

Here special possibilities of a new version of the CORALSEA-2020 are described briefly

The systematic development of the CORAL software is represented by versions: CORALSEA-2016 - CORALSEA-2017 - CORALSEA-2019 – CORALSEA-2020. There are special files comment.pdf in folders CORALSEA-2016, CORALSEA-2017 and CORALSEA-2019. User can get information on architecture and applying of the CORAL software in these files as well as in general file “ReadMe.pdf”.

The differences of CORALSEA-2020 in comparison with previous versions are the following.

- 1. Self-organized atoms pairs proportions vector.**
- 2. Molecular features contributions.**
- 3. Possibility to apply new types of the Monte Carlo optimization.**
- 4. New options for design of the quasi-SMILES codes.**
- 5. The calculation of the values of the endpoint using the current model for a list of the "dark SMILES" (or "dark quasi-SMILES").**

Below, the above-mentioned possibilities (1-5) are commented.

It is to be noted, that method.txt files for different versions are not interchangeable i.e. method.txt used for CORALSEA-2016 cannot be used for CORALSEA-2017, CORALSEA-19, and CORALSEA-2020; as well as method for CORALSEA-2017 cannot be applied for CORALSEA-2019 and CORALSEA-2020. The user should modify the CORAL-method step-by-step starting from method.txt that is provided in corresponding folder.

The example of the CORAL model represented in the folder CORALSEA-2020 related to data on anti-HIV activity from article:
[Structural Chemistry \(2020\) 31,1441–1448 \(for details please see pdf in the folder\)](#)

FYI:

You can send questions, suggestions, remarks using contact addresses:

andrey.toropov@marionegri.it,

alla.toropova@marionegri.it,

aatoropov@yahoo.com,

allatoropova@ymail.com

1. Self-organized atoms pairs proportions vector.

The interface of CORALSEA-2020 is the following. The atom pairs proportion APP vector controlled via the square indicated by red.

The screenshot displays the CORAL software interface, which is used for selecting and optimizing models for chemical structure prediction. The interface is divided into several sections:

- Top Bar:** Contains the CORAL logo and the text "CORAL: select Phase 1, Phase 2, or you can change and save method".
- Left Panel:** Shows three data input sections, each with a "CALC" label and an "EXPR" label:
 - Active Training set (TRN)
 - Passive Training set (iTRN)
 - Calibration set (CLB)
- Central Panel:** Contains two main phases:
 - Phase 1: Search for preferable model (T*,N*)
 - Phase 2: Building up preferable model (T*,N*)
- Right Panel:** Contains settings for the model selection process:
 - [quasi] - SMILES for TRN, iTRN, and CLB:** A text input field for "#TrainingSet.txt".
 - Method Selection:** Checkboxes for GRAPH, HSG, HFG, GAO, and SMILES. The SMILES checkbox is checked.
 - Atom Pairs Proportions:** A table showing proportions for various atom pairs (F, Cl, Br, N, O, S, P, =, #) across different methods (e0, e1, e2, e3, p2, p3, p4, s2, s3, nn).
 - Contributions:** Checkboxes for N, O, S, F, Cl, Br, =, #, and Cmax.
 - Model Selection:** Radio buttons for Classification model, CLASSIC SCHEME, and CORRELATION BALANCE. The CORRELATION BALANCE option is selected.
 - Index of Ideality of Correlation:** Checkboxes for Index of Ideality of Correlation and Correlation Intensity Index.
 - Parameters:** Input fields for D_start (0.5), d_limit (0.1), N_epoch (15), Threshold start (1), Threshold maximal (1), and Number of probes (1).
 - Import of current model:** A text input field for "Model Details.txt".
 - Quota:** A text input field for "Quota" (15).
 - Buttons:** "STOP and SAVE current CW's", "DemoDCW", "EvolutionCorr", and "EXIT".

