The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD (Q)SAR Toolbox v.4.4.1

Example illustrating endpoint vs. endpoint correlation using ToxCast data

Outlook

• Background

- Objectives
- The exercise
- Workflow

Background

This presentation is designed to introduce the user to:

- 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 the two types of data:
 - > AC50 vs. AC50 endpoints associated with different test type
 - > AC50 vs. Estrogen receptor binding data

Outlook

- Background
- Objectives
- The exercise
- Workflow

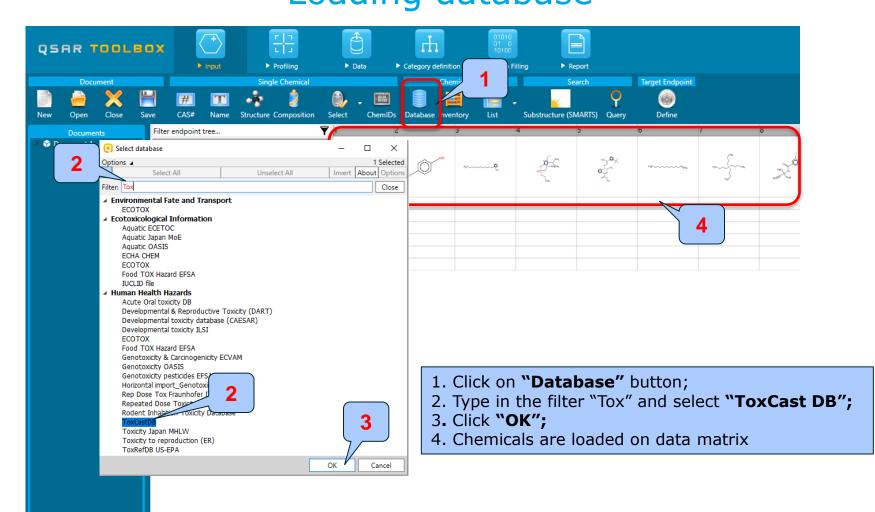
Workflow

- 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

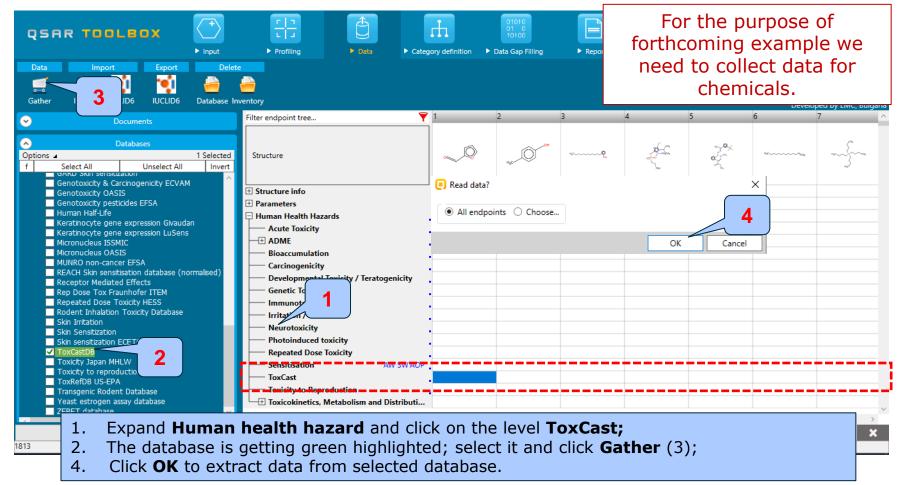
- Background
- Objectives
- The exercise
- Workflow
 - Load ToxCast database

ToxCast database Loading database



ToxCast database Sidebar of database relevancy

Once the endpoint is selected, the relevant databases become highlighted in green.



ToxCast database Data gathering

ilter endpoint tree	Ţ	1	2	3	4	5	6	7	8	9	10	11
Structure		, Q	нус	* <u>P</u>		syo ov €	460000000000000000000000000000000000000		ž	¢.0	H3CC43	Hendrydag
Parameters												
Physical Chemical Properti	es											
Environmental Fate and Tra	ansport											
Ecotoxicological Information												
Human Health Hazards												
— Acute Toxicity												
+ ADME												
— Carcinogenicity												
 Developmental Toxicity 	/ Teratog											
— Genetic Toxicity												
- Immunotoxicity												
- Irritation / Corrosion												
- Photoinduced toxicity					1							
— Repeated Dose Toxicity	, [
Sensitisation	AW SW AOP				\sim							
-🖯 ToxCast												
- + ACEA	600/660			M: 0.0601 mg/L		M: 7.06 mg/L	_	M: 6.8 mg/L	M: 2.84 mg/L			M: 0.0219 mg/
+ Apredica	425/2653					M: 1.69 mg/L				M: 32.1 mg/L		
- 🕀 Attagene	1374/11710		M: 0.88 mg/L	M: 0.113 mg/L	M: 0.627 mg/L	M: 4.82 mg/L	M: 16.2 mg/L	M: 0.033 mg/L	M: 3.79 mg/L	M: 3.4 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	M: 0.464 mg/L	M: 0.187 mg/L	
- I NCGC	1475/6890	M: 0.367 mg/L		M: 0.156 mg/L	M: 1.61 mg/L	M: 0.357 mg/L		M: 1.86 mg/L		M: 0.000358 mg/L	M: 0.0144 mg/L	M: 6.23 mg/L
+ Novascreen	975/8054		M: 2.43 mg/L		M: 0.0957 mg/L	M: 0.209 mg/L	M: 0.0122 mg/L	M: 8.61 mg/L	M: 0.0597 mg/L			
Odyssey Thera	969/2794		M: 6.89 mg/L	M: 0.121 mg/L	M: 9.54 mg/L	M: 0.592 mg/L		M: 17.1 mg/L	M: 14.1 mg/L		M: 6.03 mg/L	
⊡ Undefined Assay pro	vider 2/2											
- Texisity to Reproductio												
- Toxicokinetics, Metabo	lism and D											
									1			

