.1 Protocol

The Hand1-Luc EST protocol used for this phase was the 08E version (see Appendix 35).
Each chemical was tested in one set. One set consisted in one dose finding study followed by two or three definitive studies. The decision on positivity or negativity of the chemical was taken with the majority judgment.

4.3.2.2 Data analysis

Three parameter curve fitting

The results of phase 2 showed that there were some cases where the IC_{50} remained lower than the ID_{50} which led to think that the curve fitting method could be improved by implementing a three parameter curve fitting method instead of a two parameter curve fitting method and use the raw data instead of the normalized one. The data from phase 2a were then analyzed in two different ways. The first one was to use the Graphpad PRISM software and the second one was to use a curve fitting developed by Dr. Omori. PRISM software gave good results but for some cases, the software could not fit the curves (data not shown). The three parameter curve fitting developed by Dr. Omori gave good results and they were described in a teleconference call where the IC_{50} and ID_{50} relationship was improved. This is the method used for phase 3.

Prediction model

The prediction model was changed from phase 2 to phase 3 for two main reasons.

The first reason is that the IC_{50} and ID_{50} values used for the establishment of the model are changed because of the new curve fitting. Furthermore, the lead lab had new chemicals data which led the model to be also adapted. Indeed, the previous one was based on the data of twenty-one chemicals but the new one is now based on seventy-one chemicals.

This prediction model is the last one and is the one to be proposed for validation. A more detailed explanation of the prediction can be found in the published report (Nagahori *et al*, 2016). For a better understanding the prediction model is shortly described below:

For the prediction, three parameters are used: the IC_{50}, the ID_{50} and the MD (maximum dose). The equation is as follows:

$$\text{Score} = 1.474 \times \log(IC_{50}/ID_{50}) + 0.714 \times \log(MD/IC_{50}) - 1.109$$

And the probability is obtained by applying the inverse logit to the score.

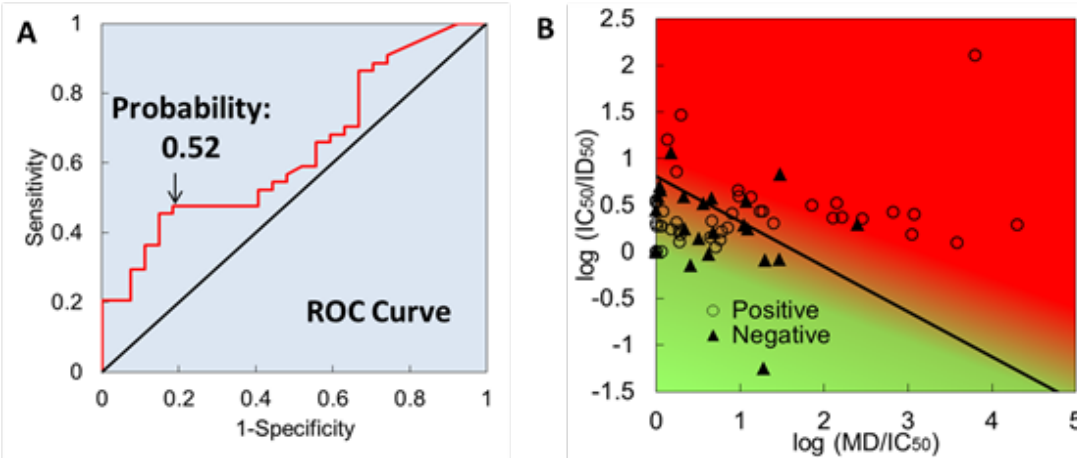


Figure 21 A: ROC curve; B: Ideal separation (0.52) between the *in vivo* positive and negative chemicals

The data shown in Figure 21B are based on the data of seventy-one chemicals obtained in the lead lab (see Appendix 36).

The plot of sensitivity versus 1-Specifity is called receiver operating characteristic (ROC) curve. This curve plays a central role in evaluating diagnostic ability of tests to discriminate the true state of subjects, finding the optimal cut off values, and comparing two alternative diagnostic tasks when each task is performed on the same subject (Karimollah Hajian-Tilaki, 2013).

The ROC curve (Figure 21A), allowed to fix a cut-off probability of 0.52 which was confirmed to be the best one by placing the seventy-one chemicals in a graph with in abscissa Log (MD/IC_{50}) and in ordinate Log (IC_{50}/ID_{50}) (Figure 21B).

4.3.3 Chemicals

The sixteen chemicals used for phase 3 and their information are presented in Table 34.

Table 34 Information of chemicals used in phase 3

No	Code No.			Test chemical	CAS No.	Supplier	Physical property	Solvent	*In vivo* classification
	FDSC	SCAS	KAO						
1	HA301	HB310	HC314	Methotrexate	59-05-2	TCI	Solid	DMSO	P
2	HA306	HB315	HC319	6-Aminonicotinamide	329-89-5	TCI	Solid	DMSO	P
4	HA316	HB301	HC310	Dexamethasone	50-02-2	TCI	Solid	DMSO	P
6	HA307	HB311	HC320	Diphenylhydantoin	57-41-0	Wako	Solid	DMSO	P
7	HA312	HB316	HC301	Hydroxyurea	127-07-1	Sigma	Solid	PBS	P
9	HA303	HB307	HC311	Boric acid	10043-35-3	sigma	Solid	PBS	P
10	HA308	HB312	HC316	Methoxyacetic acid	625-45-6	Wako	Liquid	PBS	P
11	HA313	HB317	HC302	Dimethadione	695-53-4	aldrich	Solid	PBS	P
12	HA318	HB303	HC307	Dimethyl phthalate	131-11-3	TCI	Liquid	DMSO	N
13	HA304	HB308	HC312	Cimetidine	51481-61-9	TCI	Solid	PBS	N
14	HA309	HB313	HC317	Benzophenone	119-61-9	Sigma	Solid	DMSO	N
15	HA314	HB318	HC303	Ethylene glycol methyl ether	109-86-4	TCI	Liquid	PBS	N
16	HA319	HB304	HC308	Isoniazid	54-85-3	Sigma	Solid	PBS	N
18	HA310	HB314	HC318	Acebutolol hydrochloride	34381-68-5	Sigma	Solid	PBS	N
19	HA315	HB319	HC304	Diphenhydramine HCl	147-24-0	Sigma	Solid	PBS	N
20	HA320	HB305	HC309	Acrylamide	79-06-1	Wako	Solid	PBS	N

4.3.4 Results

4.3.4.1 IC_{50} and ID_{50} values variation and final judgment

The results are presented for each chemical from tables 35 to 50.

Methotrexate

Table 35 shows the results obtained with methotrexate during phase 3.

Table 35 Results obtained with methotrexate during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Test chemical	Lab	Run	IC_{50} value (μg/ml)	ID_{50} value (μg/ml)	Runs judgment	Final judgment
Methotrexate	A	DS1	0.04	0.02	P	P
		DS2	0.06	0.02	P	
	B	DS1	0.11	0.03	P	P
		DS2	0.05	0.003	P	
	C	DS1	0.04	0.05	P	P
		DS2	0.04	0.03	P	

The IC_{50} and ID_{50} values remain close to each other. Given the fact that all the values are very low (lower than 0.11), the insertion in the prediction model gave a positive result. All the labs obtained the same judgment, verifying the between-laboratory reproducibility for methotrexate.

6-aminonicotinamide

Table 36 shows the results obtained with 6-aminonicotinamide during phase 3.