Status	Comment																																																																																										
<div>Atoms pairs proportions</div> <table><tr><td></td><td>Cl</td><td>Br</td><td>N</td><td>O</td><td>S</td><td>P</td><td>=</td><td>#</td></tr><tr><td>F..</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>Cl....</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>Br.....</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>N.....</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>O.....</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>S.....</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>P.....</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td>=.....</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr><tr><td colspan="9">All</td></tr></table>		Cl	Br	N	O	S	P	=	#	F..	0	0	0	0	0	0	0	0	Cl....	0	0	0	0	0	0	0	0	Br.....	0	0	0	0	0	0	0	0	N.....	0	0	0	0	0	0	0	0	O.....	0	0	0	0	0	0	0	0	S.....	0	0	0	0	0	0	0	0	P.....	0	0	0	0	0	0	0	0	=.....	0	0	0	0	0	0	0	0	All									<p>The button “all” gives possibility to switch status of all buttons from the square into the following status: 0, 1, and 2.</p> <p>Status 0 means that APP vector is not involved in the modeling process.</p>
	Cl	Br	N	O	S	P	=	#																																																																																			
F..	0	0	0	0	0	0	0	0																																																																																			
Cl....	0	0	0	0	0	0	0	0																																																																																			
Br.....	0	0	0	0	0	0	0	0																																																																																			
N.....	0	0	0	0	0	0	0	0																																																																																			
O.....	0	0	0	0	0	0	0	0																																																																																			
S.....	0	0	0	0	0	0	0	0																																																																																			
P.....	0	0	0	0	0	0	0	0																																																																																			
=.....	0	0	0	0	0	0	0	0																																																																																			
All																																																																																											
<div>Atoms pairs proportions</div> <table><tr><td></td><td>Cl</td><td>Br</td><td>N</td><td>O</td><td>S</td><td>P</td><td>=</td><td>#</td></tr><tr><td>F..</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>Cl....</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>Br.....</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>N.....</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>O.....</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>S.....</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>P.....</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>=.....</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr></table>		Cl	Br	N	O	S	P	=	#	F..	1	1	1	1	1	1	1	1	Cl....	1	1	1	1	1	1	1	1	Br.....	1	1	1	1	1	1	1	1	N.....	1	1	1	1	1	1	1	1	O.....	1	1	1	1	1	1	1	1	S.....	1	1	1	1	1	1	1	1	P.....	1	1	1	1	1	1	1	1	=.....	1	1	1	1	1	1	1	1	<p>Status 1 means that pair-combinations (in format Yes-Yes) of indicated eight chemical elements together with double bond (=) and triple bond (#) will be involved as components for calculation of the optimal descriptor via their correlation weights (see Section 2.3. in ReadMe.pdf).</p> <p>Examples of codes for these pair combinations are (=..#)..Y.Y. // a molecule (SMILES) contains double bonds and triple bonds (F..#)..Y.Y. // a molecule (SMILES) contains fluorine atoms and triple bonds (N..O)..Y.Y. // a molecule (SMILES) contains nitrogen atoms and oxygen atoms</p>									
	Cl	Br	N	O	S	P	=	#																																																																																			
F..	1	1	1	1	1	1	1	1																																																																																			
Cl....	1	1	1	1	1	1	1	1																																																																																			
Br.....	1	1	1	1	1	1	1	1																																																																																			
N.....	1	1	1	1	1	1	1	1																																																																																			
O.....	1	1	1	1	1	1	1	1																																																																																			
S.....	1	1	1	1	1	1	1	1																																																																																			
P.....	1	1	1	1	1	1	1	1																																																																																			
=.....	1	1	1	1	1	1	1	1																																																																																			
<div>Atoms pairs proportions</div> <table><tr><td></td><td>Cl</td><td>Br</td><td>N</td><td>O</td><td>S</td><td>P</td><td>=</td><td>#</td></tr><tr><td>F..</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>Cl....</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>Br.....</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>N.....</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>O.....</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>S.....</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>P.....</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td></tr><tr><td>=.....</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td></tr></table>		Cl	Br	N	O	S	P	=	#	F..	2	2	2	2	2	2	2	2	Cl....	2	2	2	2	2	2	2	2	Br.....	2	2	2	2	2	2	2	2	N.....	2	2	2	2	2	2	2	2	O.....	2	2	2	2	2	2	2	2	S.....	2	2	2	2	2	2	2	2	P.....	2	2	2	2	2	2	2	2	=.....	2	2	2	2	2	2	2	2	<p>Status 2 means that pair-combination of indicated eight chemical elements together with double bond (=) and triple bond (#) will be involved as components for calculation of the optimal descriptor via their correlation weights (see Section 2.3. in ReadMe.pdf).</p> <p>Examples of codes for these pair combinations are (=..#)..2.1. // a molecule (SMILES) contains two double bonds and one triple bond (F..=)..3.2. // a molecule (SMILES) contains three fluorine atoms and two double bonds (N..O)..3.7. // a molecule (SMILES) contains three atoms of nitrogen and seven atoms of oxygen</p>									
	Cl	Br	N	O	S	P	=	#																																																																																			
F..	2	2	2	2	2	2	2	2																																																																																			
Cl....	2	2	2	2	2	2	2	2																																																																																			
Br.....	2	2	2	2	2	2	2	2																																																																																			
N.....	2	2	2	2	2	2	2	2																																																																																			
O.....	2	2	2	2	2	2	2	2																																																																																			
S.....	2	2	2	2	2	2	2	2																																																																																			
P.....	2	2	2	2	2	2	2	2																																																																																			
=.....	2	2	2	2	2	2	2	2																																																																																			

User can define any configuration of the APP vector by means of click corresponding buttons in the above square, e.g.:

Atoms pairs proportions									
	Cl	Br	N	O	S	P	=	#	
F..	2	2	2	2	2	2	2	2	
Cl....	2	2	2	2	2	2	2	2	
Br.....	1	1	1	2	2	2			
N.....		1	1	2	2	2			
O.....			1	2	2	2			
S.....				1	1	1			
P.....					1	1			
=.....							1		

Atoms pairs proportions									
	Cl	Br	N	O	S	P	=	#	
F..	1	1	1	1	1	1	1	1	
Cl....	1	1	1	1	1	1	1	1	
Br.....		2	2	2	1	1	1		
N.....			2	2	1	1	1		
O.....				2	1	1	1		
S.....					2	2	2		
P.....						2	2		
=.....							2		

Atoms pairs proportions									
	Cl	Br	N	O	S	P	=	#	
F..	1	1	1	1	1	1	0	0	
Cl....	1	1	1	1	1	0	0		
Br.....		2	2	2	1	0	0		
N.....			2	2	1	0	0		
O.....				2	1	0	0		
S.....					2	2	2		
P.....						2	2		
=.....							2		

Atoms pairs proportions									
	Cl	Br	N	O	S	P	=	#	
F..	2	2	2	2	2	2	1	1	
Cl....	2	2	2	2	2	2	1	1	
Br.....		0	0	0	2	1	1		
N.....			0	0	2	1	1		
O.....				0	2	1	1		
S.....					0	0	0		
P.....						0	0		
=.....							0		

If the corresponding pair is absent or in other words, one of the chemical elements (as well as a double or triple bond) is absent the component of APP vector will not be involved for the modelling process. Thus, the self-organization of the vector is the selection of pairs or APP which indeed have a presence in the molecule (SMILES).

