The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.1

Example illustrating endpoint vs. endpoint correlation using ToxCast data

Outlook

• Background

- Objectives
- The exercise
- Workflow

Background

This presentation is designed to introduce the user with:

- ToxCast database as part of the Toolbox database
- Illustration of endpoint vs. endpoint correlations using:
 - ToxCast data
 - ToxCast and Estrogen receptor data

Outlook

- Background
- Objectives
- The exercise
- Workflow

Objectives

• This presentation demonstrates endpoint vs. endpoint correlations using ToxCast and Estrogen receptor data

Outlook

- Background
- Objectives
- The exercise
- Workflow

The exercise

- Illustration of endpoint data correlations using the ToxCast and estrogen binding data between two type data:
 - > AC50 vs. AC50 endpoints associated with different test type
 - > AC50 vs. Estrogen receptor binding data

Outlook

- Background
- Objectives
- The exercise
- Workflow

Workflow

- The Toolbox has six modules which are typically used in a workflow:
 - Chemical Input
 - Profiling
 - Endpoints
 - Category Definition
 - Filling Data Gaps
 - Report
- In this example we will use the modules in a different order, tailored to the aims of the example.

Outlook

- Background
- Objectives
- The exercise
- Workflow
 - Load ToxCast database

ToxCast database Loading database



The OECD QSAR Toolbox for Grouping Chemicals into Categories

ToxCast database Sidebar of database relevancy

Once the endpoint is selected, the relevant databases are getting green highlighted.



Click on the level ToxCast endpoint tree;
 Click "Options" and ask for Legend;

2. The database is getting green highlighted;

ToxCast database Data gathering

QSAR TOOLBOX	P Image: Ima	efinition > Data Ga	10 00 00 p Filling > Repo	Dort					
Data Import Export Gather 3 UD6 IUCLID6	1								
lents	Filter endpoint tree	1	2	3	4	5	6	7	8 9
▲ Document 1	Structure	, Q	н	** 9	n-ton	sox or €	H000000000000000000	HC	"H
Databases Options f Select All Unselect All Invert Industry of gene capitosian uneaver	Structure info Parameters Thysical Chemical Properties Environmental Fate and Transport Ecotoxicological Information Human Health Hazards Acute Toxicity Bioaccumulation Carcinogenicity Developmental Toxicity / Teratogenicity Genetic Toxicity Immunotoxicity Irritation / Corrosion Neurotoxicity Photoinduced toxicity Repeated Dose Toxi Awy SW AOP								
Keratinocyte gene expression LuSens									
Micronucleus ISSMIC Micronucleus OASIS	ACEA (600/660)			M: 0.0601 mg/L		M: 7.06 mg/L		M: 6.8 mg/L	M: 2.84 mg/L
MUNRO non-cancer EFSA	+ Apredica (425/2653)					M: 5.84 mg/L			
REACH Skin sensitisation database (no	+ Attagene (1374/11710)		M: 4.33 mg/L	M: 0.756 mg/L	M: 0.627 mg/L	M: 17.9 mg/L	M: 16.2 mg/L	M: 0.67 mg/L	M: 9.86 mg/L
Receptor Mediated Effects	H BioSeek (971/21906)	M: 0.127 mg/L	M: 3.41 mg/L		M: 4.17 mg/L	M: 0.674 mg/L	M: 2.69 mg/L		M: 1.14 mg/L
Rodent Inh:	(1475/6890)	M: 2.72 mg/L		M: 6.32 mg/L	M: 1.61 mg/L	M: 1.53 mg/L	14.00400 #	M: 12.4 mg/L	14.0.445
Skin Irritatio 🙎 🔍	(975/8054)		M: 2.43 mg/L	M. 0 200 //	M: 0.295 mg/L	M: 0.209 mg/L	M: 0.0122 mg/L	M: 8.61 mg/L	M: 0.415 mg/L
Skin Sensi	Udyssey Thera (969/2794)		M: 6.89 mg/L	M: 0.299 mg/L	M: 9.54 mg/L	M: 2.57 mg/L		M: 1/.1 mg/L	M: 14.1 mg/L
ToxCastDB	Lucieita da Basa dustina (2/2)								
Toxicity Japan MHLW	- Ioxicity to Reproduction								
Toxicity to reproduction (ER)									
ToxRefDB US-EPA									

1. Go to **Data**; 2. Check **ToxCast database**; 3. **Click** "Gather"; 4. The data appears on datamatrix on the level "ToxCast"

Outlook

- Background
- Objectives
- The exercise
- Workflow
 - Load ToxCast database
 - ToxCast database overview

ToxCast database Background

- A major part of EPA's CompTox research is the ToxCast[™] project. ToxCast is a multi-year project launched in 2007 that uses automated chemical screening technologies (called "high-throughput screening assays") to expose living cells or isolated proteins to chemicals. The cells or proteins are then screened for changes in biological activity that may suggest potential toxic effects. These innovative methods have the potential to limit the number of required laboratory animal-based toxicity tests while quickly and efficiently screening large numbers of chemicals.
- ToxCast has evaluated over 2,000 chemicals from a broad range of sources including: industrial and consumer products, food additives, and potentially "green" chemicals that could be safer alternatives to existing chemicals. Chemicals were evaluated in over 700 high-throughput assays that cover a range of high-level cell responses and approximately 300 signaling pathways.
- ToxCast results are contributed to the federal agency collaboration called Toxicity Testing in the 21st Century (Tox21). Tox21 pools chemical research, data and screening tools from multiple federal agencies including the National Toxicology Program. So far, Tox21 has compiled high-throughput screening data on nearly ten thousand chemicals.

Outlook

- Background
- Objectives
- The exercise
- Workflow
 - Load ToxCast database
 - ToxCast database overview
 - Correlation of data background

Correlation of endpoint data Background

- This functionality introduce the user with opportunity to analyze correlations between selected gap filling endpoint (endpoint used for prediction) and other endpoint data.
- It is applicable for correlation analysis of data presented in ordinary, interval or ratio scale.
- If correlated data are measured in interval or ratio scale they are transformed in ordinary scale and the strength of the correlation is estimated by Spearman correlation coefficient.
- Basically, this functionality provides a correlation between target endpoint (this is the initial endpoint selected by the user) displayed on ordinate axis (Y-axis) and other endpoint data displayed on abscissa (X-axis).

Correlation of endpoint data Spearman coefficient factor

- Spearman's rank correlation coefficient is a nonparametric rank statistic proposed by Charles Spearman as a measure of the strength of an association between two variables. It assesses how well the relationship between two variables can be described using a monotonic function.
- Spearman correlation coefficient could be used for exploring the covary between:
 - two ranked variables
 - one measurement variable and one ranked variable (in this case, the measurement variable need to be to converted to ranks)
- Spearman correlation varies from -1 to +1 and the interpretation of the coefficient factor is provided below:
 - 0.00 0.19 very weak correlation
 - 0.20 0.39 weak correlation
 - 0.40 0.59 moderate correlation
 - 0.60 0.79 strong correlation
 - 0.80 1.0 very strong

Outlook

- Background
- Objectives
- The exercise
- Workflow
 - Load ToxCast database
 - ToxCast database overview
 - Correlation of data background
 - Types endpoint correlations

Types endpoint correlations are as follows:

- Continuous vs. continuous
- Categorical vs. categorical*:
 - ✓ Categorical vs. categorical
 - ✓ Categorized continuous vs. categorical
 - ✓ Categorized continuous vs. categorized continuous

*All type categorical vs. categorical correlations are not illustrated in this presentations. These type correlations are shown in presentation "Tutorial 13 TB4.1. Example illustrating endpoint vs. endpoint correlation for apical endpoints"

