

Main instruction are available in File “ReadMe.pdf” available at <http://www.insilico.eu/coral>  
This is placed in CORALSEA.zip that can be downloaded from the above link.

**Here special possibilities of new version of the software (suggested in 2019) are described briefly.**

The systematic development of the CORAL software is represented by versions:  
CORALSEA-2016 - CORALSEA-2017 - CORALSEA-2019

There are special files comment.pdf in folders CORALSEA-2016 and CORALSEA-2017. User can get information on architecture and applying of the CORAL software in these files.

The factual difference of CORALSEA-2019 in comparison with CORALSEA-2017 is the following.

1. **The "Square of binary combinations of topological invariants" of molecular graph**
2. **Additional SMILES attributes: Cmax, Nmax, Omax, and Smax**
3. **Possibility to stop the Monte Carlo optimization**
4. **Possibility to continue the Monte Carlo optimization**

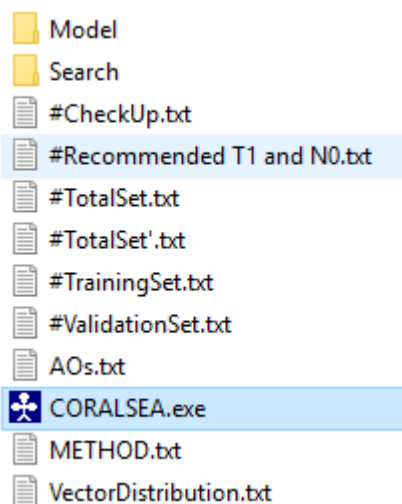
Below, the above-mentioned possibilities (1-4) are commented.

It is to be noted, that method.txt files for different versions are not interchangeable i.e. method.txt used for CORALSEA-2017 cannot be used for CORALSEA-2019.

The run of this new program (2019) is the sequence of steps:

First, enter in folder CORALSEA-2019

Second, click CORALSEA.exe



# 1. The "Square of binary combinations of topological invariants" of molecular graph

The interface of the CORALSEA-2019 is the following

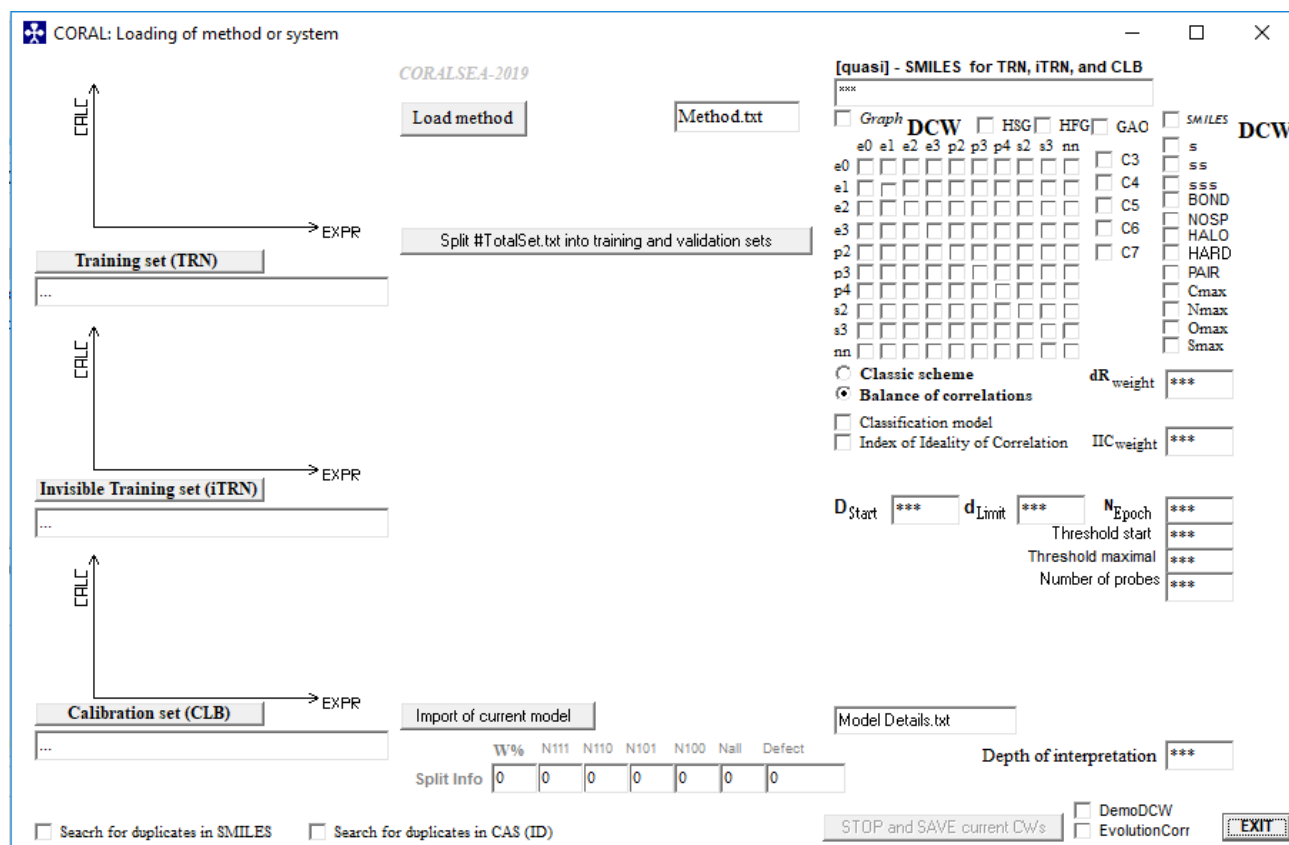


Table 1 contains the meaning for graph invariants “e0 ... nn”.

Table 1

List of Local graph invariants involved in building up the CORAL models (please see also section 2.1 in ReadMe.pdf as well as Table “Components of methods” in Comment.pdf in CORALSEA-2017).

Invariant	Comment
e0	Vertex degree (Morgan’s extended connectivity of zero order)
e1	Morgan’s extended connectivity of first order
e2	Morgan’s extended connectivity of second order
e3	Morgan’s extended connectivity of third order
p2	The number pf paths of length 2 which started from a given vertex in graph
p3	The number pf paths of length 3 which started from a given vertex in graph
p4	The number pf paths of length 4 which started from a given vertex in graph
s2	Valence shell of second order
s3	Valence shell of third order
nn	The nearest neighbours code

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

#TotalSet.txt

☒ *Graph* **DCW**
☒ HSG
 ☐ HFG
 ☐ GAO
 ☒ *SMILES* **DCW**

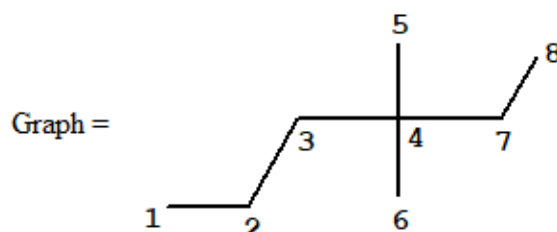
	e0	e1	e2	e3	p2	p3	p4	s2	s3	nn	
e0	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> C3
e1	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> C4
e2	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> C5
e3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> C6
p2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> C7
p3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
p4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
s2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
s3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
nn	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

☒ s
 ☒ ss
 ☒ sss
 ☒ BOND
 ☒ NOSP
 ☒ HALO
 ☒ HARD
 ☐ PAIR
 ☒ Cmax
 ☐ Nmax
 ☐ O max
 ☐ S max

Figure 1

Interface for definition of the optimal descriptor: the diagonal is place for traditional graph invariants: extended connectivity of zero (e0), first (e1), second (e2), third (e3) orders; paths of length two (p2), three (p3), four (p4), valence shells of second (s2) and third orders; and nearest neighbour code (nn). The selection of pair e1 and e2 (upper triangle) means absolute value of the arithmetic operation "e1 plus e2". The selection of pair e2 and e1 (low triangle) means absolute value for arithmetic operation "e1 minus e2". Figure 2 contains other combinations of the topological invariants of molecular graph.