2. Molecular features contributions.

The molecular contributions indicated by red frame.

CORAL: select Phase 1, Phase 2, or you can change and save method

Active Training set (TRN) → EXPR

Passive Training set (iTRN) → EXPR

Calibration set (CLB) → EXPR

Load method Save method Method.txt

Phase 1: Search for preferable model (T*,N*)

Split #TotalSet.txt into training and validation sets

Phase 2: Building up preferable model (T*,N*)

Import of current model

Model Details.txt

Quota 15

STOP and SAVE current CW's

EXIT

[quasi] - SMILES for TRN, iTRN, and CLB

#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES

☒ s

☒ ss

☒ sss

☒ BOND

☒ NOSP

☒ HALO

☒ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F...	0	0	1	0	0	0	0	1
Cl...	0	0	1	0	0	1	0	0
Br...	0	0	1	1	0	0	0	0
N...	0	1	1	0	0	0	0	0
O...	0	0	1	0	0	1	0	0
S...	0	0	0	1	0	0	1	0
P...	0	0	0	0	1	0	0	0
=	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S

☐ F ☐ Cl ☐ Br

☐ = ☐ # ☐ Cmax

☐ Classification model

☐ CLASSIC SCHEME

☒ CORRELATION BALANCE

☒ Index of Ideality of Correlation

☐ Correlation Intensity Index

dR weight 0,1

IIC_{clb} 0,5 0

D_{start} 0,5 d_{limit} 0,1 N_{epoch} 15

Threshold start 1

Threshold maximal 1

Number of probes 1

W% N111 N110 N101 N100 Nall Defect

Split Info 0 0 0 0 0 0 0

☐ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

DemoDCW EvolutionCorr

Some examples explain the above options are the following.

Status	Comment
Contributions <input type="checkbox"/> N <input type="checkbox"/> O <input type="checkbox"/> S <input type="checkbox"/> F <input type="checkbox"/> Cl <input type="checkbox"/> Br <input type="checkbox"/> = <input type="checkbox"/> # <input type="checkbox"/> Cmax	The number of chemical elements (N, O, S, F, Cl, Br); the number of double and triple bonds (=,#); and the number of cycles (rings) in the molecule (SMILES) are not involved for building up model.
Contributions <input checked="" type="checkbox"/> N <input type="checkbox"/> O <input type="checkbox"/> S <input type="checkbox"/> F <input type="checkbox"/> Cl <input type="checkbox"/> Br <input type="checkbox"/> = <input type="checkbox"/> # <input type="checkbox"/> Cmax	The total number of nitrogen atoms will be applied to develop model via correlation weights of the code <N>..0001... // the number of nitrogen atoms is one <N>..0002... // the number of nitrogen atoms is two <N>..000x... // the number of nitrogen atoms is x
Contributions <input type="checkbox"/> N <input checked="" type="checkbox"/> O <input type="checkbox"/> S <input type="checkbox"/> F <input type="checkbox"/> Cl <input type="checkbox"/> Br <input type="checkbox"/> = <input type="checkbox"/> # <input type="checkbox"/> Cmax	The total number of oxygen atoms will be applied to develop model via correlation weights of the code <O>..0003... // the number of oxygen atoms is three <O>..0005... // the number of oxygen atoms is five <O>..000x... // the number of oxygen atoms is x
Contributions <input type="checkbox"/> N <input type="checkbox"/> O <input type="checkbox"/> S <input type="checkbox"/> F <input type="checkbox"/> Cl <input type="checkbox"/> Br <input checked="" type="checkbox"/> = <input checked="" type="checkbox"/> # <input type="checkbox"/> Cmax	The total number of double bonds and the total number of triple bonds will be applied to develop model via correlation weights for codes <=>..000x... // x≠0 <#>..000y... // y≠0
Contributions <input checked="" type="checkbox"/> N <input checked="" type="checkbox"/> O <input checked="" type="checkbox"/> S <input checked="" type="checkbox"/> F <input checked="" type="checkbox"/> Cl <input checked="" type="checkbox"/> Br <input checked="" type="checkbox"/> = <input checked="" type="checkbox"/> # <input checked="" type="checkbox"/> Cmax	The number of chemical elements (N, O, S, F, Cl, Br); the number of double and triple bonds (=,#); and the number of cycles (rings) in the molecule (SMILES) are involved for building up model.

3. Possibility to apply new types of the Monte Carlo optimization.

The new possibilities of the Monte Carlo optimization are based on two new criteria of predictive potential (Table 1).

- The index of ideality of correlation (IIC);
- The correlation intensity index (CII).

CORAL: select Phase 1, Phase 2, or you can change and save method

Active Training set (TRN)

...

Passive Training set (iTRN)

...

Calibration set (CLB)

...