Outlook

- Background
- Objectives
- The exercise

Workflow

- Load ToxCast database
- ToxCast database overview
- Correlation of data background

• Types endpoint correlations

• Continuous vs. continuous

Types endpoint correlations Continuous vs. continuous

- The aim of this type correlation is to illustrate how continues type endpoint data or so called ratio data correlates each other (e.g.LC50 vs. EC50 data)
- In this example we will illustrated how AC50 data associated with two different test assays extracted from ToxCast DB correlates each other:
 - NCGC Reporter Gene Assay ERa Agonist, Estrogen receptor 1 (assay 1)
 - Tox21_Era_BLA_Agonist_ch2 (assay 2)
- Step by step workflow is presented on next few slides. Summary of the workflow steps are provided below:
 - Gather experimental data (step 1)
 - Define target endpoint (step 2)
 - Enter Gap filling (step 3)
 - Change default X-descriptor (logKow) with AC50 data (step 5)

Continuous vs. continuous

Gather experimental data – step 1

QSAR TOOLE		→ Fring → Data 1	ition Data G	ap Filling > Repo	rt					
Data Import Gather 3	Export IUCLID6									
Documents		Filter endpoint tree	1	2	3	4	5	6	7	8 9
▲ Document 1		Structure	Q	H ₂ C	******** 9	AC	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	11/2	Har and Andrew	X
		Parameters								
		Physical Chemical Properties								
		Environmental Fate and Transport								
		Ecotoxicological Information								
		Human Health Hazards								
		- Acute Toxicity								
Databases		Bioaccumulation								
Options 🖌		Developmental Toxicity / Teratogenicity								
f Select All Unselect All	l Invert	Genetic Toxicity								
Rodent Inhalation Tox	kicityDatabase \land									
Skin Irritation		Irritation / Corrosion								
Skin Sensitization ECET	100	Neurotoxicity								
✓ ToxCastDB		Photoinduced toxicity								
Toxicity Sapan in ILW	2	Repeated Dose Toxicity								
ToxRefDB US-EPA		Sensitisation AW SW AOP								
Transgenic Rodent Da	tabase	ToxCast								
Yeast estrogen assay	database	Toxicity to Reproduction								
<	>									

Follow the steps if you already load Toxcast data on data matrix. 1. **Go** to "Data" 2. **Select "**ToxCast" DB 3. **Click** "Gather"

Continuous vs. continuous

Gather experimental data – step 1

Toxicity information on the target chemical is electronically collected from the selected datasets.

A window with "Read data?" appears. Now the user could choose to collect "all" or "endpoint specific" data.



Continuous vs. continuous

Gather experimental data - step 1

QSAR TOOLBOX	Input ▶ Profiling ▶ Data ▶ Cates	Jory definition Data 6	1010 100 Sap Filling ► Rep	Port	
Data Import Export Import Import Import Import					
Documents	Filter endpoint tree	1	2	3	4
Ā Document 1 ToxCastDB	Structure) vszré	HO CHB	-300	-250
	Structure info				
	Parameters				
	Physical Chemical Properties				
	Environmental Fate and Transport				
	Ecotoxicological Information				
	Human Health Hazards	L			
	Acute loxicity Biogeographicity	•			
Databases		•	Q		×
Options 🖌	Developmental Toxicity / Teratogenicity	-	+		
f Select All Unselect All Invert	Genetic Toxicity	-	54669 point	s added across 1813 che	micals.
Skin Irritation	Immunotoxicity	-	+ .		
Skin Sensitization	Irritation / Corrosion		+		
test Iu6	Neurotoxicity				OK
✓ ToxCastDB	Photoinduced toxicity				\prec
Toxicity Japan MHLW Toxicity to reproduction (FR)	Repeated Dose Toxicity				
ToxRefDB US-EPA	Sensitisation AW SW AOP			1	
Transgenic Rodent Database	Taviaita ta Base dustian (1813/54669)	M: 2.06 mg/L	M: 0.0039 mg/L	M: 0.0545 n	1 mg/L
ZEBET database	Toxicity to Reproduction	-			F
< >	Toxiconneucs, metabolisin and Distribution	4			
Options 4					
f Select All Unselect All Invert					
Canada DSL					
COSING					
ECHA PR					
EINECS					
HPVC OECD					
NICNAS					
REACH ECB					
TSCA			· · · ·		
LIC UDV Challongo Drogram		nco tho	WIN A AVI		

NICNAS

Th

Types endpoint correlations

Continuous vs. continuous Gather experimental data – step 1

QSAR TOOLBOX	Input F []	F ► Categor	y definition Data Ga	10 00 p Filling > Repo	rt		
Data Import Export Gather Import IUCLID6							
 Documents 	Filter endpoint tree		1	2	3	4	5
 ▲ Document 1 ■ ToxCastD8 	Structure		2-2-8-4E	HO CH3			
	Structure info	ſ					
	+ Parameters						
	Physical Chemical Properties						
	Environmental Fate and Transport						
	+ Ecotoxicological Information	-					
	- Human Health Hazards	-					
	Bioaccumulation	-					
 Databases 		•					
Options 🖌	Developmental Toxicity / Teratogenicity	-					
f Select All Unselect All Invert	Genetic Toxicity	•					
Skin Irritation	Immunotoxicity	1					
Skin Sensitization	Irritation / Corro	1					
test Iu6	Neurotoxicity						
✓ ToxCastDB	Photoinduced t						
Toxicity Japan MHLW Toxicity to reproduction (FR)	Repeated Dose						
ToxRefDB US-EPA	Sensitisation AV	WSW AOP					
Transgenic Rodent Database	ToxCast		14 04 0 4	14 0 0000 //	N. 0.000505 //	N 001 /	14.0.22 #
ZEBET database		(600/660)	M: 21.2 mg/L	WI: 0.0039 mg/L	WI: 0.000585 mg/L	N: 8.01 mg/L	WI: 9.32 mg/L
	L± Apredica	(425/2053)	M: 2.06 mg/l		M:0.01/11 mg/l	M: 10.9 mg/L	M: 2.06 mg/L
		(071/21006)	wi. 2.00 mg/ c		with out of the time of time o	M: 2.44 mg/c	M: 2.00 mg/L
Inventories	TH NCGC	(1475/6890)	M: 23.3 mg/l	M: 0.000106 mg/l	M: 0.00017 mg/l	M: 0.242 mg/L	M: 16.5 mg/L
f Select All Unselect All Invest		(975/8054)	M: 0.0272 mg/L		M: 0.0121 mg/L	M: 0.185 mg/L	M: 7.8 mg/L
Canada DSI	Odyssey Thera	(969/2794)	M: 19.8 mg/L	M: 2.46 mg/L	M: 0.0545 mg/L	M: 0.107 mg/L	M: 13.8 mg/L
COSING	Undefined Assay provider	(2/2)		- 10			
DSSTOX	Toxicity to Reproduction	(_/_/					
ECHA PR	Toxicokinetics, Metabolism and Distribution	ion					
HPVC OECD		4					L