SMILES = CCCC(C)(C)CC



Adjacency Matrix

	1	2	3	4	5	6	7	8
1	0	1	0	0	0	0	0	0
2	1	0	1	0	0	0	0	0
3	0	1	0	1	0	0	0	0
4	0	0	1	0	1	1	1	0
5	0	0	0	1	0	0	0	0
6	0	0	0	1	0	0	0	0
7	0	0	0	1	0	0	0	1
8	0	0	0	0	0	0	1	0

Graph invariants

e0	e1	e2	e3	p2	p3	p4	s2	s3	nn
1	2	3	8	1	1	3	2	4	110
2	3	8	12	1	3	1	4	4	220
2	6	9	27	4	1	0	5	1	220
4	6	19	29	2	1	0	3	1	440
1	4	6	19	3	2	1	5	3	110
1	4	6	19	3	2	1	5	3	110
2	5	8	24	3	1	1	4	2	220
1	2	5	8	1	3	1	4	4	110

$$\begin{aligned}
 e1_k &= 2,3,6,6,4,4,5,2 \\
 e2_k &= 3,8,9,19,6,6,8,5 \\
 e1_k + e2_k &= 5,11,15,25,10,10,13,7 \\
 e2_k - e1_k &= 1,5,3,13,2,2,3,3
 \end{aligned}$$

$$\begin{aligned}
 p2_k &= 1,1,4,2,3,3,3,1 \\
 p3_k &= 1,3,1,1,2,2,1,3 \\
 p2_k + p3_k &= 2,4,5,3,5,5,4,4 \\
 p3_k - p2_k &= 0,2,3,1,1,1,2,2
 \end{aligned}$$

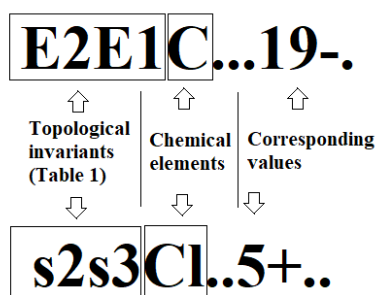
$$\begin{aligned}
 s2_k &= 2,4,5,3,5,5,4,4 \\
 s3_k &= 4,4,1,1,3,3,2,4 \\
 s3_k + s2_k &= 6,8,6,4,8,8,6,8 \\
 s3_k - s2_k &= 2,0,4,2,2,2,2,0
 \end{aligned}$$

Figure 2

Example of calculation of graph invariants and their combinations for the optimal descriptor in the case of 3,3-dimethylhexane. It should be noted, that similarly, one can select other combinations, e.g.  $e3_k \pm p4_k$ , ...  $e3_k \pm s2_k$ , ...,  $s3_k \pm nn_k$ , and others. Examples of records of these topological combinations are the following:

E2E1C...20-..  
E2E1C...21-..  
E2E1C...22-..  
E2E1C...23-..  
E2E1C...24-..  
E2E1C...25-..  
E2E1C...26-..

S2S3O...4+..  
S2S3O...5+..  
S2S3O...6+..  
S2S3O...7+..  
S2S3O...8+..  
S2S3O...9+..  
S2S3S...1+..  
S2S3S...10+..



## 2. Additional SMILES attributes: Cmax, Nmax, Omax, and Smax

Please see Table 2

Table 2

Clarifications for these global SMILES attributes

Attribute	Comment
Cmax	Total number of rings (the range 0 .. 9)
Nmax	Total number of nitrogen atoms in the molecular structure
Omax	Total number of oxygen atoms in the molecular structure
Smax	Total number of sulphur atoms in the molecular structure

## 3. Possibility to stop the Monte Carlo optimization

User can click button **STOP and SAVE current CWs** in order to stop the optimization, with save corresponding correlation weights (CWs).