Load method Method.txt

Phase 1: Search for preferable model (T*,N*)

Split #TotalSet.txt into training and validation sets

Phase 2: Building up preferable model (T*,N*)

Import of current model

Model Details.txt

Quota 15

[quasi] - SMILES for TRN, iTRN, and CLB
#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO ☒ SMILES

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ \$ ☒ ss ☐ sss ☐ BOND ☐ NOSP ☐ HALO ☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F...	0	0	1	0	0	0	0	1
Cl...	0	0	1	0	0	1	0	0
Br...	0	0	1	1	0	0	0	0
N...	0	1	1	0	0	0	0	0
O...	0	0	1	0	0	0	1	0
S...	0	0	0	0	1	0	0	0
P...	0	0	0	0	0	1	0	0
=...	0	0	0	0	0	0	1	0

Contributions

☐ N ☐ O ☐ S ☐ F ☐ Cl ☐ Br ☐ # ☐ Cmax

☐ Classification model
☐ CLASSIC SCHEME
☒ CORRELATION BALANCE

☒ Index of Ideality of Correlation IIC_{clb} 0,5 0
☒ Correlation Intensity Index CII_{clb} 0,2 0

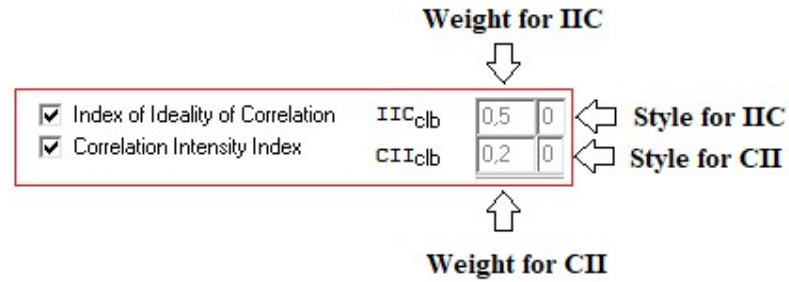
d_{start} 0,5 d_{limit} 0,1 N_{epoch} 15
Threshold start 1
Threshold maximal 1
Number of probes 1

W% N111 N110 N101 N100 Nall Defect
Split Info 0 0 0 0 0 0 0

☐ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

☒ DemoDCW ☒ EvolutionCorr

General options to control the Monte Carlo optimization



The basic equation for the target function (TF) used in Monte Carlo optimization are the following.

$$TF_1 = r_{TRN} + r_{iTRN} - |r_{TRN} - r_{iTRN}| * dr_{weight}$$

$$TF_2 = TF_1 + IIC_{CLB} * WeightForIIC$$

$$TF_3 = TF_1 + IIC_{iTRN} * WeightForIIC$$

$$TF_4 = TF_1 + (IIC_{CLB} + IIC_{iTRN}) * WeightForIIC$$

$$TF_5 = TF_1 + CII_{CLB} * WeightForCII$$

$$TF_6 = TF_1 + CII_{iTRN} * WeightForCII$$

$$TF_7 = TF_1 + (CII_{CLB} + CII_{iTRN}) * WeightForCII$$

The observed and calculated are corresponding values of an endpoint.

Status		Comment
<input type="checkbox"/> Index of Ideality of Correlation <input type="checkbox"/> Correlation Intensity Index		The optimization without applying <i>IIC</i> and <i>CII</i> , i.e. target function is <i>TF₁</i>
<input checked="" type="checkbox"/> Index of Ideality of Correlation <input type="checkbox"/> Correlation Intensity Index	<i>IIC</i> _{clb} 0,5 0	Optimization with <i>TF₂</i> , <i>WeightForIIC</i> =0.5
<input checked="" type="checkbox"/> Index of Ideality of Correlation <input type="checkbox"/> Correlation Intensity Index	<i>IIC</i> _{itrn} 0,5 1	Optimization with <i>TF₃</i> , <i>WeightForIIC</i> =0.5
<input checked="" type="checkbox"/> Index of Ideality of Correlation <input type="checkbox"/> Correlation Intensity Index	<i>IIC</i> _{itrn,clb} 0,5 2	Optimization with <i>TF₄</i> , <i>WeightForIIC</i> =0.5
<input type="checkbox"/> Index of Ideality of Correlation <input checked="" type="checkbox"/> Correlation Intensity Index	<i>CII</i> _{clb} 0,2 0	Optimization with <i>TF₅</i> , <i>WeightForCII</i> =0.2
<input type="checkbox"/> Index of Ideality of Correlation <input checked="" type="checkbox"/> Correlation Intensity Index	<i>CII</i> _{itrn} 0,2 1	Optimization with <i>TF₆</i> , <i>WeightForCII</i> =0.2
<input type="checkbox"/> Index of Ideality of Correlation <input checked="" type="checkbox"/> Correlation Intensity Index	<i>CII</i> _{itrn,clb} 0,2 2	Optimization with <i>TF₇</i> , <i>WeightForCII</i> =0.2
<input checked="" type="checkbox"/> Index of Ideality of Correlation <input checked="" type="checkbox"/> Correlation Intensity Index	<i>IIC</i> _{itrn} 0,3 1 <i>CII</i> _{itrn,clb} 0,5 2	Optimization with target function <i>TF</i> = <i>TF₃</i> + <i>TF₇</i> ; <i>WeightsForIIC</i> =0.3; and <i>WeightsForCII</i> =0.5.