1. ToxCast data has been loaded on datamatrix in a separate node of "Endpoint tree" called "ToxCast"

Continuous vs. continuous

Define target endpoint – step 2

QSAR TOOLBOX	Dut > Profiling > Data > Category definition > Data Gap Filing > Report	
Data Import Export Gather Import IUCLID6 IUCLID6		
Documents	Filter endpoint tree 1 2	3 4 5
 ▲ Document 1 ■ ToxCastD8 	Structure	
	Photoinduced toxicity	
	Repeated Dose Toxicity	
	ToxCast	
	E ACEA (600/660) M: 212 mg/l M: 0.0039 mg/l	M: 0.000585 mg/l M: 8.01 mg/l M: 9.32 mg/l
	T Apredica (425/2653)	M: 10.9 mg/L M: 27.5 mg/L
		M: 0.0141 mg/L M: 2.44 mg/L M: 2.06 mg/L
Databases		M: 4.71 mg/L M: 9.85 mg/L
Options 4		
f Select All Unselect All Invert	- DNCGC Reporter Gene Assay ERa Agonist	
Skin Irritation	Homo sapiens	
Skin Sensitization	- estrogen receptor 1	
Skin sensitization ECETOC	AC50 (374/505)	M: 0.00017 mg/L M: 9.12E-05 mg/L M: 5.42 mg/L
✓ ToxCastDB	NCGC Reporter Gene Assay ERa Antagonist (487/559) M: 19.1 mg/L M: 0.000531 mg	/L M: 7.62 mg/L
Toxicity Japan MHLW		n
Toxicity to reproduction (ER)		/L
Transgenic Rodent Database		M: 0.242 mg/L M: 10.3 mg/L
Yeast estrogen assay database	Tox21_AR_BLA_Agonist_ctiz (0//0/)	M: 0.105 mg/L
ZEBET database	Tox21_AR_BLA_Antanonist_ratio (05/05)	M: 3 mg/l
	Tox21 AR BLA Antagonist viability (207/207) M: 14.4 mg/L	
Inventories	Tox21_AR_LUC_MDAKB2_Agonist (90/90) M: 2.86 mg/L	M: 0.152 mg/L
Options	Tox21_AR_LUC_MDAKB2_Antagonist (56/56)	
f Select All Unselect All Invert	Tox21_AR_LUC_MDAKB2_Antagonist_viability (291/291) M: 0.000106 mg	/L

The target endpoint is AC50 associated with assay "NCGC Reporter Gene Assay ERa Agonist" 1. **Click** on the cell related to the investigated endpoint, below the first chemical of datamatrix

Continuous vs. continuous Define target endpoint – step 2

QSAR TOOLEOX	P ↓ ↓ Profiling > Data > Categ	ory definition	1				
G F Workflow	w S stomated						
Documents	Filter endpoint tree	1 2	3	4 5	6	7	8
À Document 1 ≣ ToxCastD8	Structure	NO NO	жн */~~~~ Р		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	46	
		1	m. 0.000 r mg/ c	1916	noo mg/c	Mi. Olo Hig/ L	m. 1.04 mg/ t
	+ Apredica (425/2653)		M:	1.69 mg/L		
	+ Attagene (1374/11710) M: 0.88 mg/L	M: 0.113 mg/L	M: 0.627 mg/L M:	11.8 mg/L M: 16.2 mg/L	M: 0.033 mg/L	M: 10.3 mg/L
	BioSeek (971/21906) M: 0.127 mg/L M: 0.16 mg/L	×	M: 0.464 mg/L M:	0.539 mg/L M: 0.243 mg/L		M: 0.663 mg/L
	NCGC Reporter Gene Assay ERa Agonist						
	Homo sapiens						
	estrogen receptor 1						
	AC50 (374/50		M: 0.156 mg/L	M:	0.478 mg/L		
			M: 10.6 mg/l	M:	0.727 mg/L	M: 5.52 mg/L	
	→ Tox21_AhR (237/237	Possible data inconsistency		^ <u>/</u>			
	Tox21_AhR_viability (319/319) Gan filling scale/unit					
	Tox21_AR_BLA_Agonist_ch1 (439/439	log(1/mol/L)					
	Tox21_AR_BLA_Agonist_ch2 (67/67	H ΟμM					
	Tox21_AR_BLA_Agonist_ratio (89/89	эн <u></u>					
 Data Gap Filling Settings 	Tox21_AR_BLA_Antagonist_ratio (150/150	Data 505/505; Chemicals 374/374			4.52 #		
Only endpoint relevant	Tuesd AB LUG ADDAVED Activity (207/207	H		M:	1.53 mg/L	IVI: 12.4 mg/L	
✓ Only chemical relevant	THI TOX21_AR_LUC_MDAKB2_Agonist (90/90	<u></u>	ОК	Cancel			
At this position:	TE Tax21_AR_LUC_MDAKB2_Antagonist (56/56		M: 15 mg/l	\longrightarrow			
At any position.	Tox21_AR_LUC_WDAKB2_AntagonIst_V(291/291		Wi: 15 mg/L	A	2 2 mg/l		
Select a cell with a rigid (bold) path	Tox21_Aromatase_Inhibition_viability (202/202			4	5.5 mg/c		
Standartized workflows 0	Tox21_AutoEluor_HEK203_Cell_rod (303/303						+
	Tox21_AutoFluor_HEK203_Gell_IEd (///						
	Tox21_AutoFluor_HEK203_Media_prop (5/5						+
	Tox21_AutoFluor_HEK203_Media_red (7/7						
	Tox21_AutoFluor_HERG2_Cell_blue(10/49	()					+
	Tox21_AutoFluor_HEPG2_Cell_pide (18/18						+

Click on "Data Gap Filling"; 2. Highlight the empty cell next to the AC50 endpoint associated with illustrated assay: "NCGC Reporter Gene Assay ERa Agonist"
 Select "Trend analysis";
 A window alerting you for data inconsistencies appears. Keep it as it is. Click "OK".

Continuous vs. continuous Define target endpoint – step 2

QSAR TOOLBOX	vput > Profiling > Data > Cat	egory definition 🕨 Data (itoŭ	eport					
Irend analysis kead across (Q)SAK Standardized Au	Eilter endpoint tree	1 [target]	3	5	0	20	28	40	41
Occuments ument 1 IoxCastD8 Finter GF(TA) with 357 chemicals, 486 data poin	Structure		۰٬۰۰۰۰۰۰۰۰ و.	****	0,00	HI® ⁻⁰⁴			HO CH3
	Mammalia (mammalis) Horno sapiens estrogen receptor 1								
	AC50 (356/4	86)	M: 0.156 mg/L	M: 0.478 mg/L	M: 2.54 mg/L	M: 2.31 mg/L	M: 2.02 mg/L	M: 7.53 mg/L	M: 2.77 mg/L
	NCGC Reporter Gene Assay ERa Antag(120/1	42)	M: 10.6 mg/L	M: 0.727 mg/L	M: 0.000358 mg/L			M: 1.97 mg/L	
		83)			NA 0.000350 //				N 0050 //
	Tax21 AR RLA Apprint ch1 (115/1	53) 15) M:0.367 m			M: 0.000358 mg/L			M: 8 57 mg/l	M: 0.859 mg/L
	Tox21_AR_BLA_Agonist_ch1 (113/1	15) W. 0.507 II	nation			×		Wi. 6.57 Hig/E	
	Tox21_AR_BLA_Agonist_ratio (45/	45)						M: 9.57 mg/L	
	Tox21_AR_BLA_Antagonist_ratio (46/	46) 19 observ	ed values for 18 chemi	cals were excluded due	to missing X descriptor val	ue(s)			
	Tox21_AR_BLA_Antagonist_viability (66/	56)						M: 0.12 mg/L	
	Tox21_AR_LUC_MDAKB2_Agonist (39/	39) M: 4.8 mg/							
< >	Tox21_AR_LUC_MDAKB2_Antagonist (10/	10)			OK				
Data Gap Filling Settings	Tox21_AR_LUC_MDAKB2_Antagonist_vial(62/	52)	M: 15 mg/L		M: 0.000358 mg/L	\wedge			
	Tox21_Aromatase_Inhibition (38/	38)		M: 3.3 mg/L				Mt 0.22 //	Mt 102 ms //
Only endpoint relevant	Tax21_Aromatase_Inhibition_Viability (62/	(6)			- 4			WI: 6.25 mg/L	W: 1.92 mg/L
	Tox21_AutoFluor_HEPG2_Cell_blue (6	/6)							
At this position:	Tox21 AutoFluor HEPG2 Media blue	/5)							
Select a cell with a rigid (bold) path	Tox21_ELG1_LUC_Agonist (38/	38)			M: 6.76 mg/L				
Automated workflows 0 Standartized workflows 0	Tox21_ELG1_LUC_Agonist_viability (52/	52)			M: 16.4 mg/L				
	Tox21_ERa_BLA_Agonist_ch1 (95/	95)	M: 0.945 mg/L	M: 0.357 mg/L				M: 7.84 mg/L	
	Tox21_ERa_BLA_Agonist_ch2 (214/2	14)	M: 0.632 mg/L					M: 3.25 mg/L	M: 4.45 mg/L
	Tox21_GR_BLA_Agonist_ch1 (81/	81)	M: 3.62 mg/L					M: 1.75 mg/L	M: 0.000102 mg/L
	Tox21_GR_BLA_Agonist_ch2 (6)	/6)							
	Tox21_GR_BLA_Agonist_ratio (68/	58)	M: 9.81 mg/L					M: 10.2 mg/L	
	Tox21_GR_BLA_Antagonist_ratio (25)	77)	M: 4.51 mg/l					M: 0.218 mg/l	
	Tox21_OrC_DEPCPrintagonist_vidulinty (///	/0)							
The message inf excluded from g	forming the user for ap filling due to mis	how masing X d	any che lescript	micals or value	with exp es appea	erimen Irs. 1. C	tal data lick "OK"	are ;	