Table 36 Results obtained with 6-aminonicotinamide during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (μg/ml)	ID_{50} value (μg/ml)	Runs judgment	Final judgment
6-aminonicotinamide	A	DS1	1.40	0.58	P	P
		DS2	1.46	0.88	P	
	B	DS1	2.03	0.67	P	P
		DS2	3.02	0.65	P	
	C	DS1	0.68	0.71	P	P
		DS2	>1.25	>1.25	P	

The IC_{50} and ID_{50} values remain close to each other. Given the fact that all the values are very low (lower than 3μg/ml), the insertion in the prediction model gave a positive results. All the labs obtained the same judgment, verifying the between-laboratory reproducibility for 6-aminonicotinamide.

Dexamethasone

Table 37 shows the results obtained with dexamethasone during phase 3.

Table 37 Results obtained with dexamethasone during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (μg/ml)	ID_{50} value (μg/ml)	Runs judgment	Final judgment
Dexamethasone	A	DS1	31.56	20.27	N	N
		DS2	34.69	27.92	N	
	B	DS1	37.35	13.98	N	N
		DS2	52.23	24.13	N	
	C	DS1	13.72	21.82	N	N
		DS2	> 125	> 125	N	

The IC_{50} and ID_{50} values remained close to each other confirming the reproducibility of the assay. All the labs obtained the same judgment, verifying the between-laboratory reproducibility for dexamethasone. The DS of Lab C was however worth noticing where no toxicity was triggered. After having checking with the lab responsible for the experiment, they acknowledged that the chemical was not added in the medium.

Diphenylhydantoin

Table 38 shows the results obtained with diphenylhydantoin during phase 3.

Table 38 Results obtained with diphenylhydanthoin during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies
(Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Diphenylhydantoin	A	DS1	> 125	9.96	P	P
		DS2	> 125	14.33	P	
	B	DS1	> 15.625	3.81	N	N
		DS2	> 15.625	4.08	N	
	C	DS1	41.65	33.4	N	N
		DS2	19.46	38.14	N	

The close data observed between the run confirms the reproducibility of the test. However, there were different judgment observed and thus, the between-laboratory reproducibility could not be verified. The ID_{50} values obtained in each lab do not vary dramatically. However, due to the different MD chosen between lab A and lab B, the gap between the IC_{50} and the ID_{50} varies and triggers then a different judgment (Positive for lab A and negative for lab B). Lab C, exhibiting a cytotoxic concentration of chemical at 41.65µg/ml for DS1 and 19.46µg/ml for DS2, triggers a negative judgment because of a smaller gap between the IC_{50} and ID_{50} values.

Hydroxyurea

Table 39 shows the results obtained with hydroxyurea during phase 3.

Table 39 Results obtained with hydroxyurea during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Hydroxyurea	A	DS1	3.06	2.14	P	P
		DS2	6.31	3.39	P	
	B	DS1	3.44	2.36	P	P
		DS2	2.97	1.83	P	
	C	DS1	1.51	1.00	P	P
		DS2	1.29	1.78	P	

The IC_{50} and ID_{50} values remained close to each other confirming the reproducibility of the test. Given the fact that all the values are very low (lower than 6), the insertion in the prediction model gave positive results. All the labs obtained the same judgment, verifying the between-laboratory reproducibility for hydroxyurea.

Boric acid

Table 40 shows the results obtained with boric acid during phase 3.

Table 40 Results obtained with boric acid during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Boric acid	A	DS1	62.59	47.40	N	N
		DS2	83.99	55.32	N	
	B	DS1	115.41	78.26	N	N
		DS2	66.07	48.76	N	
	C	DS1	94.06	40.98	N	N
		DS2	94.5	31.32	P	
		DS3	97.37	44.60	N	

The IC_{50} and ID_{50} values remain close to each other confirming the reproducibility of the test. All the labs obtained the same judgment, verifying the between-laboratory reproducibility for boric acid.

Methoxyacetic acid

Table 41 shows the results obtained with methoxyacetic acid during phase 3.

Table 41 Results obtained with methoxyacetic acid during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies
(Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (μg/ml)	ID_{50} value (μg/ml)	Runs judgment	Final judgment
Methoxyacetic acid	A	DS1	392.88	586.88	N	N
		DS2	382.57	342.91	N	
	B	DS1	656.17	475.35	N	N
		DS2	544.62	738.53	N	
	C	DS1	585.04	747.86	N	N
		DS2	> 1000	>1000	N	

The IC_{50} and ID_{50} values remain close to each other confirming the reproducibility of the test. All the labs obtained the same judgment, verifying the between-laboratory reproducibility for methoxyacetic acid

Dimethadione

Table 42 shows the results obtained with dimethadione during phase 3.

Table 42 Results obtained with dimethadione during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (μg/ml)	ID_{50} value (μg/ml)	Runs judgment	Final judgment
Dimethadione	A	DS1	925.95	> 1000	N	N
		DS2	> 1000	535.55	N	
	B	DS1	> 1000	459.3	N	N
		DS2	> 1000	228.34	N	
	C	DS1	902.68	727.71	N	N
		DS2	> 1000	> 1000	N	

The IC_{50} and ID_{50} values remained close to each other between the runs confirming the reproducibility of the test. All the labs obtained the same judgment, verifying the between-laboratory reproducibility for dimethadione

Dimethyl phtalate

Table 43 shows the results obtained with dimethyl phtalate during phase 3.

Table 43 Results obtained with dimethyl phthalate during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies

(Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Dimethyl phtalate	A	DS1	> 250	130.57	N	N
		DS2	> 250	69.56	N	
	B	DS1	> 250	28.41	P	P
		DS2	> 250	24.17	P	
	C	DS1	10.57	> 1000	N	N
		DS2	94.98	> 1000	N	

The IC_{50} and ID_{50} values remained close to each other between the runs confirming the reproducibility of the test.

It seems the ID_{50} values situated between 20 and 70 µg/ml when an IC_{50} and MD are at 250 µg/ml, are in the critical window, a slight variation of ID_{50} could trigger a positive or a negative result according to the prediction model. This is what happens between lab A and lab B which triggers a different judgment.

In the case of lab C, the IC_{50} values are greatly lower than ID_{50} (Figure 22 and Table 43). This can be explained as follows: Dimethyl phthalate is a liquid chemical that only dissolves in DMSO and precipitate (or forms small bubbles) in the assay medium at high concentration. Lab C set the highest concentration to 1000 µg/ml which consists in almost the pure substance. It is possible that during the precipitation evaluation the bubbles of DMSO that did not dissolve were not clearly observable. When pipetting assay medium from the 2ml assay block, the chemical may have stayed (bubbles on the top of the liquid and pipetting was done in the middle) in the assay block due to its incapacity to dissolve at 1000 and 500 µg/ml leading to non toxic effect at highest concentration.

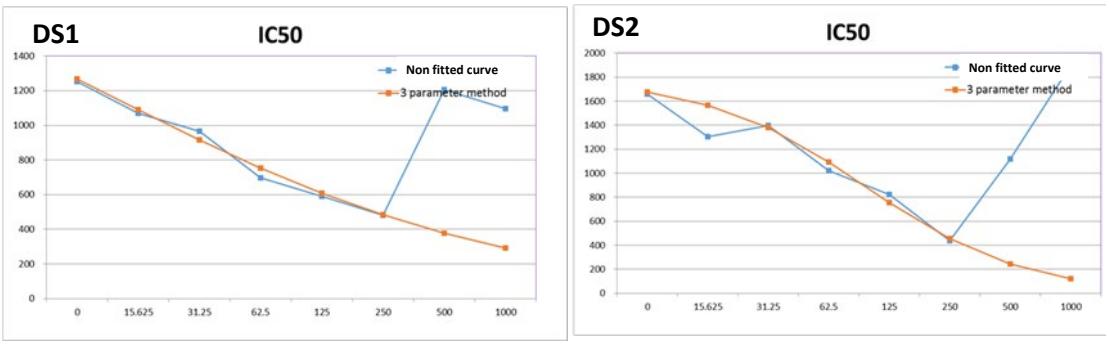


Figure 22 Results of cytotoxicity for dimethyl phtalate obtained in lab C

(Comparison of cytotoxicity curves between DS1 and DS2)

Cimetidine

Table 44 shows the results obtained with cimetidine during phase 3.