Table 1

Statistical criteria of the predictive potential for QSPR/QSAR models

Criterion of the predictive potential	Reference
$R = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$	Hemmateenejad et al. 2012
$Q^2 = 1 - \frac{\sum (y_k - \hat{y}_k)^2}{\sum (y_k - \bar{y}_k)^2}$	Shayanfar and Shayanfar 2014
$Q_{F1}^2 = 1 - \frac{[\sum_{i=1}^{N_{EXT}} (\hat{y}_i - y_i)^2] / N_{EXT}}{[\sum_{i=1}^{N_{EXT}} (y_i - \bar{y}_{TR})^2] / N_{EXT}}$	Chirico and Gramatica 2011
$Q_{F2}^2 = 1 - \frac{[\sum_{i=1}^{N_{EXT}} (\hat{y}_i - y_i)^2] / N_{EXT}}{[\sum_{i=1}^{N_{EXT}} (y_i - \bar{y}_{EXT})^2] / N_{EXT}}$	Chirico and Gramatica 2011
$Q_{F3}^2 = 1 - \frac{[\sum_{i=1}^{N_{EXT}} (\hat{y}_i - y_i)^2] / N_{EXT}}{[\sum_{i=1}^{N_{TR}} (y_i - \bar{y}_{TR})^2] / N_{TR}}$	Chirico and Gramatica 2011
$\overline{R}_m^2 = \frac{R_m^2(x, y) + R_m^2(y, x)}{2}$ $\Delta R_m^2 = R_m^2(x, y) - R_m^2(y, x) $	Roy and Kar 2014
$CCC = \frac{2 \sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2 + \sum (y - \bar{y})^2 + n(\bar{x} - \bar{y})^2}$	Lin 1992
$HC_{CLB} = r_{CLB} \frac{\min(-MAE_{CLB}, +MAE_{CLB})}{\max(-MAE_{CLB}, +MAE_{CLB})}$ $-MAE_{CLB} = \frac{1}{-N} \sum_{k=1}^{-N} \Delta_k , \text{ } -N \text{ is the number of } \Delta_k < 0$ $+MAE_{CLB} = \frac{1}{+N} \sum_{k=1}^{+N} \Delta_k , \text{ } +N \text{ is the number of } \Delta_k \geq 0$ $\Delta_k = observed_k - calculated_k$	Toropov and Toropova 2017
$CII = 1 - CCI;$ where $CCI = \sum (\Delta R_j^2 > 0)$ $\Delta R_j^2 = R_j^2 - R^2$	Toropov and Toropova 2019

References

- Chirico N, Gramatica P (2011) Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J Chem Inf Model* 51(9): 2320-2335. DOI: 10.1021/ci200211n
- Hemmateenejad B, Javidnia K, Miri R, Elyasi M (2012) Quantitative structure-retention relationship study of analgesic drugs by application of combined data splitting-feature selection strategy and genetic algorithm-partial least square. *J Iran Chem Soc* 9(1): 53-60. DOI: 10.1007/s13738-011-0005-z
- Lin LI-K (1992) Assay validation using the concordance correlation coefficient. *Biometrics* 48(2): 599-604. DOI: 10.2307/2532314
- Roy K, Kar S (2014) The r_m^2 metrics and regression through origin approach: Reliable and useful validation tools for predictive QSAR models (Commentary on 'Is regression through origin useful in external validation of QSAR models?') *Eur J Pharm Sci* 62:111-114. DOI: 10.1016/j.ejps.2014.05.019
- Shayanfar A, Shayanfar S (2014) Is regression through origin useful in external validation of QSAR models? *Eur J Pharm Sci* 59(1): 31-35. DOI: 10.1016/j.ejps.2014.03.007
- Toropov AA, Toropova AP (2017) The index of ideality of correlation: A criterion of predictive potential of QSPR/QSAR models? *Mutat Res Genet Toxicol Environ Mutagen* 819: 31-37. DOI: 10.1016/j.mrgentox.2017.05.008
- Toropov, A.A., Toropova, A.P. QSAR as a random event: criteria of predictive potential for a chance model (2019) *Structural Chemistry*, 30 (5), pp. 1677-1683. DOI: 10.1007/s11224-019-01361-6

4. New options for design of the quasi-SMILES codes.

Quasi-SMILES in the CORALSEA-2020 can be defined with sequences of symbols in square brackets, e.g.

ID [alpha][Betta][Gamma]...[epsilon] Endpoint.

For an example

```
1 [low_conc][t25C][light] endpoint1
2 [middle_conc][t30C][dark] endpoint2
3 [high_conc][t25C][dark] endpoint3
...
```

In real task it can be represented as

```
*1 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o24][p0,001] 92.530
*2 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o24][p0,01] 96.130
*3 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o24][p0,1] 93.560
#4 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o24][p1] 97.680
#5 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o24][p5] 94.850
#6 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o24][p10] 96.390
#7 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o24][p20] 92.270
-8 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o24][p50] 92.270
#9 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o24][p100] 82.730
*10 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o12][p0,001] 95.690
+11 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o12][p0,01] 94.480
+12 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o12][p0,1] 96.900
-13 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o12][p1] 99.480
-14 [aAl2O3][b39,7][c267][d36,3][e64,7][f-17,345][g-1,51][h-9,81][i5,67][jMTT][kHCMEC][LHuman][mBlood][nNormal][o12][p5] 94.660
```

• • •

T1

The screenshot displays the CORAL software interface, which is used for selecting Phase 1, Phase 2, or changing and saving the method. The interface is divided into several sections:

- Top Bar:** Contains the CORAL logo and the text "CORAL: select Phase 1, Phase 2, or you can change and save method".
- Left Panel:** Features three vertical axes labeled "CALC" and "EXPR". The first axis is associated with the "Active Training set (iTRN)" and the second with the "Passive Training set (iTRN)". The third axis is associated with the "Calibration set (CLB)".
- Central Panel:** Contains two main phases:
 - Phase 1:** Search for preferable model (T*,N*). It includes buttons for "Load method", "Save method", and a text field for "Method.txt".
 - Phase 2:** Building up preferable model (T*,N*). It includes a button for "Split #TotalSet.txt into training and validation sets".
- Right Panel:** Contains a section titled "[quasi] - SMILES for TRN, iTRN, and CLB" with a text field for "#TrainingSet.txt". Below this are checkboxes for "GRAPH", "HSG", "HFG", and "GAO". A table of checkboxes for "e0", "e1", "e2", "e3", "p2", "p3", "p4", "s2", "s3", and "nn" is present. To the right of this table are checkboxes for "R3", "R4", "R5", "R6", and "R7". Further right are checkboxes for "SMILES", "s", "ss", "sss", "BOND", "NOSP", "HALO", and "HARD". Below these are checkboxes for "Atoms pairs proportions" and a table of checkboxes for "Cl", "Br", "N", "O", "S", "P", and "#".
- Bottom Panel:** Contains a section titled "Contributions" with checkboxes for "N", "O", "S", "F", "Cl", "Br", "=", and "#". Below this are checkboxes for "Classification model", "CLASSIC SCHEME", and "CORRELATION BALANCE". A table of checkboxes for "Index of Ideality of Correlation" and "Correlation Intensity Index" is present. To the right of this table are checkboxes for "dR weight", "IIC_{itrn}", "CII_{itrn,clb}", "D_{start}", "d_{limit}", "N_{epoch}", "Threshold start", "Threshold maximal", and "Number of probes".
- Bottom Right:** Contains a section titled "Import of current model" with a text field for "Model Details.txt". Below this are checkboxes for "Search for duplicates in SMILES" and "Search for duplicates in CAS (ID)". To the right of these checkboxes are checkboxes for "DemoDCW" and "EvolutionCorr".

However, if one is using quasi-SMILES where each quasi-SMILES contains fragment which is the traditional SMILES, all SMILES codes (indicated by green) are available. This can be a sequence of lines similar to

-1 O=S(=O)(c1cccn1=O)Cc2ccccc2.[Dark][Low_Conc] 3.840

*2 O=S(=O)(c1cccn1=O)Cc2ccccc2C.[Light][High_Conc] 4.060

...

5. The calculation of the values of the endpoint using the current model for a list of the "dark SMILES" (or "dark quasi-SMILES").

Having preliminary built up a model one can click button "Import of current model"

CORAL: Loading of method or Import of model

Load method

Active Training set (TRN)

Passive Training set (iTRN)

Calibration set (CLB)

[quasi] - SMILES for TRN, iTRN, and CLB

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO ☐ SMILES

e0 e1 e2 e3 p2 p3 p4 s2 s3 nm ☐ R3 ☐ \$

☐ R4 ☐ \$\$

☐ R5 ☐ \$\$\$

☐ R6 ☐ BOND

☐ R7 ☐ NOSP

☐ HALO

☐ HARD

Atoms pairs proportions

Cl Br N O S P = #

F...

Cl...

Br...

N...

O...

S...

P...

=...

Contributions

☐ N ☐ O ☐ S

☐ F ☐ Cl ☐ Br

☐ = ☐ # ☐ Cmax

☐ Classification model

☐ CLASSIC SCHEME

☒ CORRELATION BALANCE

☐ Index of Ideality of Correlation

☐ Correlation Intensity Index

d_start d_limit N_epoch

Threshold start

Threshold maximal

Number of probes

Model Details.txt

Quota

☐ DemoDCW

☐ EvolutionCorr

Split Info

W%	N111	N110	N101	N100	Nall	Defect
0	0	0	0	0	0	0

☐ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

and to calculate the model value for a list of dark SMILES (clicking “Calculation Endpoint for dark SMILES”) placed in file list.txt. The results of the calculation will be placed in file Endpoint.txt.

CORAL: Calculation of model for external substances

Active Training set (TRN)
 n=116: R2=0,7029; s=0,951; MAE=0,739; F=270

Passive Training set (iTRN)
 n=110: R2=0,6905; s=0,856; MAE=0,672; F=241

Calibration set (CLB)
 n=111: R2=0,5890; s=0,953; MAE=0,759; F=156

DEMO plots "EXPR vs. CALC"

Phase 1: Search for preferable model (T*,N*)

Split #TotalSet.txt into training and validation sets

Phase 2: Building up preferable model (T*,N*)

Calculation Endpoint for dark SMILES

list.txt Endpoint.txt

CO = -2,5523094 C1 = 0,2183696

Insert a SMILES for calculation of DCW and EndPoint

Demo of calculation of DCW and endpoint will be in file

DCW(1,15)= Endpoint =

Start of DCW and Endpoint calculation for SMILES from file

Import of current model

W% N111 N110 N101 N100 Nall Defect

Split Info 85 92 3 6 7 108 10,0872

[quasi] - SMILES for TRN, iTRN, and CLB

#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES ☒ s ☒ ss ☒ sss ☒ BOND ☒ NOSP ☒ HALO ☒ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F...	0	0	1	0	0	0	0	1
Cl...	0	0	1	0	0	1	0	0
Br...	0	0	1	1	0	0	0	0
N...	0	1	1	0	0	0	0	0
O...	0	0	1	0	0	1	0	0
S...	0	0	0	1	0	0	1	0
P...	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0