Continuous vs. continuous Enter Gap filling – step 3

	• Data • Catagory definition	X 0 5 4 0
Workflow		The OECD QSAR Toolbox for Grouping Chemicals into Categories
Trend analysis Read across (Q)SAR Standardized Automated		Developed by LMC, Bulgaria
▲ Documents ▲ Document 1 ▲	Filter endpoint tree I target] 3 4 5 6 10 14 Structure $3 \times 3^{\circ}$ $3 \times 3^{\circ}$ $3 \times 3^{\circ}$ $5 \times 3^{\circ}$ </th <th></th>	
Data Gap Filling Settings	L AC50 (356/48) M: 0.00017 mg/L M: 0.0016 mg/L M: 152 mg/L M: 443 mg/L M: 433 mg/L M: 432 mg/L M: 433 mg/L M: 432 mg/L M: 432 mg/L M: 433 mg/L M: 432	M: 13.3 mg/L
Only endpoint relevant Only chemical relevant At this position: Select a cell with a rigid (bold) path Automated workflows O Standartized workflows O	Descriptors Prediction Adequecy Cumulative frequency Residuals Statistics Descriptors Descriptors Prediction Adequecy Cumulative frequency Residuals Statistics Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors Descriptors D	Select / filter data Gap filling approach Descriptors / data Model/QSAR Calculation options Visual options Information Miscellaneous

Enter Gap filling applying trend analysis. Trend analysis is applied because the target endpoint is in continues range of data and there is enough data to build a linear regression.1. Data Gap filling stage 2. Trend analysis approach is applied 3. AC50 endpoint related to ER enzyme

assay ia plotted on Y-axis 4. Pay attention that default descriptor displayed on X-axis is log Kow.

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with AC50 data – step 4

SAR TOOLBOX	► Profiling ► Data ► Category definit	01010 01 00 01 00 01000 00 00 00 00 00 00 00 00 00 00 0			X 0 5 4 0
Gap Filling Workflow					The OECD QSAR Toolbox for Grouping Chemicals into Categories
	Filter endpoint tree	1 [target] 3 5	9 20	28 40 4	Developed by LMC, Bulgaria
Cocuments Documents Documents Cocuments Cocuments Cocument2 # CANS:62533 CAS:62533 CAS:6253 CAS:62533 CAS:6253 CAS:625 CAS:625 CAS:6253 CAS:625 CAS	Structure	~ ~~ ~~	0,0°°		
Document 3	AC50 (💽 S	elect endpoint descriptor	– – × –	M: 2.02 mg/L M: 7.53 mg/L	M: 2.77 mg/L 🗧 M: 0.75 mg/L
ToxCastDB ToxCast	NCGC Reporter Gene Assay ERa Antag(Below are the tree positions which contain experie	nental data for analogues.	M: 1.97 mg/L	
	HTTox21_ARR	Select one of them to use it as a X	descriptor:		M: 0.859 mg/L
	Tox21_AR_BLA_Agonist_ch1 (4 H	uman Health Hazards (357/16310)		M: 8.57 mg/L	
	Tox21_AR_BLA_Agonist_ch2	▷ ACEA (189/230)	\sim		
	Tox21_AR_BLA_Agonist_ratio	 Apredica (110/813) Attagene (200/2040) 		3 M: 9.57 mg/L	
	Tox21 AR BLA Antagonist viability	 BioSeek (210/6558) 		M: 0.12 mg/L	
	Tox21_AR_LUC_MDAKB2_Agonist	 NCGC (357/2569) Novascreen (233/2078) 			
	Tox21_AR_LUC_MDAKB2_Antagonist	 Odyssey Thera (238/1122) 			
	Tox21_AR_LUC_MDAKB2_Antagonist_vi				M
	Tox21_Aromatase_Inhibition			M: 8.23 mg/L	M: 1.92 mg/L
	Tox21_AutoFluor_HEK293_Media_blue				
Data Gap Filling Settings	Tox21_AutoFluor_HEPG2_Cell_blue		OK Count		
Only endpoint relevant Only chemical relevant	I LEI Tav21 AutaEluar UEDQ3 Madia blua				>
At this position:	Descriptors	Trend analysis pro	ediction for AC50, based on 356 values		Select / filter data
Select a cell with a rigid (bold) path	Desidentia	Model equation: AC50 = 4.	1971 22 (±0.176) + 0.0382 (±0.0422) * log Kow, log(1/mol/L)		Gap filling approach
Automated workflows 0	Fredicion				
Stanuardized Worknows 0	Adequacy				Descriptors / data
	Completing for some state of the solution of t				Change descriptor units
					Edit descriptor options
	Residuals				Salact and point trea descriptor
	Statistics		· · · · · · · · · · · · · · · · · · ·		Madel/OSAP
	•				MODEVQSAR
					Calculation options
					Visual options

1. **Click** on "Descriptors /data"; 2. **Go** on "Select endpoint tree descriptor"; 3. A window with arranged "Endpoint data tree" appears. Expand the endpoint tree;

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with other AC50 data – step 4



1. **Click** on "NCGC" node to open the sub-nodes; 2. **Select** endpoint, which will be placed on X-axis circled in red box; point the mouse on the level of AC50 (214/214); 3. **Click** "OK" button

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with other AC50 data – step 4