Table 44 Results obtained with cimetidine during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (μg/ml)	ID_{50} value (μg/ml)	Runs judgment	Final judgment
Cimetidine	A	DS1	> 250	> 250	N	N
		DS2	> 250	> 250	N	
	B	DS1	> 250	> 250	N	N
		DS2	> 250	> 250	N	
	C	DS1	> 250	> 250	N	N
		DS2	> 250	> 250	N	

The IC_{50} and ID_{50} values remained close to each other between the runs confirming the reproducibility of the test. All the labs obtained the same judgment, verifying the between-laboratory reproducibility for cimetidine.

Benzophenone

Table 45 shows the results obtained with benzophenone during phase 3.

Table 45 Results obtained with benzophenone during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (μg/ml)	ID_{50} value (μg/ml)	Runs judgment	Final judgment
Benzophenone	A	DS1	> 62.5	> 62.5	N	N
		DS2	> 62.5	> 62.5	N	
	B	DS1	> 250	3.32	P	P
		DS2	103.47	7.51	P	
	C	DS1	16.92	143.13	N	N
		DS2	50.79	208.65	N	

The IC_{50} and ID_{50} values remain close to each other between the runs confirming the reproducibility of the test. Lab B obtained a positive result while the two other labs obtained a negative result. The positive results obtained in lab B is due to a larger gap between the IC_{50} and the ID_{50} values compared to lab A (where no gap was observed) and an IC_{50} value lower than the ID_{50} value for lab C. For the latter, the fact that IC_{50} values that were considerably lower than the ID_{50} values is due to

precipitation. Indeed, as shown in Figure 23, the concentration of 166.66 and 250 µg/ml do not completely kill the cells leading to a flat curve at the end triggering a curve fitting shifted to the right. DS1 displayed the same pattern (data not shown).

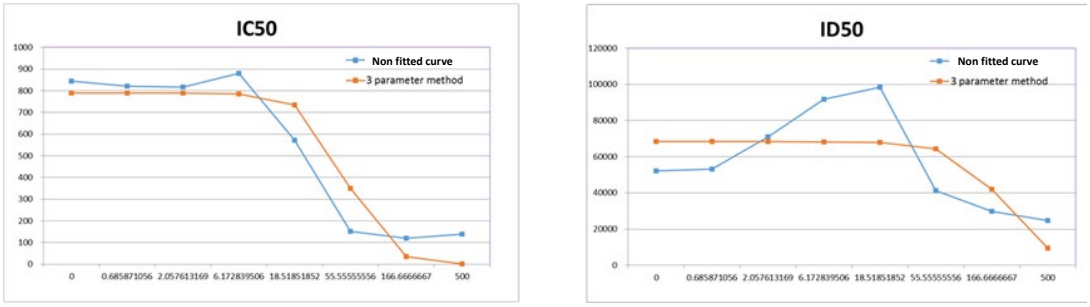


Figure 23 Results obtained by lab C (DS2) with benzophenone

Ethylene glycol methyl ether

Table 46 shows the results obtained with ethylene glycol methyl ether during phase 3.

Table 46 Results obtained with ethylene glycol methyl ether during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies
(Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Ethylene glycol methyl ether	A	DS1	>1000	>1000	N	N
		DS2	>1000	>1000	N	
	B	DS1	>1000	>1000	N	N
		DS2	>1000	>1000	N	
	C	DS1	>1000	>1000	N	N
		DS2	>1000	>1000	N	

No toxicity was observed between the runs for all the laboratories confirming the reproducibility of the test. All the labs obtained the same judgment, verifying the between-laboratory reproducibility for ethylene glycol methyl ether.

Isoniazid

Table 47 shows the results obtained with isoniazid during phase 3.

Table 47 Results obtained with isoniazid during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Isoniazid	A	DS1	287.04	376.41	N	N
		DS2	269.78	356.65	N	
	B	DS1	406.45	421.92	N	N
		DS2	263.94	312.65	N	
	C	DS1	177.82	280.12	N	N
		DS2	303.54	385.88	N	

The IC_{50} and ID_{50} values remain close to each other between the runs confirming the reproducibility of the test. All the labs obtained the same judgment verifying the between-laboratory reproducibility for isoniazid.

Acebutolol HCl

Table 48 shows the results obtained with acebutolol HCl during phase 3.

Table 48 Results obtained with acebutolol HCl during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	**Lab**	**Run**	**IC_{50} value (µg/ml)**	**ID_{50} value (µg/ml)**	**Runs judgment**	**Final judgment**
Acebutolol HCl	A	DS1	182.16	220.38	N	N
		DS2	206.92	202.01	N	
	B	DS1	95.46	38.42	P	N
		DS2	95.84	74.24	N	
		DS3	95.48	69.06	N	
	C	DS1	9.45	49.08	N	N
		DS2	35.89	94.33	N	

The IC_{50} and ID_{50} values remain close to each other between the runs confirming the reproducibility of the test. However, for lab C the ID_{50} values of DS1 and DS1 are much higher (49.08 and 94.33µg/ml respectively) than the IC_{50} values (9.45 and 35.89µg/ml respectively) which is not relevant. This is due to the fact that the 10 dilution fold has been chosen where a lower fold dilution would have shown more accurate results (see Figure 24)

All the labs obtained the same judgment verifying the between-laboratory reproducibility for isoniazid.

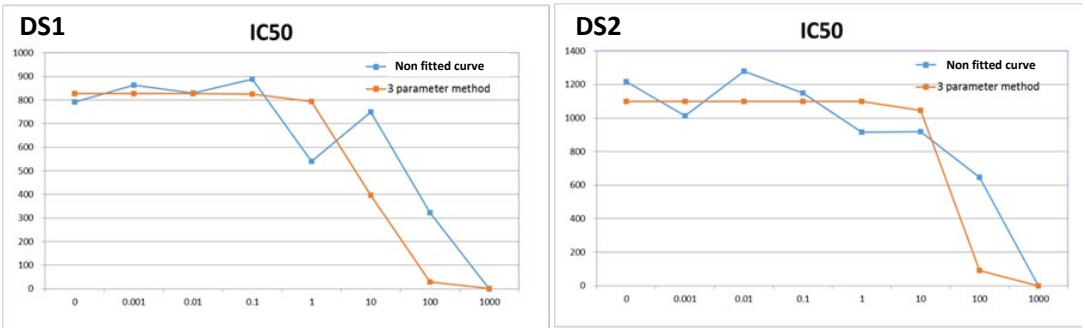


Figure 24 Results of acebutolol HCl obtained in lab C. Comparison of cytotoxicity curves between DS1 and DS2

Diphenhydramine

Table 49 shows the results obtained with diphenhydramine during phase 3.