Contributions

☒ N ☒ O ☐ S ☐ F ☐ Cl ☐ Br ☒ Cmax

☐ Classification model ☐ CLASSIC SCHEME ☒ CORRELATION BALANCE

☒ Index of Ideality of Correlation ☒ Correlation Intensity Index

dR weight 0,1 IICitn 0,3 1 CIIitn,clb 0,5 2

D_start 0,5 d_limit 0,1 N_epoch 15

DemoDCW.txt Threshold start 1

#ValidationSet.txt Threshold maximal 1

Model Details.txt Number of probes 1

#ModelForValidationSet.txt

Quota 15

☐ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

Continue optimization **STOP and SAVE current CWs** ☒ DemoDCW ☒ EvolutionCorr **EXIT**

The “list.txt” can be

```
O[C@@H](C(=O)N1COC2([C@H]1C(=O)NCc1c(C)nsc1C)CCCC2)C[C@H](C(=O)N[C@H]1[C@@H](O)COc2c1cccc2)Cc1cccc1  
O[C@@H](C(=O)N1COC2([C@H]1C(=O)NCc1c(C)noc1C)CCCC2)C[C@H](C(=O)N[C@H]1[C@@H](O)COc2c1cccc2)Cc1cccc1  
O[C@@H](C(=O)N1COC2([C@H]1C(=O)NCc1c(C)cccc1C)CCCC2)C[C@H](C(=O)N[C@H]1[C@@H](O)COc2c1cccc2)Cc1cccc1
```

The corresponding “Endpoint.txt” will be

```
O[C@@H](C(=O)N1COC2([C@H]1C(=O)NCc1c(C)nsc1C)CCCC2)C[C@H](C(=O)N[C@H]1[C@@H](O)COc2c1cccc2)Cc1cccc1 : 10,1285  
O[C@@H](C(=O)N1COC2([C@H]1C(=O)NCc1c(C)noc1C)CCCC2)C[C@H](C(=O)N[C@H]1[C@@H](O)COc2c1cccc2)Cc1cccc1 : 10,0883  
O[C@@H](C(=O)N1COC2([C@H]1C(=O)NCc1c(C)cccc1C)CCCC2)C[C@H](C(=O)N[C@H]1[C@@H](O)COc2c1cccc2)Cc1cccc1 : 10,0100
```

References // related to applying of the CORALSEA-2020

INDEX OF IDEALITY OF CORRELATION

- Toropov, A.A., Toropova, A.P. The index of ideality of correlation: A criterion of predictive potential of QSPR/QSAR models? (2017) Mutation Research - Genetic Toxicology and Environmental Mutagenesis, 819, pp. 31-37. Cited 57 times. DOI: 10.1016/j.mrgentox.2017.05.008
- Toropova, A.P., Toropov, A.A. The index of ideality of correlation: A criterion of predictability of QSAR models for skin permeability? (2017) Science of the Total Environment, 586, pp. 466-472. DOI: 10.1016/j.scitotenv.2017.01.198
- Toropova, A.P., Toropov, A.A. Fullerenes C60 and C70: a model for solubility by applying the correlation intensity index (2020) Fullerenes Nanotubes and Carbon Nanostructures, 28 (11), pp. 900-906. DOI: 10.1080/1536383X.2020.1779705
- Toropov, A.A., Sizochenko, N., Toropova, A.P., Leszczynska, D., Leszczynski, J. Advancement of predictive modeling of zeta potentials (ζ) in metal oxide nanoparticles with correlation intensity index (CII) (2020) Journal of Molecular Liquids, 317, art. no. 113929, . DOI: 10.1016/j.molliq.2020.113929
- Toropov, A.A., Toropova, A.P. Correlation intensity index: Building up models for mutagenicity of silver nanoparticles (2020) Science of the Total Environment, 737, art. no. 139720, . DOI: 10.1016/j.scitotenv.2020.139720
- Stošić, B., Janković, R., Stošić, M., Marković, D., Stanković, D., Sokolović, D., Veselinović, A.M. In silico development of anesthetics based on barbiturate and thiobarbiturate inhibition of GABAA (2020) Computational Biology and Chemistry, 88, art. no. 107318, . DOI: 10.1016/j.combiolchem.2020.107318
- Kumar, A., Kumar, P. Construction of pioneering quantitative structure activity relationship screening models for abuse potential of designer drugs using index of ideality of correlation in monte carlo optimization (2020) Archives of Toxicology, 94 (9), pp. 3069-3086. DOI: 10.1007/s00204-020-02828-w
- Kumar, P., Kumar, A. Nucleobase sequence based building up of reliable QSAR models with the index of ideality correlation using Monte Carlo method (2020) Journal of Biomolecular Structure and Dynamics, 38 (11), pp. 3296-3306. DOI: 10.1080/07391102.2019.1656109
- Bagri, K., Kumar, A., Nimbhal, M., Kumar, P. Index of ideality of correlation and correlation contradiction index: a confluent perusal on acetylcholinesterase inhibitors (2020) Molecular Simulation, 46 (10), pp. 777-786. DOI: 10.1080/08927022.2020.1770753