1. The data appears in the datamatrix under 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 introduces the user to the opportunity to analyze correlations between selected gap filling endpoints (endpoints 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 a target endpoint (this is the initial endpoint selected by the user) displayed on ordinate axis (Y-axis) and other endpoint data displayed on the 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 TB 4.4.1 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 correlate with 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 correlate with 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 the next few slides. Summary of the workflow steps are provided below:
 - Gather experimental data (step 1)
 - Selection of target endpoint (step 2)
 - Enter Gap filling (step 3)
 - Change default X-descriptor (logKow) with AC50 data (step 4)

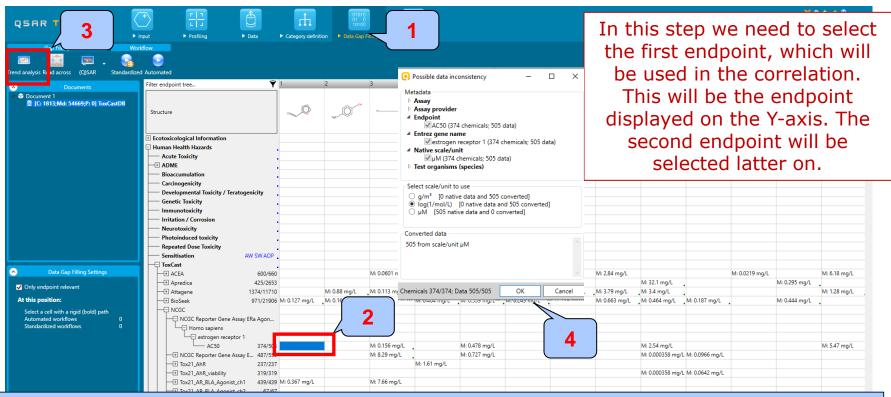
Continuous vs. continuous

Gather experimental data – step 1

	Filter endpoint tree	Y 1	2	3	4	5	6	7	8	9	10	11	Deve 12	loped by LN 13
nent 1 : 1813;Md: 54669;P: 0] ToxCastDB	Structure			٠٠٠٠٠ <u>۴</u>	e de la compañía de	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	14		zà		ryc Hoars	Handrandones	-500	na*
	± Structure info													
	Parameters													
	Physical Chemical Properties													
	Environmental Fate and Transport													
	Ecotoxicological Information													
	- Human Health Hazards													
	Acute Toxicity	•												
	ADME Bioaccumulation													
	Carcinogenicity													
	Developmental Toxicity / Terat	•												
	Genetic Toxicity													
	Immunotoxicity													
	Irritation / Corrosion													
Databases	Neurotoxicity													
1 Selecte														
Unselect All Inver gene expression LuSens	Repeated Dose Toxicity													
SSMIC	Sensitisation	AW SW AOP												
DASIS	ToxCast													
cer EFSA itisation database (no	- E ACEA	600/660		M: 0.0601 mg/L		M: 7.06 mg/L		M: 6.8 mg/L	M: 2.84 mg/L			M: 0.0219 mg/L		M: 6.18 r
d Effects		425/2653				M: 1.69 mg/L	•			M: 32.1 mg/L	•		M: 0.295 mg/L	
nhofer ITEM	- + Attagene	1374/11710	M: 0.88 mg/L	M: 0.113 mg/L	M: 0.627 mg/L	M: 4.82 mg/L	M: 16.2 mg/L	M: 0.033 mg/L	M: 3.79 mg/L	M: 3.4 mg/L	•			M: 1.28 r
	BioSeek	971/21906 M: 0.127 mg			M: 0.464 mg/L	M: 0.539 mg/L	M: 0.243 mg/L	•	M: 0.663 mg/L	M: 0.464 mg/L		•	M: 0.444 mg/L	
	E NCGC	1475/6890 M: 0.367 mg, 975/8054	•	M: 0.156 mg/L		M: 0.357 mg/L	•	M: 1.86 mg/L		M: 0.000358 mg/l	M: 0.0144 mg/L	. M: 6.23 mg/L	M: 0.33 mg/L	M: 5.47 r
			M: 2.43 mg/L		M: 0.0957 mg/L M: 9.54 mg/L	M: 0.209 mg/L M: 0.592 mg/L	M: 0.0122 mg/L	M: 8.61 mg/L M: 17.1 mg/L	M: 0.0597 mg/L M: 14.1 mg/L	•	M: 6.03 mg/L			M: 1.64 r M: 10.7 r
Toxicity Database	Novascreen		M 6.99 m = //					with the mg/L	wi: 14.1 mg/L		m. 0.05 mg/c			WI: 10.7 P
cicity Database	- + Odyssey Thera	969/2794	M: 6.89 mg/L	M: 0.121 mg/L										
cicity Database	⊕ Odyssey Thera ⊕ Undefined Assay provider		M: 6.89 mg/L	M: 0.121 mg/L										
oxicity HESS Toxicity Database	- + Odyssey Thera	969/2794 2/2	M: 6.89 mg/L	M: 0.121 mg/L										