**CORAL: Wait please...**

*CORALSEA-2019*

**Training set (TRN)**  
n=67: R2=0.8324; s=0.541; MAE=0.418; F=323

**Invisible Training set (iTRN)**  
n=67: R2=0.8490; s=0.522; MAE=0.379; F=366

**Calibration set (CLB)**  
n=39: R2=0.8885; s=0.433; MAE=0.349; F=295

Selected threshold is 1  
Epoch.....3 to 15

Phase 2: Building up preferable model (T\*,N\*)  
Define preferable threshold and press Continue 1 Continue  
C0 = 0 C1 = 1

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

☒ Graph **DCW** ☒ HSG ☐ HFG ☐ GAO ☒ SMILES **DCW**

e0 ☐ e1 ☐ e2 ☐ e3 ☐ p2 ☐ p3 ☐ p4 ☐ s2 ☐ s3 ☐ nn ☐ C3 ☐ C4 ☐ C5 ☐ C6 ☐ C7

☐ Classic scheme ☒ Balance of correlations  
☐ Classification model ☒ Index of Ideality of Correlation

dR\_weight 0.1  
IIC\_weight 0.2

D\_start 1.0 d\_Limit 0.1 N\_Epoch 15  
Threshold start 1  
Threshold maximal 1  
Number of probes 1