Table 49 Results obtained with diphenhydramine during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies

(Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (μg/ml)	ID_{50} value (μg/ml)	Runs judgment	Final judgment
Diphenhydramine	A	DS1	5.01	7.67	P	P
		DS2	17.36	6.31	P	
	B	DS1	7.81	2.78	P	P
		DS2	10.82	1.43	P	
	C	DS1	10.20	6.14	P	P
		DS2	2.36	12.59	N	
		DS3	4.8	4.15	P	

The IC_{50} and ID_{50} values remained close to each other between the runs confirming the reproducibility of the test. All the labs obtained the same judgment verifying the between-laboratory reproducibility for diphenhydramine.

Acrylamide

Table 50 shows the results obtained with acrylamide during phase 3.

Table 50 Results obtained with acrylamide during phase 3 showing the IC_{50} and ID_{50} values with the final judgment obtained in the DS studies (Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Lab	Run	IC_{50} value (µg/ml)	ID_{50} value (µg/ml)	Runs judgment	Final judgment
Acrylamide	A	DS1	40.81	64.23	N	N
		DS2	41.53	48.74	N	
	B	DS1	43.61	14.86	P	P
		DS2	34.04	13.52	P	
	C	DS1	14.02	33.61	N	N
		DS2	32.88	48.00	N	

The IC_{50} and ID_{50} values remain close to each other between the runs confirming the reproducibility of the test. Acrylamide did not satisfy the between-laboratory reproducibility because lab B obtained a positive result. This seems to be due to a lower accuracy because they used a 5 dilution ratio while the other lab chose a lower one (2 and 3). This triggered an ID_{50} value lower than the other labs and thus, a positive result.

4.3.4.2 Summary of the between-laboratory reproducibility

Since each chemical was tested in only one set, only the between-laboratory reproducibility was tested with the new parameter curve fitting (three parameters) and the new prediction model.
As shown in Table 51, between-laboratory reproducibility reached 75% (12/16) which again satisfies the criterion fixed in the study plan (75%).

Table 51. Results on the between-laboratory reproducibility obtained in phase 3
(Lab A: FDSC, Lab B: SCAS, Lab C: KAO)

Chemical name	Laboratory			*In vivo* classification
	A	B	C	
Methotrexate	P	P	P	P
6-aminonicotinamide	P	P	P	P
Dexamethasone	N	N	N	P
5,5 diphenylhydantoin	P	N	N	P
Hydroxyurea	P	P	P	P
Boric acid	N	N	N	P
Methoxyacetic acid	N	N	N	P
Dimethadione	N	N	N	P
Dimethyl phthalate	N	P	N	N
Cimetidine	N	N	N	N
Benzophenone	N	P	N	N
EGME	N	N	N	N
Isoniazid	N	N	N	N
Acetobutolol HCl	N	N	N	N
Diphenhydramine	P	P	P	N
Acrylamide	N	P	N	N

Among the sixteen chemicals tested in phase 3, four of them (5,5 diphenylhydantoin, dimethyl phthalate, benzophenone and acrylamide) did not satisfied the between-laboratory reproducibility. The global between-laboratory reproducibility for phase 3 is therefore raised at 75% (12/16) which satisfies the criteria fixed in the study plan confirming the relevancy of the Hand1-Luc EST.

4.3.5 Conclusion

Phase 3, with the three parameter curve fitting and the new prediction model safely satisfied the between-laboratory reproducibility criterion (75%). However, some problems were pointed out and the protocol was improved with minor changes to avoid this kind of problems to happen again. The improvement are listed below

- Since precipitation was observed, especially for liquid chemicals, the precipitation evaluation in the medium should be even more taken care of (watch for DMSO bubbles appearance for example).

- Once the dose finding study is done, the dilution ratio chosen shall be lowest when possible (1.5 or 2).

- When the IC_{50} and/or the ID_{50} of D1 and DS2 are different from a factor 7 (value decided

by analyzing the historical data), then another experiment shall be done and the data used will be the ones the closest to the ones previously obtained. This point was considered to avoid the cases where manipulation errors (for example, forgetting to add chemicals in the medium or dilution mistake) are analyzed for the final judgment.

- The quality of the cells was also pointed out by the participant laboratories. Thus, before sending each lot, one tube shall be thawed and the quality of cells shall be verified.

All those improvements should raise the reproducibility of the protocol. The very final version of the protocol can be found in Appendix 37.

5 DISCUSSION

The discussion is proposed following the modular approach as recommended by ECVAM (Hartung, 2004).

5.1 Test definition

As previously described in this report, the Hand1-Luc EST is defined as a developmental toxicity test *in vitro* assessing the cytotoxicity and the differentiation toxicity by using engineered mouse embryonic stem cells and make them differentiate into cardiomyocyte during 120 hrs (5 days). The cells, called KOB1 ES cells, are stably transfected with a vector containing the luciferase gene modulated by the Hand1 gene promoter. The Hand1 gene being a significant gene at the first steps of the development of the organism is used to reveal the gene disruption of embryotoxic chemicals. The endpoints of the test are the IC_{50}, the ID_{50} which are measured at 120 hr (day5) and the maximum dose (MD) consisting of the concentration of the chemical that dissolves in the assay medium (assessed before the test). The IC_{50} is measured thanks to a commercial kit and represents the concentration of chemical for the one 50% of the cells die. The ID_{50} is measured via a luciferase assay system and shows the concentration of chemical 50% of the signal is inhibited.
The test has been developed to respond to needs of the 3Rs. It finds its origins in trying to solve the problems faced by the original EST validated by ECVAM in 2002 which failed to be raised as an OECD guideline. Among those, Hand1-Luc EST improved the time, the cost and the difficulty of the original test.

5.2 Definition of the performance standards

The performance standards are evaluated at each experiment with the *in vivo* positive chemical 5-FU. The acceptance criteria for this chemical are the following:

- IC_{50}: 0.0345 µg/mL ± 5SD (0.003 - 0.065 µg/mL) and
- ID_{50}: 0.0355 µg/mL ± 5SD (0.003 - 0.067 µg/mL)

5.3 Predictive capacity

5.3.1 Prediction model

The prediction model consists in a formula using IC_{50}, ID_{50} and MD values. The score given by the formula is then converted into a probability. If the probability is higher than 0.52 then the chemical

is considered as positive. If the probability is lower than 0.52 then the chemical is considered as negative.

$$\text{Score} = 1.474 \times \log(IC_{50}/ID_{50}) + 0.714 \times \log(MD/IC_{50}) - 1.109$$

$$\text{Probability} = \text{Inverse logit (Score)}$$

$$\text{Positive: Probability} \geq 52\%$$

$$\text{Negative: Probability} < 52\%$$

The improvements the prediction model undertook are described in the manuscript. The first version is available in section **3.3.2.2**, the second one in section **4.1.2.2** and the final version in **4.3.5**. Shortly, the prediction model was improved and adapted because of the following main reasons:

- IC_{50} and ID_{50} values obtained in the lead lab were changed because of the implementation of the curve fitting and the prediction model needed to be adapted to those changes

- New data from other chemicals (21 chemicals for the first and the second prediction model and 71 chemicals for the last version) were measured in the lead lab and those data were used to improve the prediction model so as to widen the applicability range of the Hand1-Luc EST (see Appendix 36).

- Maximum dose (MD) was implemented and used in the prediction model.