QSAR TOOLEOX	F In Category definition	01010 01 0 10100 n	► Report							X 0 5 6 8
Gap Filing Workflow										The OECD QSAR Toolbox for Grouping Chemicals into Categories
	Filter endpoint tree	1 [target]	3	5	9	20	28	40	41	Developed by LMC, Bulgaria
Document 1 Document 2 Construction Document 2 Coss 6033 Coss 6033	Structure		۰٬۰۰۰۰۰	°,2°,×	0	HEQ ^{-CH}	C Nat	npcC	HO CH3	Na*
A Document 3	AC50 (356/486		M: 0.156 mg/L	M: 0.478 mg/L	M: 2.54 mg/L	M: 2.31 mg/L	M: 2.02 mg/L	M: 7.53 mg/L	M: 2.77 mg/L	M: 0.75 mg/L
✓	NCGC Reporter Gene Assay ERa Antag(120/142		M: 10.6 mg/L	M: 0.727 mg/L	M: 0.000358 mg/L			M: 1.97 mg/L		
					M: 0.000358 mg/l				M: 0.859 mg/l	
	Tox21 AR BLA Agonist ch1 (115/115)	M: 0.367 mg/L	M: 7.66 mg/L		ini ciccosso nigi c			M: 8.57 mg/L	in clossing/c	
	Tox21_AR_BLA_Agonist_ch2 (36/36									
	Tox21_AR_BLA_Agonist_ratio (45/45	Possible dat	inconsistency		×			M: 9.57 mg/L		
	Tox21_AR_BLA_Antagonist_ratio (46/46							M: 0.12 mg/l		
	Tox21_AR_LUC_MDAKB2_Agonist (39/39	M: - Gap filling scale	e/unit							
	Tox21_AR_LUC_MDAKB2_Antagonist (10/10	0 μM	-)							
	Tox21_AR_LUC_MDAKB2_Antagonist_vial(62/62	D-1- 214/214 (h		ig/L					M:
	Tox21_Aromatase_Inhibition viability (62/62	Data 214/214; C	nemicais 214/214					M: 8.23 mg/L	M: 1.92 mg/L	
× 7	Tox21_AutoFluor_HEK293_Media_blue (6/6			ОК -	Cancel	<u> </u>				
Data Gap Filling Settings	Tox21_AutoFluor_HEPG2_Cell_blue (6/6									
 ✓ Only endpoint relevant ✓ Only chemical relevant 	I I Tav21 AutoEluor UEDG2 Modia blua (6/6									>
At this position:	Descriptors			Trend analysis predict	ion for AC50, based on 3	o6 values				Select / filter data
Select a cell with a rigid (bold) path Automated workflows 0	Prediction		Ma	del equation: AC50 = 4.92 (±0	.176) + 0.0382 (±0.0422) * log K	ow, log(1/mol/L)				Gap filling approach
Standardized workflows 0	Adequacy		* <u>*</u>	•						Descriptors / data
	Cumulative frequency		•	•						Change descriptor units
		• •	•	•••						Edit descriptor options
	Residuals		• • • • •	•••						Select endpoint tree descriptor
	Statistics	• • •								Model/QSAR
									•	Calculation ontions
		So Bang and		B B B B B B	••••	• • •				Visual options
	-6 -4	-2 0	2	4 6	8 10 log Kow	12	14 16	18	20	Accept prediction
					-					V
1. Click "OK" on	the message ale	rtina v	ou foi	r data	incons	istency	/:			

The aim of this example is to see how the data correlates.

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with other AC50 data – step 4

QSAR TOOLBOX	Frofiling Data Category definition	01010 01 0 10100 on Data Gap Filing	► Report							Xoses	
Gap Filling Workflow Trend analysis Read across (QISAR Standardized Automated										The OECD QSAR 1 for Grouping Che into Categories Developed by LM	Foolbox micals IC, Bulgaria
 Documents 	Filter endpoint tree	1 [target]	3	40	41	86	87	93	94	112	
ent 1 nple_file_AW_SW_Ecotox bample_file_AW_SW_Ecotox ent 2 : 62533 CAS: 65733	Structure	~0	۰٬۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰۰	H5 5	H0 H0 OH		Ż	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	otito		
ent 3	AC50 (214/343		M: 0.156 mg/L	M: 7.53 mg/L	M: 2.77 mg/L	M: 5.55 mg/L	M: 6.86 mg/L	M: 1.11 mg/L	M: 2.24 mg/L	M: 0.000267 mg/L	
LastDB Enter GF(TA) with 357 chemicals, 486 data points		2	M: 10.6 mg/L	M: 1.97 mg/L							<u> </u>
Endpoint X-axis descriptor: AC50; Undefined Entrez gene	Tox21_AhR_viability (38/38	2			M: 0.859 mg/L			M: 10.7 mg/L			
	Tox21_AR_BLA_Agonist_ch1 (70/70) M: 0.367 mg/L	M: 7.66 mg/L	M: 8.57 mg/L					M: 0.00131 mg/L		
	Tox21_AR_BLA_Agonist_ch2 (34/34	2		M: 9.57 mg/l						M: 0.231 mg/L	M:
	Tox21_AR_BLA_Antagonist_ratio (33/33)			in sist ng/c						M: 0.0754 mg/L	
) Information			^ _						M:
	Tox21_AR_LUC_MDAKB2_Agonist (32/32)				M: 10.5 mg/L				M: 0.2 mg/L	
	Tox21_AR_LUC_MDAKB2_Antagonist_vial(35/35)	143 observed values	s for 142 chemicals were	excluded due to missing	X descriptor value(s)						M:
	Tox21_Aromatase_Inhibition (22/22)									
<	Tox21_Aromatase_Inhibition_viability (36/36	2			ОК	- 1					
 Data Gap Filling Settings 	Tox21_AutoFluor_HEPG2_Cell_blue (5/5					-					
✓ Only endpoint relevant	Tav21 AutaEluar BEDG2 Madia blua (2/2	1)			_	
✓ Only chemical relevant											
At this position:	Descriptors			Trend analysis predicti Predicted: 1.09 mg/L	on for AC50, based on 3	56 values				Select / filter data	^
Select a cell with a rigid (bold) path	Prediction		Moc	sel equation: AC50 = 4.92 (±0	176) + 0.0382 (±0.0422) * log I	(ow, log(1/mol/L)				Gap filling approach	
Standardized workflows 0				•						Descriptors / data	
	Adequacy		•								
	Cumulative frequency	•	•	•						Change descriptor uni	ts
	Pariduale E	•	•	•						Edit descriptor option	IS
			• • • • • •	** • • •						Select endpoint tree desc	riptor
	Statistics 56	• • •		@ 2° %	. 8 .					Model/OSAR	
		•							•	C L L L	
		Se and the				• •				Calculation options	
								+++++++++	┝┿┿╋┿╋	Visual options	

1. **Click** "OK" on the message informing you excluded number of chemicals due to missing X-descriptor data. They are analogues with no such type AC50 data. This will not affect the value of correlation coefficient;

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with other AC50 data – step 4