Types endpoint correlations

Continuous vs. continuous Selection of target endpoint – step 2



1. Go to *Data Gap Filling* module;

2. Highlight the empty cell next to the AC50 endpoint associated with assay: "NCGC Reporter Gene Assay ERa Agonist"

- 3. Click "Trend analysis";
- 4. A window alerting you for data inconsistencies appears. Keep it as it is. Click "OK".

Continuous vs. continuous Selection of target endpoint – step 2

	d Automated Filter endpoint tree	💙 1 [target]	2	5	0	21	29	41	42	59	65	69	Deve 87	loped by LMC,
Documents 1 Document 1 ■ [C: 1813/Md: 54669;P: 0] ToxCastDB ⊞ [C: 352;Md: 16101;P: 0] Enter GF(TA)	Structure		٠ <u>۴</u>	°. °	x ⁰ ?								O	т. Т.
	Ecotoxicological Information Human Health Hazards Acute Toxicity													
	ADME Bioaccumulation Carcinogenicity		formation					×						
	Developmental Toxicity / Teratogenicity Genetic Toxicity Immunotoxicity Initration / Corrosion Neurotoxicity Photoinduced toxicity				chemicals were escriptor value		lue to missin	ıg X						
	Repeated Dose Toxicity Sensitisation AW SW A ToxCast													
Data Gap Filling Settings	ACEA 186/ Apredica 108/	/226						ок	M: 4.13 mg/L			M: 3.29 mg/L	M: 1.98 mg/L	•
Only endpoint relevant	+ Attagene 286/2		M: 0.113 mg/L	M: 4.82 mg/L		M: 0.0102 mg/L		M: \$83 mg/L			M: 3.33 mg/L	M: 3.87 mg/L	M: 0.389 mg/L	
this position: OSARs 0	BioSeek 208/6	5468 M: 0.127 mg/L	•	M: 0.539 mg/L	M: 0.464 mg/L	M: 0.307 mg/L	• (M: 2.56 mg
Automated workflows 0	NCGC Reporter Gene Assay ERa Ago	n						1						
Standardized workflows 0	Homo sapiens							•)						
nodes below:	estrogen receptor 1 AC50 351/	/479	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	M: 4.33 mg/L	M: 0.899 mg/L	M: 5.55 mg/L	M: 6.86 mg
QSARs 0 Automated workflows 0	NCGC Reporter Gene Assay E 120/		M: 8.29 mg/L	M: 0.727 mg/L	M: 0.000358 mg/l			M: 1.97 mg/L				M: 4.19 mg/L		
Standardized workflows 0		3/83			M: 0.000358 mg/l				M: 0.859 ma/L		M: 3.27 ma/L	M: 1.24 mg/L		
		/114 M: 0.367 mg/L	M: 7.66 mg/L		ini. e.eeesse mg/t			M: 8.57 mg/L	mi ologije ingre		With Starr Hight	M: 14.8 mg/L		
		6/36												
		4/44						M: 9.57 mg/L						
	Tox21_AR_BLA_Antagonist_ratio 44	6/46	M: 8.74 mg/L	M: 1.53 mg/L				M: 0.12 mg/L				M: 7.6 mg/L		
	Tox21_AR_LUC_MDAKB2_Agonist 39		, in on Fig. 2	in the trigit				in one ngre					M: 10.5 mg/L	