Furthermore, for some chemicals, no IC_{50} or ID_{50} were triggered given values higher than the maximum dose. Actually, those values do not indicate real IC_{50} or ID_{50} ones. However, as described in the formula, the developmental toxicity is still evaluated thanks to the two ratios IC_{50}/ID_{50} and MD/IC_{50}. If the ID_{50} is higher than the MD, then both ratios will be equal to 1 ($IC_{50}=ID_{50}=MD$) and the score will be -1.109 leading to a negative results (not toxic). However, there are cases where the IC_{50} is equal to the MD but the ID_{50} is triggered and then lower than the MD. In this case, the part $0.714 \times \log(MD/IC_{50})$ of the formula will be erased and the evaluation of the toxicity will be done only by the ratio IC_{50}/ID_{50} (gap between cytotoxicity and Hand1 disruption). For those reasons, the evaluation of developmental toxicity is still possible even though some values are higher than the MD.

The other point was to prove that Hand1-Luc EST is not just simply a cytotoxicity test. The explanation can be given by analyzing in details the discriminant equation of the prediction model:

Two main aspects are analyzed in the equation that are the ratio IC_{50}/ID_{50} (Ratio A) and MD/IC_{50} (ratio B).

Ratio A is used to define the potency of a chemical to disrupt the Hand1 gene compared to cytotoxicity. If there is a gap between the IC_{50} and the ID_{50} values, then at non cytotoxic

concentration, the chemical is affecting the expression of the Hand1 gene and the greater is the ratio A, the greater would be the ability for the chemical to induce genetic disruption that would lead to fetus malformations. However, it is important to keep in mind, that even if the concentration of chemicals is not cytotoxic, a disruption in the gene expression balance can still lead to the fetus death due to organ function suppression or disruption (protein expression or cell migration disruption).

When ratio A is close to 1 ($IC_{50} \simeq ID_{50}$), cells are dying before the effect on the Hand1 gene expression can be observed. In this case, Hand1-Luc EST may be thought not to be sufficient to predict embryotoxicity. However, cytotoxic compounds can also be embryotoxic by preventing the fetus to grow up normally or even lead to fetal death. Toxicology is based on the dose applied to the organism. A dose that kills for example 20% of the cells may induce malformations but not necessary fetus death. However, a dose that kills 80% of the cells has greater chances to induce fetal death.

The other parameter is ratio B which is expressed as the ratio maximum dose (MD), IC_{50}. The maximum dose, as previously defined as the concentration of chemical that dissolves in the assay medium, can be thought to be soluble in blood and therefore have greater chances to reach the fetus and exert its effect. With the same way of thinking as in the case of ratio A, the greater is the gap between the MD and the IC_{50} the greater is the possibility for the chemical to be cytotoxic. If the ratio B is close to one, then the chemical, at the IC_{50} dose, may not necessarily be soluble in blood and thus may have fewer chances to reach the fetus by crossing the blood vessel wall and/or the placental barrier to exert its toxic effect on the fetus.

In other available *in vitro* developmental toxicity tests, the evaluation of chemical precipitation is not described in the protocol. Precipitation may affect the results and thus the results obtain at the concentration precipitation occur may not be the correct ones. Hand1-Luc EST is therefore more accurate concerning this aspect.

In addition, the cut-off criterion to separate positive and negative chemicals has been set to 52% because first, this value was proposed on the ROC curve (see Figure 21) and also because this value best separated the *in vivo* positive and the *in vivo* negative chemicals used for the prediction model.

5.3.2 Predictive capacity

The predictive capacity of the Hand1-Luc EST estimated with seventy-one chemicals is presented in Table 52.

Table 52. Predictive capacity of the Hand1-Luc EST with the most recent prediction model

		Hand1-Luc EST	
		Positive	Negative
In vivo	Positive (44)	21	23
	Negative (27)	5	22
		(26)	(45)

Predictivity parameter	Value
Sensitivity	47.7% (21/44)
Specificity	81.5% (22/27)
Positive predicted value	80.8% (21/26)
Negative predicted value	48.9% (22/45)
Accuracy	60.6% (43/71)

As shown in Table 52, sensitivity and negative predictive value is low which means that Hand1-Luc EST negative predicted chemicals are predicted, as such, with a low confidence. However, the high specificity and positive predictive value shows that a Hand1-Luc EST positive predicted chemical is very likely to be actually positive *in vivo*. This result supports that the Hand1-Luc EST is powerful tool to predict embryotoxic positive chemicals *in vitro*.

Accuracy sets at 60.6% which is considered as insufficient according to Spielmann *et al*., 2006. The other protocols that evaluate embryotoxicity *in vitro* and that undertook a validation process had better accuracy. Indeed, the accuracy reached 70% for the rat limb bud micromass test, 80% for the whole embryo culture test (by combining 2 PMs) and 78% for the original EST (Spielmann *et al*., 2006). However, they only tested twenty chemicals while our predictive capacity is determined by seventy-one chemicals. All those seventy-one chemicals have different mode of action and therefore different effect on development. For chemicals that do not affect Hand1 gene expression will be hard to detect. Furthermore, developmental toxicity involves complex mechanisms and establishing criteria about accuracy to know whether an embryotoxicity test is reliable or not (Spielmann *et al*, 2006) should be revised according to the test definition. The discussion about developmental mechanisms and applicability domain is described in details in part 5.7.

5.4 Transferability

The three participant laboratories, although having expertise in tests involving cell culture and manipulation, were given training. All the participant laboratories got the training the same time at

the same place (SC, Osaka). Hand1-Luc EST aiming to be an easy test with no complicated manipulation requiring special expertise, the training consisted only in showing the test steps and the points to be taken care of. Phase 0, as described in part 3.2, showed reproducible results with low variation of the IC_{50} and ID_{50} values.

5.5 Within-laboratory reproducibility

Within-laboratory reproducibility of the protocol was confirmed in phases 1, 2b, 2c where a test substance was tested in three sets (a set consisting in one dose finding study and one definitive study for phase 1 and for phase 2b, c, one dose finding study and two or three definitive studies). The acceptance criterion for within-laboratory reproducibility was set in the study plan as 75% (see Appendix 31)

As shown in part 3.3.4.2, phase 1 showed a within-laboratory reproducibility of 89% (8/9 tests) and phase 2b, 2c showed a within- laboratory reproducibility of 83.3% (20/24 tests) (part 4.2.4.2) which largely fulfills the acceptance criterion of 75%. The content of the test for phase 1 is however different from phases 2b and 2c. Indeed, at that time, no curve fitting was used and the evaluation of precipitation was not done either. However, it is reasonable to think that the reproducibility would still remain as a high level if the same chemicals were tested since the protocol was improved to avoid the cases that triggered one result that did not fulfill the reproducibility.

The protocol for phase 2 and phase 3 is basically the same except minor changes leading to think that the reproducibility would be better thanks to the improvements applied (see part 4.3.5) in the last version of the protocol.

5.6 Between-laboratory reproducibility

The between-laboratory reproducibility has been tested in phases 2 and 3 (parts 4.1, 4.2 and 4.3). Each compound was tested in one dose finding study and two or three definitive studies and the positive or the negative judgment was decided thanks to the majority judgment. Twelve coded chemicals were tested in phase 2 and sixteen coded chemicals in phase 3.

The between-laboratory reproducibility of phase 2 was 83.3% (10/12 chemicals) and the one for phase 3 was 75% (12/16 chemicals). The reason for discrepancy of some chemical is described in details in parts 4.2.4 and 4.3.4. The protocol has been improved according to the problem raised when analyzing the cases and thus, with the present protocol, although the reproducibility already attained the 75%, it is expected to be much better in the future.

5.7 Applicability domain

As many current *in vitro* available tests, Hand1-Luc EST has limitations. Those limitations can be due to the material used or the biological parameter evaluated.