	► Profiling ► Data ► Category definition	01010 01 0 10100 n Data Gap Filling	► Report							X 8 5 4 8
Gap Filling Workflow										The OECD QSAR Toolbox for Grouping Chemicals into Categories
Decuments	Filter endpoint tree	1 [target]	3	40	41	86	87	93	94	Developed by LMC, Bulga
Int 1 Interpretation Interpretatio Interpretation Interpretation Interpretation Interpretation I	Structure		»،۹		HO CH3			~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	0-1-70°	->>° []
CAS: 62533 ent 3 CastDB Enter GF(TA) with 357 chemicals 486 data points	AC50 (214/343		M: 0.156 mg/L M: 10.6 mg/L	M: 7.53 mg/L	M: 2.77 mg/L	M: 5.55 mg/L	M: 6.86 mg/L	M: 1.11 mg/L		M: 0.000267 mg/L
Endpoint X-axis descriptor: AC50; Undefined Entrez gene	Image: Disk21_ARR (49/49 Image: Disk21_ARR_viability (38/38 Image: Disk21_ARR_BLA_Agonist_ch1 (70/70	M: 0.367 mg/L	M: 7.66 mg/L	M: 8.57 mg/L	M: 0.859 mg/L			M: 10.7 mg/L		
	← Tox21_AR_BLA_Agonist_ch2 (34/34 ← Tox21_AR_BLA_Agonist_ratio (25/25 ← Tox21_AR_BLA_Antagonist_ratio (33/33			M: 9.57 mg/L						M: 0.231 mg/L M: M: 0.0754 mg/L
	Tox21_AR_BLA_Antagonist_viability (34/34 Tox21_AR_LUC_MDAKB2_Agonist (32/32 Tox21_AR_LUC_MDAKB2_Antagonist (4/4	M: 4.8 mg/L	M: 8.74 mg/L	M: 0.12 mg/L		M: 10.5 mg/L				M: 0.2 mg/L
< >	Tox21_AR_LUC_MDAKB2_Antagonist_via(35/35 Tox21_Aromatase_Inhibition (22/22 Tox21_Aromatase_Inhibition_viability (36/36		M: 15 mg/L	M: 8.23 mg/L	M: 1.92 mg/L					M:
Data Gap Filling Settings	Tox21_AutoFluor_HEK293_Media_blue (5/5									
Only enapoint relevant Only chemical relevant						2			_	,
At this position:	Descriptors		M	Trend analys Predicted: N lodel equation: AC50 = 1.08 (±	is prediction for AC50, /A 0.344) + 0.799 (±0.0648) * AC	50, log(1 nol/L)				Select / filter data
Automated workflows 0 Standardized workflows 0	Prediction							9		Descriptors / data
	Cumulative frequency				• •					Change descriptor units
	Residuals		-		•			\leq	1	Edit descriptor options
	Statistics	•							Se	lect endpoint tree descriptor
										Calculation options
				••••	8 ••					Visual options
	4 4.5	5	5.5 6	AC50	7 7 5 [log(1/mol/L)]	8	8.5 9	9.5 10	0 10.5	Accept prediction

The graph obtained after replacing log Kow with Toxcast endpoint is visualized;
 The equation including endpoint data is rebuild;

Types endpoint correlations Continuous vs. continuous *Interpretation of correlation results*

- In this example, we have correlated two AC50 endpoints associated with different type assay
- As seen from the graph, a linear relationship between two endpoints has been observed
- In order to assess only the chemicals having positive estrogen activity we remove the "Non-binders" chemicals based on subcategorization by "Estrogen receptor binding by OASIS" profiler (illustrated on next slide)

Continuous vs. continuous

Subcategorization by Estrogen receptor binding profiler Sidebar of profiles relevancy

Once the endpoint is selected, the relevant profiles and metabolic transformations are highlighted.



- Suitable developed using data/knowledge for the target endpoint;
- **Plausible** structure-based; form broader group of analogues;
- Unclassified all profilers, which are not classified in any of the categories above.

Continuous vs. continuous

Subcategorization by Estrogen receptor binding profiler

QSF	IR TOOLBOX	-	# #		h	01010 01 0 10100								X 0 5 4 0	
🕘 Subca	tegorization	▶ Input	► Profiling	► Data	Category definition		Report							The OECD QSAR To for Grouping Cherr	oolbox nicals
Options 2	<u>ا</u>		Adjust opti	ons]									into Categories	
f Se	lect All Unselect All Invert	About								1993 -				Developed by LMC	, Bulgaria
Bio Bio Bio Bio Bio Bio Bio Bio Bio	egradation prohability (Blowin 1) egradation probability (Blowin 1) egradation probability (Blowin 5) egradation probability (Blowin 5) egradation probability (Blowin 7) egradation timate (Blowin 7)	2	Non binder, without OH o	or NH2 group				86 O		93 "~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	94 07-7-7-0		114 YO		0
DN	A binding by OASIS	٦ -			(214/343)	M: 7.53 mg/L	M: 2.77 mg/L	M: 5.55 mg/L	M: 6.86 mg/L	M: 1.11 mg/L	M: 2.24 mg/L	M: 0.000267 mg/L	M: 0.929 mg/L	M: 8.85 mg/L	
Est	rogen Receptor Binding				ERa Antagoni(69/81)	M: 1.97 mg/L							M: 0.00222 mg/L	M: 0.000236 mg/L	<u> </u>
Hy	drolysis half-life (Ka, pH 8)(Hydrowin)				(38/38)		M: 0.859 mg/L			M: 10.7 mg/L			M: 0.000198 mg/L	M: 0.000236 mg/L	
Hyd	drolysis half-life (Kb, pH 7)(Hydrowin) drolysis half-life (Kb, pH 8)(Hydrowin)				(70/70)	M: 8.57 mg/L				-	M: 0.00131 mg/L		M: 2.18 mg/L		
Hy	drolysis half-life (pH 6.5-7.4)				(34/34)							M: 0.231 mg/L	M: 1.02 mg/L		
Ion	nization at pH = 1 nization at pH = 4		Differ from target by		atic	: 9.57 mg/L						M: 0.179 mg/L M: 0.0754 mg/L	M: 1.44 mg/L		
<	· ·· · · · ·	>	 At least one category All categories 	[STOP]	3	: 0.12 mg/L							M: 0.407 mg/L		
Options ,	4				h	J		M: 10.5 mg/L				M: 0.2 mg/L			
f Se	lect All Unselect All Invert				tagonist (4/4)							M: 2.49 mg/L	Mt 0.000102 mg/l		_
Do not	t account metabolism nented	^	(7) Moderate binder, OH 9) Non binder, impaired	grooup	(22/22)								M: 0.000 198 Mg/L		
Ob	served Mammalian metabolism		9) Non binder, MW>500		iability (36/36)	M: 8.23 mg/L	M: 1.92 mg/L							M: 0.00118 mg/L	
Ob	served Microbial metabolism served Rat In vivo metabolism		(34) Non binder, non cycl	ic structure	edia_blue (5/5)										
Ob	served rat liver metabolism with quar	ntitative dai	(90) Non binder, without	OH or NH2 group	L_blue (5/5)										~
⊿ Simula	ated		(26) Strong binder, NH2 gr	oup											>
Au Au Die	toxidation simulator toxidation simulator (alkaline medium) sociation simulator)	(22) Very strong binder, OH (3) Weak binder, NH2 grou	Very strong binder, OH group				Trend analysis Predicted: N/A	prediction for AC50,					Select / filter data	
Hy	drolysis simulator (acidic)		(13) Weak binder, OH gro	up C			Mod	el equation: AC50 = 1.08 (±0.3	44) + 0.799 (±0.0648) * AC50, k	og(1/mol/L)				Subcategorize	
Hyo Hyo in y	drolysis simulator (basic) drolysis simulator (neutral) vivo Rat metabolism simulator		<		4						••••8*		M	ark chemicals by WS	1
Mic Rat	robial metabolism simulator t liver S9 metabolism simulator		Selected 142 (72/214) Select differ	ent				•	•	•			Mark ch	emicals by descriptor	
<	n matahalian sinulatar	>	Remove sele	cted				•	8					Mark outliers	
				[log	•	•	• • ••	•••		•			Filter	points by test condition	15
			Statistics	900	• •				••••		•		M	ark focused chemical	
_							•							lark focused points	
	1. Open "S	elect	/filter dat	a" me	enu iter	n, then	click "	Subcate	gorize";	2	. Se	elect	"Estro	jen data	~
	recentor bir	ndina	" nrofiler		3 50	ect on	ly Non I	hinder c	ategorie	s hy la	ft mou	se click	and h	old rediction	n
		uniy	promer	1	J. 3ei				acegone		it mou				
	"Ctrl" butto	n;	4. Click	("Rer	nove" b	utton;									