The

Types endpoint correlations

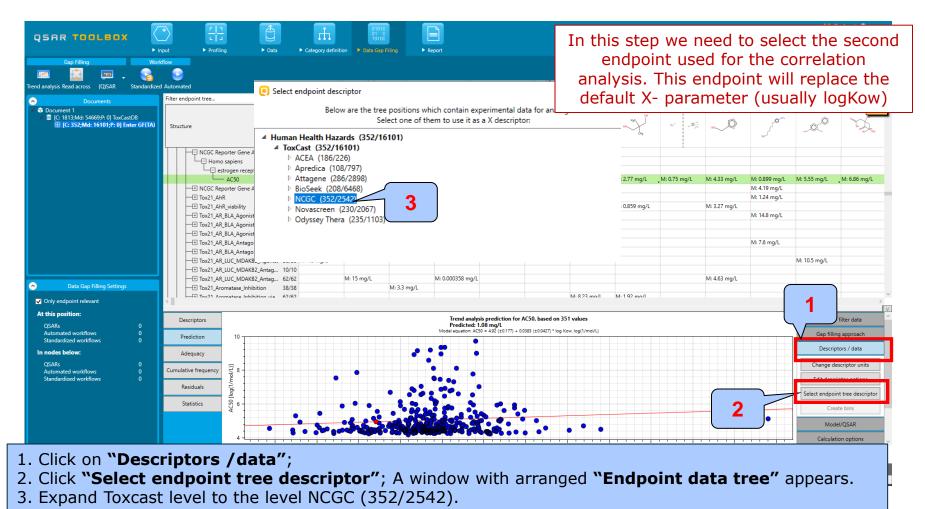
Continuous vs. continuous

Enter Gap filling – step 3

[C: 1813;Md: 54669;P: 0] ToxCastDB [C: 352;Md: 16101;P: 0] Enter GF(TA)								21	-			pic	lieu	on Y	-dxis	•
	Struc	ture	۰	, Ø	٠ę	×°×	0,00	"/~~~~® ^{rai}		**~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	H¢ Dhj	ni _0Ç				Υ,
	$\left \right $	NCGC Reporter Gene Assay ERa	Agon													
		AC50	351/479		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	M: 4.33 mg/L	M: 0.899 mg/L	M: 5.55 mg/L	M: 6.86 mg/L
		NCGC Reporter Gene Assay E			M: 8.29 mg/L	M: 0.727 mg/L	M: 0.000358 mg/L		Live ing/t	M: 1.97 mg/L	Line ing/c		in the ring t	M: 4.19 mg/L	in one of the	
		Tox21_AhR	83/83								N 0.050			M: 1.24 mg/L		
		Tox21_AhR_viability Tox21_AR_BLA_Agonist_ch1	62/62 114/114 M: 0	0.367 ma/L	M: 7.66 mg/L		M: 0.000358 mg/L			M: 8.57 mg/L	M: 0.859 mg/L		M: 3.27 mg/L	M: 14.8 mg/L		
		Tox21_AR_BLA_Agonist_ch2	36/36													
		Tox21_AR_BLA_Agonist_ratio	44/44							M: 9.57 mg/L						
		Tox21_AR_BLA_Antagonist_ratio Tox21_AR_BLA_Antagonist_viabi. Tox21_AR_BLA_Antagonist_viabi.			M: 8.74 mg/L	M: 1.53 mg/L				M: 0.12 mg/L				M: 7.6 mg/L		
		Tox21_AR_BLA_Antagonist_viabi.		4.8 mg/L	m. o.r. mg/L	init hoo hig/t				m. v.rz mg/t					M: 10.5 mg/L	
		Tox21_AR_LUC_MDAKB2_Antag	. 10/10	-												
Data Gap Filling Settings		To 21 AP LUC MDAVB2_Antag	. 62/62 38/38		M: 15 mg/L	M: 3.3 mg/L	M: 0.000358 mg/L						M: 4.63 mg/L			
			38/38			willow mg/L				M- 8 23 mg/l	M-192 mo/l					~
nly endpoint relevant	<	H 3 ition via) V
his position:	C	Descriptors							n for AC50, based	on 351 values					Se	lect / filter data
SARs 0 utomated workflows 0		Prediction					Model equatio	ed: 1.08 mg/L n: AC50 = 4.92 (±0.17	7) + 0.0383 (±0.0427)	* log Kow, log(1/mol/L	.)				Gap	filling approach
tandardized workflows 0								•								escriptors / data
odes below:		Adequacy				•		•								
SARs 0 utomated workflows 0	Cumu	ative frequence					• •									Model/QSAR
tandardized workflows 0	_	<u>ē</u>		•	•		00 ee	•							Cal	culation options
		Residuals				• •		•							·····	√isual options
		Statistics				8.00	9	b o	•							Information
		¥ .	•	•				89								
				S-0		CONTRACTOR OF										Miscellaneous
		4	•							_						
			4	-2	Ó	2	4	6	8	74		14	16	18	20	A
									log Kow						V	Accept prediction
		plying trend an														×