5.7.1 Limitations due to materials

The evaluation of the cytotoxicity is evaluated thanks to the CellTiter-Fluor™ Cell Viability Assay. Viability of cells is measured via the protease activity happening in living cells. Thus, all the chemicals with a protease inhibitory activity cannot be tested in the Hand1-Luc EST.

5.7.2 Limitations due to the measured biological parameters

Hand1-Luc EST is evaluating developmental toxicity by the disruption of Hand1 gene expression. Therefore, chemicals that do not affect directly or indirectly Hand1 gene expression may not be detected. However, high cytotoxic chemicals would still be predicted as positive (see part 5.3.1).
Figure 25 shows the possible pathways and gene related to Hand1. If a chemical affects one of these genes then it could be detectable via Hand1 gene expression measurement. This scheme has been established thanks to all the papers enumerated in Appendix 38.

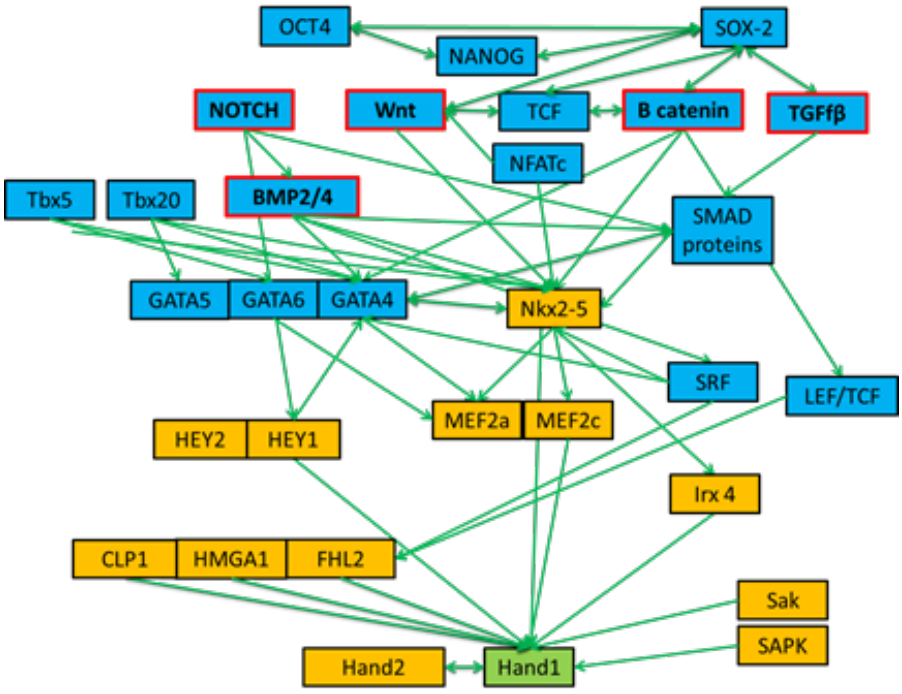


Figure 25 Hand1 gene related pathways (references can be found in Appendix 38)

Chemicals have their own mode of action. Some chemicals could affect some specific pathways (enzyme inhibition), some others could affect a particular physiological function (for example, ion channel inhibition) or some others could affect directly DNA synthesis (enzyme inhibition or intercalating agent). When tested *in vivo*, the appearance of malformations maybe very selective to some special organs while some other may possibly affect a relatively larger number of organs or eventually lead to fetal death. Hand1 gene is reported to be involved in the heart, the limb and the facial bones development (see introduction). Thus, chemicals reported to trigger these organs' malformations may be detected with Hand1. However, it is important to keep in mind that some chemicals may alter genes expressed after 120 hr (5 days) after the beginning of differentiation. In this case, Hand1 may not be able to detect the effect of those chemicals.
As previously emphasized when the other *in vitro* developmental toxicity test were assessed, effects of metabolites should be assessed. Unfortunately, Hand1-Luc EST cannot evaluate metabolites developmental effects of chemicals due to the incapacity of the ES cells to metabolize compounds.
Another point which is important to consider, is the placental barrier. Chemicals that do not cross the barrier will be considered as negative. However, the Hand1-Luc EST does not provide this system and therefore misclassification may occur because of this aspect.

Development is a very complex phenomenon that involves different mechanisms. The seventy-one chemicals we used have their own mode of action and they can disrupt many different genes, some related to Hand1 and some unrelated.
Therefore, given all the points developed above, the accuracy of 60.6% for the Hand1-Luc EST is reliable for the assessment of the developmental toxicity of chemicals in screening stages.

5.7.3 Justifications on misclassified chemicals

Phase 3 revealed six chemicals misclassified and the possible biological reason is detailed below for each of them.

5.7.3.1 False negative chemicals

5.7.3.1.1 Boric Acid

Boric acid can be used as an antiseptic, insecticide, flame retardant, neutron absorber, or precursor to other chemical compounds.
Boric acid is positive *in vivo* but negative in the Hand1-Luc EST. In rats and mice, the most

frequently observed malformations were enlarged lateral ventricles of the brain and agenesis or shortening of the thirteenth rib (Heindel, 1992). Boric acid is also reported to affect the development of the skeleton axis by altering the expression of Hox genes (Hox 4,5,6) leading to cervical malformations (Narotsky, 2004).
Moreover, boric acid can inhibit the activity of HDAC (Histone deacetylases) as can do valproic acid (VPA). This inhibition is usually related to teratogenicity and the malformations have been classified as fusions and homeotic transformation of the axial skeleton segments, quite similar to those induced by VPA (Di Renzo, 2007).
There is no paper showing cardiac teratogenicity for the chemical. Therefore, it can be suggested that boric acid does not affect the development of the heart and thus consequently the expression of the Hand1 gene.
In addition, in a novel neural EST that assesses the expression of two different marker genes (Reln and Tubb) (results soon to be published), boric acid is classified as positive chemical supporting the fact this chemical is actually toxic for neuronal development.

5.7.3.1.2 Dexamethasone

Dexamethasone, synthetic glucocorticoid, is widely used to treat inflammatory conditions such as allergies, skin conditions, ulcerative colitis, arthritis and breathing disorders. This chemical is positive *in vivo* but predicted as negative in the Hand1-Luc EST.
In rodents, dexamethasone elicited cleft palate *in vivo* and prevented neural tube closure *in vitro* (Hansel, 1999). According to in-house data (DNA chip analysis), the glucocorticoid receptor is not yet expressed in the ES cells after 120 hrs of differentiating. Therefore, the effect of dexamethasone on Hand1 may not be observable.

5.7.3.1.3 Dimethadione

Dimethadione is used as an antiepileptic drug. It inhibits T-type calcium channel in the thalamus. It is predicted as negative in the Hand1-Luc EST but positive *in vivo*. The main defects observed in the fetus are facial dysmorphism, cardiac defect and brain retardation (Van Boxtel, 2008). Dimethadione is reported to have a toxic effect on the placenta and not on embryonic tissues that may link to the malformations observed in the fetus (Ozolins *et al*, 2015). This correlates with our observation where dimethadione did not exhibit any cytotoxicity (IC_{50}>1000µg/ml). Another fact seems to be the ability by dimethadione to produce free radicals that could lead to embryotoxicity (Hood, 2006). Hand1 gene being not related to this phenomenon, Hand1-Luc EST may thus not be able to detect the embryotoxicity of dimethadione.

5.7.3.1.4 Methoxyacetic acid

Methoxyacetic acid (MAA) is the active metabolite of the widely used industrial chemical ethylene glycol monomethyl ether.