Continuous vs. continuous

Correlation of active Estrogen receptor categories vs.AC50 endpoint

QSAR TOOLBOX	↓ Input	r 1 L J ► Profiling	► Data	Category definition	01010 01 0 10100 > Data Gap Filling	► Report							X 0 5 4 0
Q Subcategorization			- 🗆 ×										The OECD QSAR Toolbox for Grouping Chemicals
Options 🖌			Adjust options										into Categories
f Select All Unselect All Invert	About	Options	ridjust options										Developed by LMC Bulgaria
Biodegradation probability (Biowin 1)		^	New Mindee without ON		131	152	173	191	257	262	266	292	295
Biodegradation probability (Biowin 2) Biodegradation probability (Biowin 5)			Non binder, without OH										
Biodegradation probability (Biowin)					0	6		0.1		.0		
Biodegradation probability (Biowin)	1				~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	~ <u>_</u>		Q	<u> </u>	Q	1 X	00	
DNA binding by OASIS						~~		6		NH2	×		····
DNA binding by OECD				(72//22)	NL 0.001 //	M 0.000175 //	M-224	M 2 27 //	ht 0.000224 //	ALC 52	NL 0.666 0	M-206	M.5.50
Estrogen Receptor Binding Hydrowsis balf-life (Ka. pH 7)(Hydrowin)				(72/123) EPo Antogoni(25/21)	M: 0.961 mg/L M: 8.51 mg/l	M: 0.000175 mg/L	WI: 2.24 mg/L	W: 5.27 mg/L	M: 0.000224 mg/L	M: 5.52 mg/L	M: 0.000 mg/L	WI: 2.90 mg/L	W: 5.56 mg/L W:
Hydrolysis half-life (Ka, pH 8)(Hydrowin)				(24/24)	Wi: 6.51 mg/c		M: 0.0681 mg/l			M: 1.21 mo/l	M: 2.7 mg/l	M: 2.12 mg/l	M: 0.177 mg/l M:
Hydrolysis half-life (Kb, pH 7)(Hydrowin)				(12/12)			in oroson ng/c	M: 0.0879 mg/L		in the trigite	in ching c	in the right	in on ring, c
Hydrolysis half-life (KD, pH 8)(Hydrowin) Hydrolysis half-life (pH 6 5-7 4)		_		(26/26)	M: 0.596 mg/L	M: 0.0162 mg/L	M: 0.000268 mg/L			M: 0.214 mg/L	M: 0.0102 mg/L	M: 0.000144 mg/L	
Ionization at pH = 1	()	< >	(8/8)		M: 0.326 mg/L							
Ionization at pH = 4	1		Differ from tar	(4/4)			<u> </u>						
Ionization at pH = 9			At least on recom	atio (19/19)		M: 0.118							M:
Protein hinding by OASIS		Ľ	O All categor	iability (11/11)			3					M: 8.23 mg/L	
Options 🖌		$\overline{}$		jonist (17/17)							M: 5.75 mg/L		M:
f Select All Unselect All Invert				tagonist (3/3)			M 14.2 mg/l						
Do not account metabolism		^	(7) Moderate binder, OH (1) Stores binder, NH2	(10/10)		M: 10.9 mg/l	Wi: 14.2 Mg/L						
Observed Mammalian metabolism			(1) Strong binder, NH2 g (26) Strong binder, OH g	(10/10)		With Toto Highe	M: 5.3 mg/L	M: 2.67 mg/L					
Observed Microbial metabolism			(20) Strong binder, Orig	edia blue (2/2)									
Observed Rat In vivo metabolism	ativo data		(3) Weak binder, NH2 or	II blue (2/2)						M: 4.45 mg/L			
Observed Rat Liver S9 metabolism	acive data		(13) Weak binder, OH gr	dia blue (1/1)									
▲ Simulated				-									,
Autoxidation simulator Autoxidation simulator (alkaline medium)							Trend analysis	prediction for AC50.					Select / filter data
Dissociation simulator							Predicted: N//	A	0 1==(1/===1/1)				Sciect / inter data
Hydrolysis simulator (acidic)						MC	del equation: ACS0 = 0.409 (±0	(285) + 0.955 (±0.0497) - AC5	0, log(1/movc)				Subcategorize
Hydrolysis simulator (basic) Hydrolysis simulator (neutral)			< >			<u> </u>							Mark chemicals by WS
in vivo Rat metabolism simulator			Selected 7 (65/72)		_						68 -		wark chemicals by wo
Microbial metabolism simulator			Select different	1						•		Mark o	hemicals by descriptor value
Skin metabolism simulator			Descere underend										Mark outliers
Tautomorism			Remove selected	J				• • • •					
			<u> </u>			•	•••					Filte	r points by test conditions
		Statistics	- ACSI	•			F	•			"Mode	erate b	inders"
						•							
			4	• 35 9							VS.	<u>AC50</u>	aata
4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10 AC50 [log(1/mol/L)]										Accept prediction			

- 1. Click again on Estrogen receptor binding profiler
- 2. Select "Moderate binder" categories
- 3. The chemicals corresponding to the selected categories are highlighted in green; 4 and in light blue on the graph

Continuous vs. continuous

Correlation of active Estrogen receptor categories vs.AC50 endpoint

QSAR TOOLEOX	► Input	► Profiling	► Data	Category definitio	01010 01 0 10100 n	► Report							Xesel)
Subcategorization			- 🗆 ×										The OECD QSAR for Grouping Ch	Toolbox emicals
Options	Abrut	Ontinus	Adjust options										into Categories	
Biodegradation probability (Biowin 1)	About	Options ^	Non binder without OH		191	257	262	266	292	295	323	330	Developed by LN 355	IC, Bulgaria
Biodegradation probability (Biowin 2) Biodegradation probability (Biowin 5)			Non binder, without off			2.10								
Biodegradation probability (Biowin 6) Biodegradation probability (Biowin 7)					Q		O	~ ²	00		Hig O	<u>O</u>	NG O	
DNA binding by OASIS					6	<u></u>	192-	Y.	No.		MC Vanj	H ¹ Ca. MH2	Hyce Aris	
Estrogen Receptor Binding				(72/123)	M: 3.27 mg/L	M: 0.000224 mg/L	M: 5.52 mg/L	M: 0.666 mg/L	M: 2.96 mg/L	M: 5.58 mg/L	M: 1.27 mg/L	M: 0.131 mg/L	M: 6.82 mg/L	M:
Hydrolysis half-life (Ka, pH 7)(Hydrowin) Hydrolysis half-life (Ka, pH 8)(Hydrowin)				ERa Antagon(25/31) (24/24)			M: 1.21 mg/L	M: 2.7 mg/L	M: 2.12 mg/L	M: 0.177 mg/L	M: 11 mg/L	M: 7.12 mg/L M: 0.227 mg/L		M:
Hydrolysis half-life (Kb, pH 7)(Hydrowin) Hydrolysis half-life (Kb, pH 8)(Hydrowin)				(12/12)	M: 0.0879 mg/L		Mr 0 214 mg/l	M: 0.0102 mg/l	Mt 0.000144 mg/l				Mr 0.000230 mg/l	
Hydrolysis half-life (pH 6.5-7.4) Ionization at pH = 1			< >	(20/20) (8/8)			Wi. 0.214 Hig/L	Wi. 0.0102 mg/c	1W. 0.000 144 Hig/L				M. 0.000325 Hig/E	M:
Ionization at pH = 4 Ionization at pH = 7.4			Differ from targ	(4/4) (19/19)							M: 6.57 mg/l			
Ionization at pH = 9 Protein hinding by OASIS		v	At least on [STOP] All categor	iability (11/11)					M: 8.23 mg/L					
Options		^		ponist (17/17) ntagonist (3/3)				M: 5.75 mg/L			M: 8.49 mg/L	M: 2.58 mg/L		
Do not account metabolism			(7) Moderate binder, OH	ntagonist_vial(13/13)									M: 8.95 mg/L	
 Documented Observed Mammalian metabolism 			 Strong binder, NH2 g Strong binder, OH g 	viability (13/13)	M: 2.67 mg/L								M: 5.93 mg/L	
Observed Microbial metabolism Observed Rat In vivo metabolism			(22) Very strong binder,	edia_blue (2/2)			M: 4.45 mg/l							
Observed rat liver metabolism with quan Observed Rat Liver S9 metabolism	titative data		(13) Weak binder, NH2 gr	dia bluo (1/1)			in the highe							> v
 A Simulated Autoxidation simulator 					//		~							
Autoxidation simulator (alkaline medium) Dissociation simulator	1	1			/	Mode	Frend analysis Predicted: N/A equation: AC50 = 0.409 (±0	prediction for AC50, A (285) + 0.953 (±0.0497) * AC50	0. log(1/mol/L)				Select / filter data	
Hydrolysis simulator (acidic) Hydrolysis simulator (basic)													Subcategorize	
hydrolysis simulator (neutral) in vivo Rat metabolism simulator			Selected 16 (56/72)	<u> </u>							• 8 •		Mark chemicals by WS	
Microbial metabolism simulator Rat liver S9 metabolism simulator			Select different	2				•				Mark	chemicals by descriptor	value
Skin metabolism simulator Tautomorism		~	Remove selected	_ _				•					Mark outliers	
		Statistics	20100				••••				_	Filt	er points by test conditi	ons
	Statistics Version of the second seco									Mark focused chemical				
			• • • • • • • • • • • •		6 • •			Δ.	C50 dat				Mark focused points	
			4			55 6		7 75		.a		+-+-+	Remove marked data	~
				- 4.J	,	5.5 0	0.5 AC50	[log(1/mol/L)]	v	0.J 9	2.3		Accept predicti	on