- Toropova, A.P., Duchowicz, P.R., Saavedra, L.M., Castro, E.A., Toropov, A.A. The Use of the Index of Ideality of Correlation to Build Up Models for Bioconcentration Factor (2020) *Molecular Informatics*, 39 (7), art. no. 1900070, . DOI: 10.1002/minf.201900070
- Kostić, T., Deljanin Ilić, M., Perišić, Z., Milić, D., Đorđević, M., Golubović, M., Koraćević, G., Šalinger Martinović, S., Ćirić Zdravković, S., Živić, S., Lazarević, M., Stanojević, D., Dakić, S., Lilić, J., Veselinović, A. Design and development of novel therapeutics for coronary heart disease treatment based on cholesteryl ester transfer protein inhibition–in silico approach (2020) *Journal of Biomolecular Structure and Dynamics*, 38 (8), pp. 2304-2313. DOI: 10.1080/07391102.2019.1630319
- Kumar, P., Kumar, A. CORAL: QSAR models of CB1 cannabinoid receptor inhibitors based on local and global SMILES attributes with the index of ideality of correlation and the correlation contradiction index (2020) *Chemometrics and Intelligent Laboratory Systems*, 200, art. no. 103982, . DOI: 10.1016/j.chemolab.2020.103982
- Zivkovic, M., Zlatanovic, M., Zlatanovic, N., Djordjevic Jocic, J., Golubović, M., Veselinović, A.M. Development of novel therapeutics for the treatment of glaucoma based on actin-binding kinase inhibition -in silico approach (2020) *New Journal of Chemistry*, 44 (17), pp. 6923-6931. DOI: 10.1039/c9nj05967a
- Ničković, V.P., Mitić, N.R., Krdžić, B.D., Krdžić, J.D., Nikolić, G.R., Vasić, M.Z., Ranković, G., Babović, P., Sokolović, D., Veselinović, A.M. Design and development of novel therapeutics for brucellosis treatment based on carbonic anhydrase inhibition (2020) *Journal of Biomolecular Structure and Dynamics*, 38 (6), pp. 1848-1857. DOI: 10.1080/07391102.2019.1619626
- Nimbhal, M., Bagri, K., Kumar, P., Kumar, A. The index of ideality of correlation: A statistical yardstick for better QSAR modeling of glucokinase activators (2020) *Structural Chemistry*, 31 (2), pp. 831-839. DOI: 10.1007/s11224-019-01468-w
- Toropova, A.P., Toropov, A.A., Leszczynska, D., Leszczynski, J. The index of ideality of correlation: Models of the flash points of ternary mixtures (2020) *New Journal of Chemistry*, 44 (12), pp. 4858-4868. DOI: 10.1039/d0nj00121j
- Jain, S., Amin, S.A., Adhikari, N., Jha, T., Gayen, S. Good and bad molecular fingerprints for human rhinovirus 3C protease inhibition: identification, validation, and application in designing of new inhibitors through Monte Carlo-based QSAR study (2020) *Journal of Biomolecular Structure and Dynamics*, 38 (1), pp. 66-77. DOI: 10.1080/07391102.2019.1566093
- Kumar, A., Sindhu, J., Kumar, P. In-silico identification of fingerprint of pyrazolyl sulfonamide responsible for inhibition of N-myristoyltransferase using Monte Carlo method with index of ideality of correlation (2020) *Journal of Biomolecular Structure and Dynamics*, in press, DOI: 10.1080/07391102.2020.1784286

CORRELATION CONTRADICTION INDEX

- Toropov, A.A., Toropova, A.P. QSAR as a random event: criteria of predictive potential for a chance model (2019) *Structural Chemistry*, 30 (5), pp. 1677-1683. DOI: 10.1007/s11224-019-01361-6
- Toropov, A.A., Toropova, A.P. The Correlation Contradictions Index (CCI): Building up reliable models of mutagenic potential of silver nanoparticles under different conditions using quasi-SMILES (2019) *Science of the Total Environment*, 681, pp. 102-109. DOI: 10.1016/j.scitotenv.2019.05.114
- Kumar, P., Kumar, A. CORAL: QSAR models of CB1 cannabinoid receptor inhibitors based on local and global SMILES attributes with the index of ideality of correlation and the correlation contradiction index (2020) *Chemometrics and Intelligent Laboratory Systems*, 200, art. no. 103982, . DOI: 10.1016/j.chemolab.2020.103982
- Toropova, A.P., Toropov, A.A. Whether the validation of the predictive potential of toxicity models is a solved task? (2019) *Current Topics in Medicinal Chemistry*, 19 (29), pp. 2643-2657. DOI: 10.2174/1568026619666191105111817
- Bagri, K., Kumar, A., Nimbhal, M., Kumar, P. Index of ideality of correlation and correlation contradiction index: a confluent perusal on acetylcholinesterase inhibitors (2020) *Molecular Simulation*, 46 (10), pp. 777-786. DOI: 10.1080/08927022.2020.1770753

CORRELATION INTENSITY INDEX

- Toropova, A.P., Toropov, A.A. Fullerenes C60 and C70: a model for solubility by applying the correlation intensity index (2020) *Fullerenes Nanotubes and Carbon Nanostructures*, 28 (11), pp. 900-906. DOI: 10.1080/1536383X.2020.1779705
- Toropov, A.A., Sizochenko, N., Toropova, A.P., Leszczynska, D., Leszczynski, J. Advancement of predictive modeling of zeta potentials (ζ) in metal oxide nanoparticles with correlation intensity index (CII) (2020) *Journal of Molecular Liquids*, 317, art. no. 113929, . DOI: 10.1016/j.molliq.2020.113929
- Toropov, A.A., Toropova, A.P. Correlation intensity index: Building up models for mutagenicity of silver nanoparticles (2020) *Science of the Total Environment*, 737, art. no. 139720, . DOI: 10.1016/j.scitotenv.2020.139720
- Jafari, K., Fatemi, M.H., Toropova, A.P., Toropov, A.A. Correlation Intensity Index (CII) as a criterion of predictive potential: Applying to model thermal conductivity of metal oxide-based ethylene glycol nanofluids (2020) *Chemical Physics Letters*, 754, art. no. 137614, . DOI: 10.1016/j.cplett.2020.137614