This chemical is positive *in vivo* but predicted as negative in the Hand1-Luc EST.

In utero, 2-MAA exposure increases the incidence of hydrocephalus, hydronephrosis, heart, tail and limb malformations in species ranging from rodents to nonhuman primates (Brown, 1984; Scott *et al*., 1989). In 1986, another study reports open or irregular fusion of the neural tube and irregular somite segmentation (Yonemoto, 1986). More recent studies show that limb formation is greatly disrupted and the cell cycle is affected by effect on histones proteins (Dayan at al, 2014). Finally, methoxyacetic acid is also reported to disrupt protein kinase pathway, oxidative stress and DNA metabolic process. This chemical has more effect on neural tube, limbs and branchial arches than in the heart (Robinson *et al*, 2010).

In the article of Brown (1984), two types of experiments have been done. The first was to inject MAA in pregnant rat from D8 to D14 and see the effect at D20. The second experiment was to inject MAA to pregnant rats at D10 of pregnancy and see the effect on the fetuses at D12.

In the first experiment, the earlier was the injection, the greater was the fetus' death. The malformations observed were skeletal malformations, hydrocephalus, dilated kidney pelvis and bladder atrophy. At observation days 10 and 12, heart and renal malformations were observed but those are likely to be the results of fetal resorption or growth retardation. For the second experiment, the main malformations observed were incomplete neural suture and forelimb bud malformations.

Given those observations, the effect of MAA seems to be more likely happening in the latter stage of development (D12 to D20). Hand1-Luc EST assessing the gene perturbation at the earlier stages may therefore be not adequate to assess the developmental toxicity of MAA. As described for boric acid, in the neural EST results we are going to submit soon, MAA is also classified as positive stressing again its toxicity for neuronal development. According to the above, Hand1-Luc EST cannot detect the embryotoxicity of MAA because of the early expression of Hand1 gene and MAA specific effect on neuronal and skeletal development.

5.7.3.1.5 Phenytoin

Phenytoin is an antiepileptic drug which can be useful in the treatment of epilepsy. It is used to block sodium channels.

This chemical is positive *in vivo* but predicted as negative in the Hand1-Luc EST.

According to many papers, the main reason for phenytoin to be embryotoxic is the reactive oxygen species generation due to metabolism (Winn *et al*., 2003; Azarbayjani *et al*., 2006). It is also reported

that phenytoin altered genes and pathway related to Hand1 such as TGF-β and the proto-oncogene Wnt-1 (Musselman *et al*., 1994). An inhibition of the Hand1 gene with no cytotoxicity was actually observed. The non cytotoxicity may be explained by the absence of ROS due to the missing of metabolism in the system. The alteration of Hand1 would therefore be linked to TGFβ and Wnt-1 pathways disruption. The fact that the negativity is triggered may be due to the low solubility of the chemical lowering the IC_{50} (=MD)/ID_{50} ratio of the prediction model.

5.7.3.2 False positive chemicals

5.7.3.2.1 Diphenhydramine

Diphenhydramine is an antagonist of the histamine receptor H1. It is predicted as positive in the Hand1-Luc EST but is negative *in vivo*.

The EST also misclassified this chemical suggesting that this type of *in vitro* system may be more sensitive to this compound class (Genschow *et al*, 2004; Panzica-Kelly *et al,* 2012). Indeed, in the Hand1-Luc EST, a relatively high cytotoxicity was observed (IC_{50}~15µg/ml).

In addition, as published in 2012, the positivity of diphenhydramine in the original EST was discussed (Riebeling *et al*, 2012). It seems the origin of the positivity could be the inhibition of potassium voltage gated channel (type H) affecting the beating potential of the differentiated cardiomyocytes. This effect is reversible when diphenhydramine is withdrawn. The Hand1-Luc EST, although it does not assess cardiomyocytes beating, shows a certain inhibition of the Hand1 gene suggesting that the positivity of the chemical in vitro may also have its origin via other mechanisms. The *in vivo* negative results may also be due to metabolism.

6 CONCLUSIONS

The aim of the study was to validate Hand1-Luc EST, in a formal between-laboratory reproducibility study, following the modular approach of Hartung *et al*. (2004).

During the three years of validation, the protocol has been improved and reproducibility (between and within) has been verified. Furthermore, thanks to the many data obtained by the lead lab and the recommendations of the VMT members, a prediction model using the IC_{50}, ID_{50} and MD values was established and improved.

Hand1-Luc EST in its current form has been confirmed to be a transferable and reproducible (>75%) protocol. Also, with a positive predictive value of 80.8% along with an accuracy of 60.6%, the Hand1-Luc EST is defined as a new a powerful tool to detect embryotoxicants and can be used for screening chemicals.

Also, compared to the original EST, Hand1-Luc EST has been recognized to improve the following points:

- Time of the experiment (10 days reduced to 120 hr (5 days))
- Manipulations: it is a very easy protocol that does not require medium change, hanging drops method or subjective microscopic observations
- Chemical quantity: just a few milligrams (50mg maximum) of chemicals are needed
- Prediction model: The prediction model uses only one equation and the solubility of the chemicals is taken into account for the prediction which becomes a way to express the solubility of chemicals in the body fluids. The prediction model has been established with a high number embryotoxicants (71) with different toxicological mechanisms.

With the conclusions asserted by the VMT and considering that developmental toxicity is quite a complicated phenomenon, we consider that the current protocol, mainly focusing on heart developmental toxicity, could be in the future combined with similar protocol evaluating the developmental toxicity of other organs, or some other protocols evaluating metabolism *in vitro*.

This protocol also respecting the 3R and could considerably reduce the use of animals, the VMT recommend that an OECD test guideline for Hand1-Luc EST to be drafted to encourage regulatory acceptance and universal implementation of this validated, reliable, cost- and time-efficient test method.

7 ACKNOWLEDGMENTS

The VMT members would like to thank the participant laboratories for their precious collaboration and efforts to obtain valuable data to allow the present protocol reach the current validation status.

8. REFERENCES

Azarbayjani F, Borg LA, Danielsson BR. (2006): Increased susceptibility to phenytoin teratogenicity: excessive generation of reactive oxygen species or impaired antioxidant defense? Basic Clin Pharmacol Toxicol. 99:305-11.

Brown NA, Holt D, Webb M. (1984): The teratogenicity of methoxyacetic acid in the rat. Toxicol Lett.;22:93–100.

Brown NA. (2002): Selection of test chemicals for the ECVAM international validation study on in vitro embryotoxicity tests. European Centre for the Validation of Alternative Methods. Altern Lab Anim.;30:177-98.

Buesen R, Visan A, Genschow E, Slawik B, Spielmann H, Seiler A. Trends in improving the embryonic stem cell test (EST): an overview. ALTEX. 2004;21:15-22.

Dayan C, Hales BF. (2014): Effects of ethylene glycol monomethyl ether and its metabolite, 2-methoxyacetic acid, on organogenesis stage mouse limbs in vitro. Birth Defects Res B Dev Reprod Toxicol;101:254-61.

Di Renzo F, Cappelletti G, Broccia ML, Giavini E, Menegola E. (2007): Boric acid inhibits embryonic histone deacetylases: a suggested mechanism to explain boric acid-related teratogenicity. Toxicol Appl Pharmacol. 2007 220:178-85.