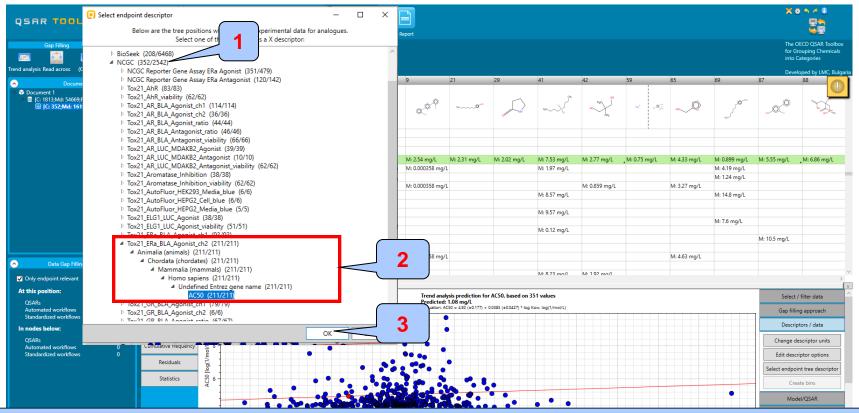
Continuous vs. continuous

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



Continuous vs. continuous

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



- 1. Click on "NCGC" node to open the sub-nodes;
- Select endpoint, which will be placed on X-axis circled in red box; point the mouse on the level of AC50 (211/211);
- 3. Click "OK" button.

Continuous vs. continuous

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

z 🔝 📼 . 😪	stow > Assay > > Assay provider		The OECD QSAR Toolb for Grouping Chemical into Categories
analysis Read across (Q)SAR Standardize Documents Document 1 G(: 1813;Md: 54669;P: 0) ToxCastD8 (C: 352;Md: 16101;P: 0) Enter GF(TA)	3 Automated Filter endpoint tree ▼ 1 [target] 3 5 9 2 Structure Image: Constraint of tree Structure Image: Constraint of tree Structure Image: Constraint of tree Image: Constraint of tree Image: Constraint of tree Image: Constraint of tree	69 	Developed by LMC, Bu
Data Gap Filing Settings Only endpoint relevant	Image: Select scale/unit to use Image: Select scale/unit to use <td>M: 0.899 mg/L M: 4.19 mg/L M: 1.24 mg/L M: 1.4.8 mg/L M: 7.6 mg/L</td> <td>M: 555 mg/L M: 686 mg/L M: 105 mg/L</td>	M: 0.899 mg/L M: 4.19 mg/L M: 1.24 mg/L M: 1.4.8 mg/L M: 7.6 mg/L	M: 555 mg/L M: 686 mg/L M: 105 mg/L
t this position: QSARs 0 Automated workflows 0	Descriptors Trend analysis prediction for AC50, based on 351 values Predicted: 1.08 mg/r Model equation: AC50 = 4.82 (±0177) + 0.0583 (±0.0427) * log Kow, log(1/mo/l/)		Select / filter data Gap filling approach
Standardized workflows 0 nodes below: QSARs 0 Automated workflows 0	Prediction Adequecy Cumulative frequency Image: Space of the state of the		Descriptors / data Change descriptor units Edit descriptor options
	Residuals Statistics 000 6 -		Edit descriptor options Select endpoint tree descripto Create bins Model/OSAR

1. Click **"OK"** on the message alerting you for data inconsistency; 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