Select "Weak binder" categories (left mouse click and hold "Ctrl" button);
 The chemicals corresponding to the selected categories are highlighted in green;

Continuous vs. continuous

Correlation of active Estrogen receptor categories vs.AC50 endpoint

			A		01010								🗙 🖨 🛧 🖉 🗊
QSAR TOOLBOX					10100								
	Input	Profiling	► Data	Category definition		Report							
Subcategorization			- 🗆 ×										The OECD QSAR Toolbox for Grouping Chemicals
Options 🖌			Adjust options										into Categories
f Select All Unselect All Invert	About	Options											Developed by LMC, Bulgaria
Biodegradation probability (Biowin 1) Biodegradation probability (Biowin 2)		^	Non binder, without OH		3	86	93	112	118	131	152	173	191
Biodegradation probability (Biowin 5)						<u>^</u>						ĩ	
Biodegradation probability (Biowin 7)					······································		······	~00	0.0	+~~~~Q ⁶⁴	~0	0	
Biodegradation ultimate (Biowin 3)								-0-		6	- Com	0	
DNA binding by OECD				(72//22)	M-0155	M. E.E.E. mar //	M. 1.11	M. 0.000267 //	A4-0.05	M-0.001 (l	M. 0.000175	M. 2.24 //	1 M 2 27 m = //
Estrogen Receptor Binding Hydrolysis half-life (Ka, pH 7)(Hydrowin)				(72/123) FRa Antagoni(25/31)	M: 0.156 mg/L M: 10.6 mg/L	M: 5.55 mg/L	M: I. I I mg/L	W: 0.000207 mg/L	M: 0.000236 mg/L	M: 0.981 mg/L M: 8.51 mg/L	M: 0.000175 mg/L	M: 2.24 mg/L	M: 5.27 mg/E M:
Hydrolysis half-life (Ka, pH 8)(Hydrowin)				(24/24)								M: 0.0681 mg/L	
Hydrolysis naif-life (Kb, pH 7)(Hydrowin) Hydrolysis half-life (Kb, pH 8)(Hydrowin)				(12/12)			M: 10.7 mg/L		M: 0.000236 mg/L				M: 0.0879 mg/L
Hydrolysis half-life (pH 6.5-7.4)				(26/26)	M: 7.66 mg/L			Mi 0 221 mg/l		M: 0.596 mg/L	M: 0.0162 mg/L	M: 0.000268 mg/L	
Ionization at pH = 1 Ionization at pH = 4				(8/8)				M: 0.251 mg/L M: 0.179 mg/L			W: 0.520 mg/L		
Ionization at pH = 7.4 Ionization at pH = 9			Differ from targ	atio (19/19)				M: 0.0754 mg/L			M: 0.118 mg/L		2
Protein hinding.by.QASIS		~	O All categor	iability (11/11)	M: 8.74 mg/L							<	
Options 🖌				ponist (17/17)		M: 10.5 mg/L		M: 0.2 mg/L			M: 1.15 mg/L		
f Select All Unselect All Invert			(7) Moderate binder, OH	ntagonist vial(13/13)	M: 15 mg/L			Wi. 2.45 Hig/C				M: 14.2 mg/L	
 Documented 			(1) Strong binder, NH2 g	(10/10)							M: 10.9 mg/L		
Observed Mammalian metabolism Observed Microbial metabolism			(26) Strong binder, OH g	viability (13/13)					M: 0.00118 mg/L			M: 5.3 mg/L	M: 2.67 mg/L
Observed Rat In vivo metabolism			(22) Very strong binder, ((2) Week binder, NH2 es	edia_blue (2/2)									
Observed rat liver metabolism with quan Observed Rat Liver S9 metabolism	titative	1	(3) Weak binder, NH2 gr (13) Weak binder, OH gr	dia bluo (1/1)									v
✓ Simulated													>
Autoxidation simulator (alkaline medium)							Trend analysis	prediction for AC50,					Select / filter data
Dissociation simulator Hydrolysis simulator (acidic)						Mod	del equation: AC50 = 0.409 (±0	A 0.285) + 0.953 (±0.0497) * AC50	0, log(1/mol/L)				Subatagoria
Hydrolysis simulator (basic)				\square									Subcategonze
in vivo Rat metabolism simulator			Selected 49 (23/72)		2					P			lark chemicals by WS
Microbial metabolism simulator Rat liver S9 metabolism simulator			Select different					•		•		Mark ch	nemicals by descriptor value
Skin metabolism simulator		~	Remove selected										Mark outliers
		nesiadais	jõo -			•		•				Filter	points by test conditions
		Statistics			- 0		<u> </u>						lark focused chemical
			◄	• •	008			"Stron	a and v	erv stro	na		lark locased enemical
					3000			لمنامط	o"				Mark focused points
			4				┉┉┈	Dinder	<u>s vs./</u>	HCOU da		╪╪╤╤╝╽	Remove marked data 🗸 🗸
			· · · ·	4 4.5	5	5.5 6	6.5 AC50	7 7.5 [log(1/mol/L)]	8	8.5 9	9.5	10	Accept prediction
1 Select "Strop	na ang	d ver	v strong	hinder	" cated	iories (I	eft mou	ise click	and ho	old "Ctrl	" huttor	1)	

Select "Strong and very strong binder" categories (left mouse click and hold "Ctrl" button
 The chemicals corresponding to the selected categories are highlighted in green;

Types endpoint correlations Continuous vs. continuous *Correlation results*

- The two AC50 endpoints associated with different type assay have been correlated each other
- Non binders according to Estrogen receptor binding profiler have been eliminated from the correlation
- User can analyse the distribution of remaining ER binders (Very strong, Strong, Moderate and Weak) across selected AC50 endpoint