Firulli, BA., Fuchs, RK., Vincentz, JW., Clouthier, DE., Firulli, AB. Hand1 phosphoregulation within the distal arch neural crest is essential for craniofacial morphogenesis. (2014) Development141:3050-61

Genschow E, Spielmann H, Scholz G, Pohl I, Seiler A, Clemann N, Bremer S, Becker K. (2004): Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. Altern Lab Anim. 32:209-44.

Hansen DK, LaBorde JB, Wall KS, Holson RR, Young JF. (1999): Pharmacokinetic considerations of dexamethasone-induced developmental toxicity in rats. Toxicol Sci ;48:230-9

Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M.,

Hoffmann, S., Roi, A,J., Prieto, P., Sabbioni, E., Scott, L., Worth, A., Zuang, V.(2004): A modular approach to the ECVAM principles on test validity. Altern Lab Anim. 32:467-72.

Heindel JJ, Price CJ, Field EA, Marr MC, Myers CB, Morrissey RE, Schwetz BA. (1992): Developmental toxicity of boric acid in mice and rats. Fundam Appl Toxicol. ;18:266-77.

Hood.R.D. (2006): Developmental and Reproductive Toxicology: A Practical Approach
2nd edition

Karimollah H-T. (2013): Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation Caspian J Intern Med. Spring; 4: 627–635.

Kishi Y, Fujihara H, Kawaguchi K, Yamada H, Nakayama R, Yamamoto N, Fujihara Y, Hamada Y, Satomura K, Masutani M. (2015): PARP Inhibitor PJ34 Suppresses Osteogenic Differentiation in Mouse Mesenchymal Stem Cells by Modulating BMP-2 Signaling Pathway. Int J Mol Sci.;16:24820-38.

Laurie, L.E., Hiroki Kokubo, H., Masataka Nakamura, M., Yumiko Saga, Y., Noriko Funato, N. (2016): The Transcription Factor Hand1 Is Involved In Runx2-Ihh-Regulated Endochondral Ossification PLOS10.1371

Le Coz F, Suzuki N, Nagahori H, Omori T, Saito K. (2015): Hand1-Luc embryonic stem cell test (Hand1-Luc EST): a novel rapid and highly reproducible in vitro test for embryotoxicity by measuring cytotoxicity and differentiation toxicity using engineered mouse ES cells. J Toxicol Sci. ;40:251-61.

Marx-Stoelting P, Adriaens E, Ahr HJ, Bremer S, Garthoff B, Gelbke HP, Piersma A, Pellizzer C, Reuter U, Rogiers V, Schenk B, Schwengberg S, Seiler A, Spielmann H, Steemans M, Stedman DB, Vanparys P, Vericat JA, Verwei M, van der Water F, Weimer M, Schwarz A (2009): Review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect Workshop. M.Altern Lab Anim.;37:313-28.

McFadden, DG., Barbosa, AC., Richardson, JA., Schneider, MD., Srivastava, D., Olson, EN. (2005): The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. Development. 132:189-201

Musselman AC, Bennett GD, Greer KA, Eberwine JH, Finnell RH. (1994): Preliminary evidence of phenytoin-induced alterations in embryonic gene expression in a mouse model. Reprod Toxicol. 8:383-95.

Nagahori H, Suzuki N, Coz FL, Omori T, Saito K (2016) : Prediction of *in vivo* developmental toxicity by combination of Hand1-Luc embryonic stem cell test and metabolic stability test with clarification of metabolically inapplicable candidates. Toxicol Lett. 259: 44-51

Narotsky, M G., N. Wery, B. T. Hamby, D S. Best, N. Pacico, J. J. Picard, F. Gofflot, R J. Kavlock. (2004): Effects of Boric Acid on Hox Gene Expression and the Axial Skeleton in the Developing Rat. Chapter 24, The Skeleton: Biochemical Genetic and Molecular Interactions in Development and Homeostasis. Humana Press Incorporated, Totowa, NJ, 17:361-372,.

Ozolinš TR, Weston AD, Perretta A, Thomson JJ, Brown NA. (2015): Dimethadione embryotoxicity in the rat is neither correlated with maternal systemic drug concentrations nor embryonic tissue levels. Toxicol Appl Pharmacol. 15;289:89-97.

Panzica-Kelly JM1, Brannen KC, Ma Y, Zhang CX, Flint OP, Lehman-McKeeman LD, Augustine-Rauch KA (2013): Establishment of a molecular embryonic stem cell developmental toxicity assay. Toxicol Sci. 131:447-57.

Reamon-Buettner, S,M., Ciribilli, Y., Traverso, I., Kuhls, B., Alberto Inga, A., and Borlak, J.(2009): A functional genetic study identifies HAND1 mutations in septation defects of the human heart. Hum. Mol. Genet. 18:3567-3578.

Riebeling C., Hayess K., Peters AK., Steemans M., Spielmann H., Luch A., Seiler AE.(2012): Assaying embryotoxicity in the test tube: current limitations of the embryonic stem cell test (EST) challenging its applicability domain. Crit Rev Toxicol. 42:443-464.

Robinson JF, van Beelen VA, Verhoef A, Renkens MF, Luijten M, van Herwijnen MH, Westerman A, Pennings JL, Piersma AH. (2010): Embryotoxicant-specific transcriptomic responses in rat postimplantation whole-embryo culture. Toxicol Sci. 118:675-85.

Scott WJ, Fradkin R, Wittfoht W, Nau H. Teratologic potential of 2-methoxyethanol and transplacental distribution of its metabolite, 2-methoxyacetic acid, in non-human primates. (1989) Teratology. 39:363–373.

Spielmann, H., Pohl, I., Doering, B., Liebsch, M., Moldenhauer, F. (1997). The embryonic stem cell test, an in vitro embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. In Vitro Toxicol. 10, 119–127.

Spielmann H, Seiler A, Bremer S, Hareng L, Hartung T, Ahr H, Faustman E, Haas U, Moffat GJ, Nau H, Vanparys P, Piersma A, Sintes JR, Stuart J. (2006): The practical application of three validated in vitro embryotoxicity tests. The report and recommendations of an ECVAM/ZEBET workshop (ECVAM workshop 57). Altern Lab Anim.34:527-38.

Suzuki N, Ando S, Sumida K, Horie N, Saito K. (2011a): Analysis of altered gene expression specific to embryotoxic chemical treatment during embryonic stem cell differentiation into myocardiac and neural cells. J Toxicol Sci.36:569-85.

Suzuki N, Ando S, Yamashita N, Horie N, Saito K. (2011b): Evaluation of novel high-throughput embryonic stem cell tests with new molecular markers for screening embryotoxic chemicals in vitro. Toxicol Sci.;124:460-71.

Suzuki N, Yamashita N, Koseki N, Yamada T, Kimura Y, Aiba S, Toyoizumi T, Watanabe M, Ohta R, Tanaka N, Saito K. (2012): Assessment of technical protocols for novel embryonic stem cell tests with molecular markers (Hand1- and Cmya1-ESTs): a preliminary cross-laboratory performance analysis. J Toxicol Sci.37:845-51.

Van Boxtel C.J., Santoso B. (2008): Edwards Drug Benefits and Risks: International Textbook of Clinical Pharmacology - Revised 2nd Edition

Winn LM, Kim PM, Nickoloff JA. (2003): Oxidative stress-induced homologous recombination as a novel mechanism for phenytoin-initiated toxicity. J Pharmacol Exp Ther. ;306:523-7.

Yonemoto J, Brown NA, Webb M. (1984): Effects of dimethoxyethyl phthalate, monomethoxyethyl phthalate, 2-methoxyethanol and methoxyacetic acid on post implantation rat embryos in culture. Toxicol Lett 21:97–102.