Gap Filling World	2				o Filling 🛛 🕨 i	Report								for Gr	ECD QSAR Tooll ouping Chemica ategories
ed analysis Read across (Q)SAR Standardized Documents Document 1	Automated Filter endpoint tree		💙 1 [target]	3	41	42	87	88	94	95	113	115	121	Devel 123	oped by LMC, Br 134
 ☑ C: 1813;Md: 54669;P: 0] ToxCastD8 ☑ [C: 352;Md: 16101;P: 0] Enter GF(TA) ☑ [C: 212;Md: 9962;P: 0] Endpoint X 	Structure		~	٠ę	**~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	H5C CH DH3		×X.	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	07.7C		× .0	Q~Q	<u>_</u> ^	·~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
	Homo sap	er Gene Assay ERa Agon iens en receptor 1										1			
	AC	50 211/3		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	M: 0.929 mg/L		M: 5.55 mg/L	M: 0.981 mg/
	+ NCGC Report	er Gene Assay ERa 69/4 49/4		M-8 20 mo/l	M-107 mo/l		_		×			M: 0.00222 mg/L	M: 0.000236 mg/L		M: 8.51 mg/l
	Tox21_AhR_vi			mation					mg/L			M: 0.000198 mg/	L M: 0.000236 mg/L		
	Tox21_AR_BLA		70 M:							M: 0.00131 mg/l		M: 2.18 mg/L			M: 0.596 mg
	Tox21_AR_BLA										M: 0.231 mg/L M: 0.179 mg/L	M: 1.02 mg/L M: 1.44 mg/L			
		_Antagonist_ratio 33/3		observed val	lues for 140 cl	riptor value(le to missing	^		M: 0.0754 mg/L	in the tig/c			
		Antagonist_viabi 34/			ueso	criptor value(»)					M: 0.407 mg/L			
		_MDAKB2_Agonist 32/: _MDAKB2_Antagonist 4									M: 0.2 mg/L M: 2.49 mg/L				
>		_MDAKB2_Antagonist 4						01			Wi. 2.45 Mg/L	M: 0.000198 mg/	L		
								OK				-			
Data Gap Filling Settings	+ Tox21_Aroma														
		ase_Inhibition 22/3 are Inhibition via 36/3			M-8.25 mn/l	M-192 mn/l		Λ					M-0.00118 mg/l		
Only endpoint relevant	<				M: 823 mn/l			1					M-0.00118 mg/l		_
Only endpoint relevant this position:					M: 873 mo/l	Trend ana Predicted	lysis prediction fo : 1.08 mg/L		51 values				M-0.00118 mo/l	Select /	filter data
Only endpoint relevant this position: QSARs 0 Automated workflows 0	<				M: 8 23 mn/l	Trend ana	: 1.08 ma/L		51 values Kow, log(1/mol/L)				M-0.00118 mo/L		filter data g approach
Dnly endpoint relevant this position: 2SARs O Uutomated workflows O tandardized workflows O	C Descriptors Prediction	isca Inhihition uis 36/			M: 8 / 5 mn/l	Trend ana Predicted	: 1.08 ma/L						M-0.00118 mo/l	Gap fillir	
Only endpoint relevant this position: QSARs 0 Automated workflows 0 Standardized workflows 0 nedes below:	Descriptors Prediction Adequacy	10			M: 8.24 mn/l	Trend ana Predicted	: 1.08 ma/L						M-0.00118 mo/l	Gap fillir Descrip	g approach tors / data
Only endpoint relevant this position: QSARs 0 Automated workflows 0 standardized workflows 0 nodes below: QSARs 0 Automated workflows 0	Descriptors Prediction Adequacy Cumulative frequency	10 8 8			M: 8./4 mo/l	Trend ana Predicted	: 1.08 ma/L						M-0.00118 mo/l	Gap fillir Descrip Change de	g approach tors / data scriptor units
Only endpoint relevant this position: QSARs 0 Automated workflows 0 standardized workflows 0 nodes below: QSARs 0 Automated workflows 0	Descriptors Prediction Adequacy Cumulative frequency	10 8 8			M- 8 /4 mo/l	Trend ana Predicted	: 1.08 ma/L						M-0.00118.ma/l	Gap fillir Descrip Change de	g approach tors / data
Only endpoint relevant this position: QSARs 0 Automated workflows 0 nodes below: QSARs 0 Automated workflows 0	Descriptors Prediction Adequacy Cumulative frequency Residuals	10 8 8 1 8 1 1 1 1 1 1 1 1 1 1 1 1 1			M-X /4 mn/l	Trend ana Predicted	: 1.08 ma/L						M-00118 mo/l	Gap fillir Descrip Change de Edit descr	g approach tors / data scriptor units iptor options
Only endpoint relevant this position: QSARs 0 Automated workflows 0 nodes below: QSARs 0 Automated workflows 0	Descriptors Prediction Adequacy Cumulative frequency Residuals	10 8 8			Mr X / 4 mm/l	Trend ana Predicted	: 1.08 ma/L						M- 000118 mo/	Gap fillir Descrip Change de Edit descr Select endpoi	g approach tors / data scriptor units iptor options
Only endpoint relevant this position: QSARs 0 Automated workflows 0 nodes below: QSARs 0 Automated workflows 0	Descriptors Prediction Adequacy Cumulative frequency Residuals	10 8 8 1 8 1 1 1 1 1 1 1 1 1 1 1 1 1			M-X/4 mm/l	Trend ana Predicted	: 1.08 ma/L						M- 00118 mo/	Gap fillin Descrip Change du Edit descr Select endpoi	g approach tors / data scriptor units iptor options nt tree descrip te bins
Only endpoint relevant t this position: QSARs QSARs QSARs Q sandardized workflows Q solds below: QSARs QSARs Q subtomated workflows Q	Descriptors Prediction Adequacy Cumulative frequency Residuals	10 8 8 1 8 1 1 1 1 1 1 1 1 1 1 1 1 1			Mr X / 1 mm/l	Trend ana Predicted	: 1.08 ma/L							Gap fillir Descrip Change du Edit descr Select endpoi Crea Mod	g approach tors / data scriptor units iptor options nt tree descript

1. Click **"OK"** on the message informing you for the number of excluded chemicals due to missing X-descriptor data. They are analogues which do not have AC50 data for the assay "Tox21....", plotted on X-axis. This will not affect the value of correlation coefficient;

Continuous vs. continuous

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

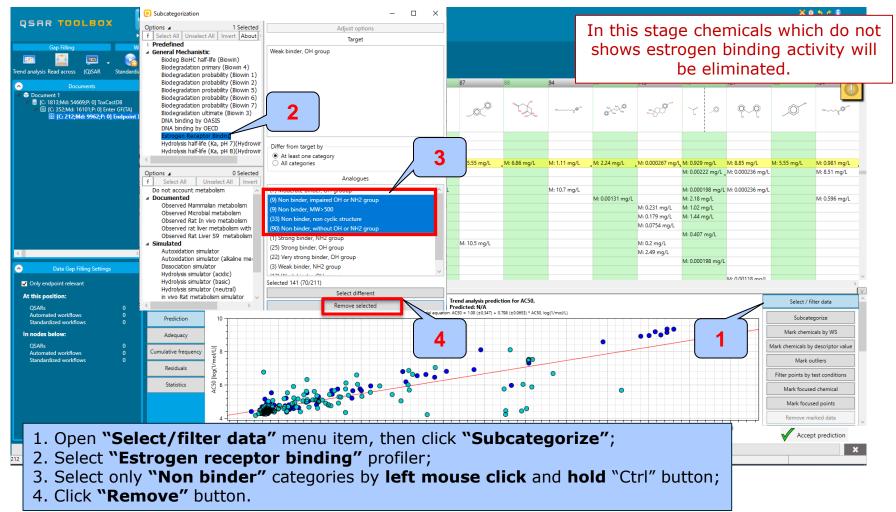
	Input rkflow	► Profiling ► Da	ita	Category definit	tion 🕨 Data Gap		Report								for G	DECD QSAR Toolt rouping Chemica Categories
analysis Read across (Q)SAR Standardizi	ed Autor	nated														loped by LMC, Bi
	Filter	endpoint tree	Ŷ	1 [target]	3	41	42	87	88	94	95	113	115	121	123	134
Document 1	Stru	cture		~0	*q	196~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	HOC CH3	_0_0	۲Ş.		0.4.4.C		~ ~	Q_Q	<u>,</u> ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
	Π	NCGC Reporter Gene Assay ER	a Agon													
		AC50	211/338		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_M: 0.929 mg/L	M: 8.85 ma/L	M: 5.55 mg/L	M: 0.981 mg/
		NCGC Reporter Gene Assay ER			M: 8.29 mg/L	M: 1.97 mg/L								M: 0.000236 mg/L		M: 8.51 mg/L
			49/49													
		Tox21_AhR_viability	37/31	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	M: 0.00131 mg/L		M: 0.000198 mg/l M: 2.18 mg/L	. M: 0.000236 mg/L		M: 0.596 mg
		Tox21_AR_BLA_Agonist_ch1 Tox21_AR_BLA_Agonist_ch2	34/34	-	M: 7.00 Mg/L	Wi: 8.57 mg/c					W: 0.00131 Mg/L	M: 0.231 mg/L	M: 1.02 mg/L			W: 0.590 mg.
		Tox21_AR_BLA_Agonist_ratio	25/25			M: 9.57 mg/L						M: 0.179 mg/L	M: 1.44 mg/L			
		Tox21_AR_BLA_Antagonist_rati	o 33/33									M: 0.0754 mg/L				
		Tox21_AR_BLA_Antagonist_vial			M: 8.74 mg/L	M: 0.12 mg/L							M: 0.407 mg/L			
		Tox21_AR_LUC_MDAKB2_Agon Tox21_AR_LUC_MDAKB2_Antag						M: 10.5 mg/L			_	M: 0.2 mg/L M: 2.49 mg/L				
,		Tox21_AR_LUC_MDAKB2_Antag			M: 15 mg/L							Mi 2AS Hig/C	M: 0.000198 mg/l	-		
Data Gap Filling Settings		Tox21_Aromatase_Inhibition	22/22								<u> </u>					
Only endpoint relevant		+ Toy21 Aromatace Inhibition vi	s 26/2/			M- 8 23 ma/l	M-192 mo/l							M-0.00118 mg/l		
this position:	_									_ 2						
DSARs 0		Descriptors						Trend analysis pre Predicted: N/A	ediction for AC50,	~ ~					Select	/ filter data
Automated workflows 0		Prediction					Model equation	n: AC50 = 1.09 (±0.347)	+ 0.798 (±0.0653) * AC	50, log(1/mol/L)					Gap filli	ng approach
itandardized workflows 0		Prediction												•		
nodes below:		Adequacy													Descri	otors / data
QSARs 0	Cum	ulative frequency 🗧 8						•	•		• • • • • • • • • • • • • • • • • • • •				Change d	escriptor units
automated workflows 0 tandardized workflows 0	Cum								~						Edit desc	riptor options
		Residuals	•				• • •	•								
		Alative frequen y TO 0 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4		•											Select endpo	nt tree descript
		Statistics Q	•		8000		•		••••		•				Cre	ate bins
			222			•									Mod	el/QSAR
				adand	•		•		8 ••	-						
		4 4 4		4.5	5 5.	5 6	6.	AC50 [log	(1/mol/L)]		1	9	9.5	10		tion options

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 binding activity we remove the "Non-binders" chemicals based on subcategorization by "Estrogen receptor binding by OASIS" profiler (illustrated on next slide)

Types endpoint correlations

Continuous vs. continuous



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

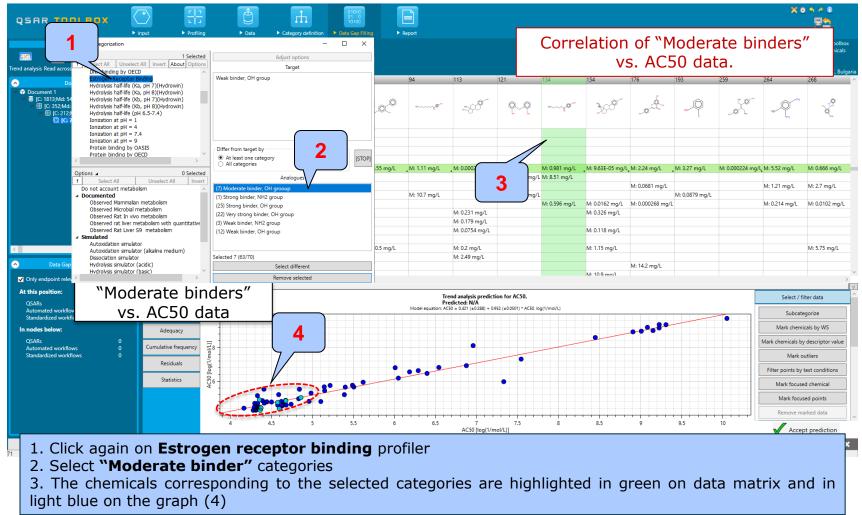
In the forthcoming slides are illustrated three endpoint vs. endpoint correlations:

- Correlation of "Moderate ER binders" vs. AC50 data;
- Correlation of "Weak ER binders" vs. AC50 data;
- Correlation of "Strong ER binders" vs. AC50 data.

The aim of the slides is to illustrate how the chemicals possessing ER binding potency correlate with AC50 data.

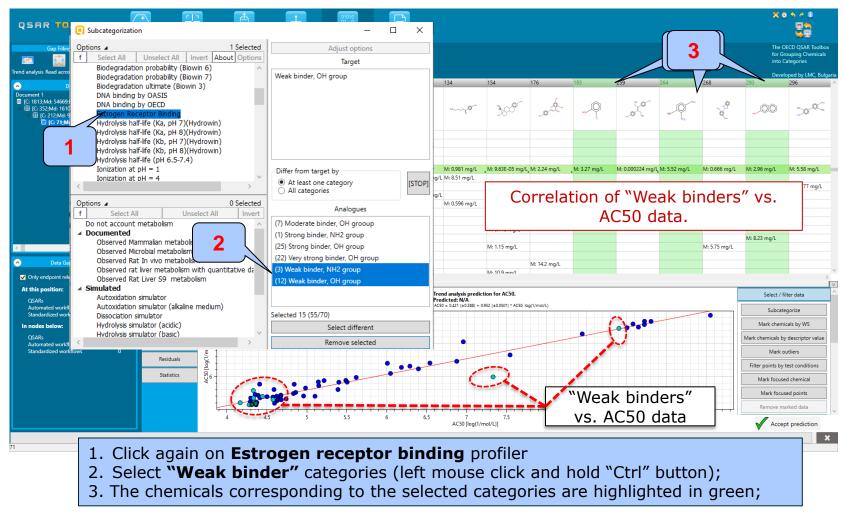
Types endpoint correlations

Continuous vs. continuous



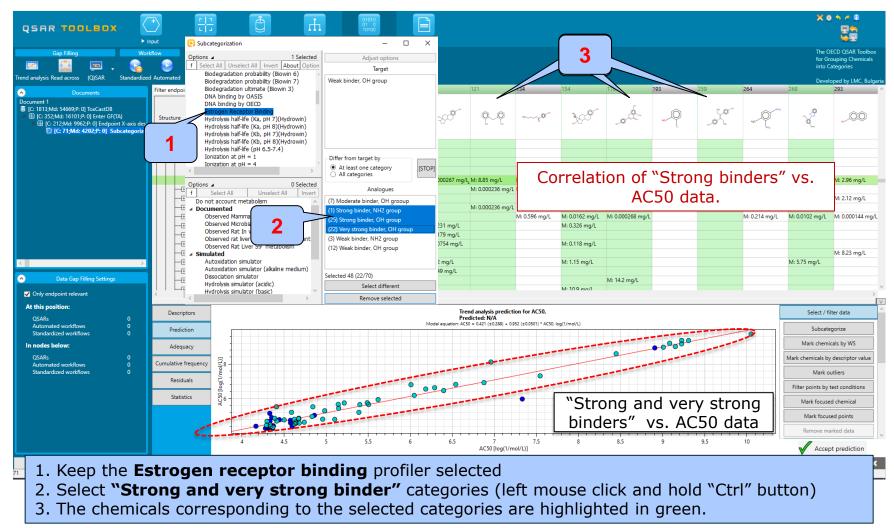
Types endpoint correlations

Continuous vs. continuous



Types endpoint correlations

Continuous vs. continuous



Types endpoint correlations Continuous vs. continuous *Correlation results*

- The two AC50 endpoints associated with different types of assays have been correlated each other
- Non binders according to the 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