Depth of interpretation 15

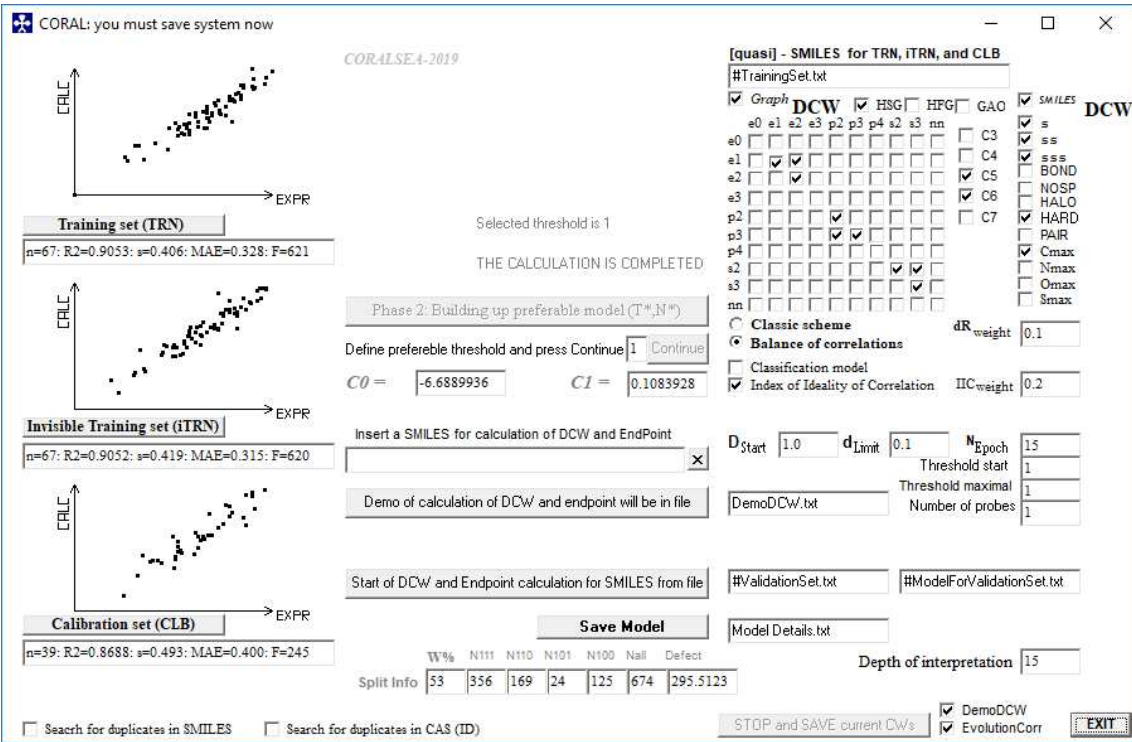
Split Info

W%	N111	N110	N101	N100	Nail	Defect
53	356	169	24	125	674	295.5123

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

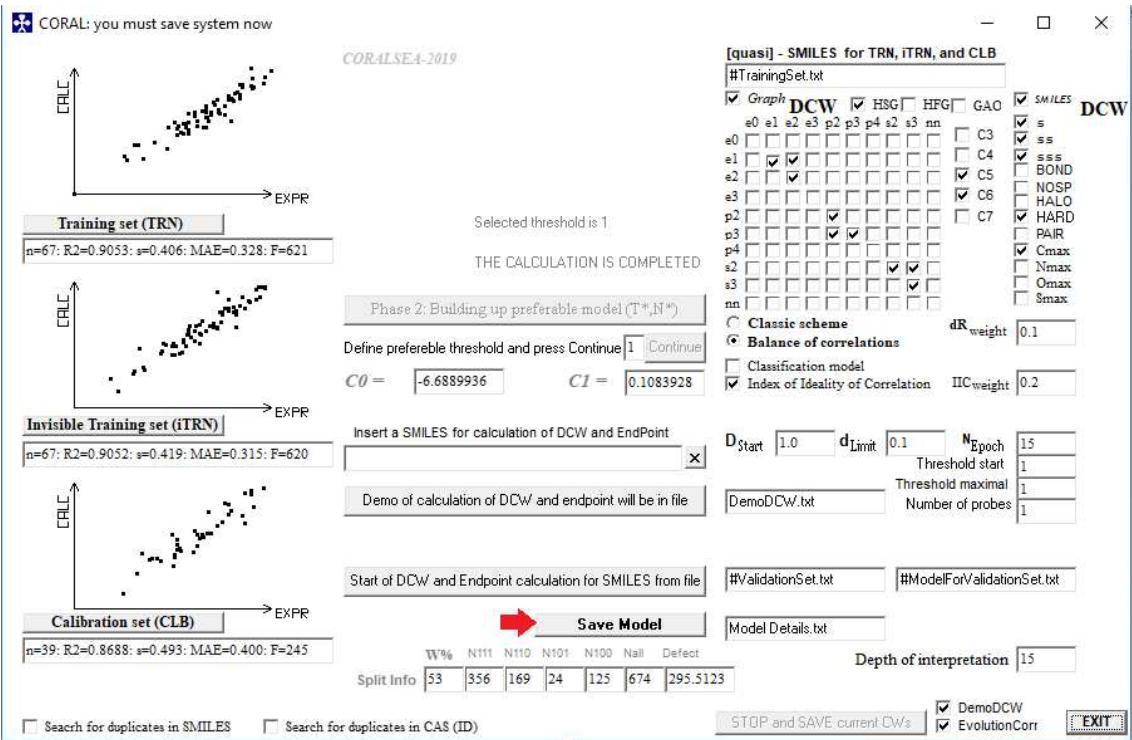
**STOP and SAVE current CWs** ☒ DemoDCW ☒ EvolutionCorr **EXIT**

Further event is the following:



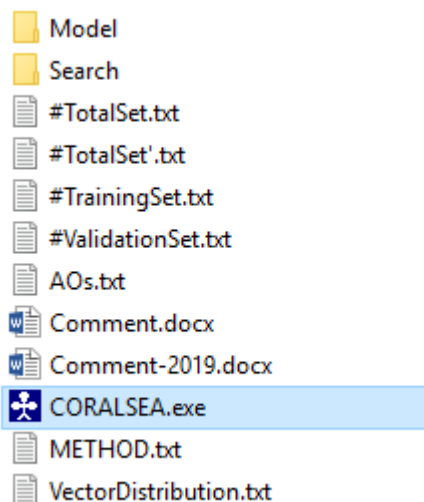
If you want to save the current status of the model, please click button

Save Model

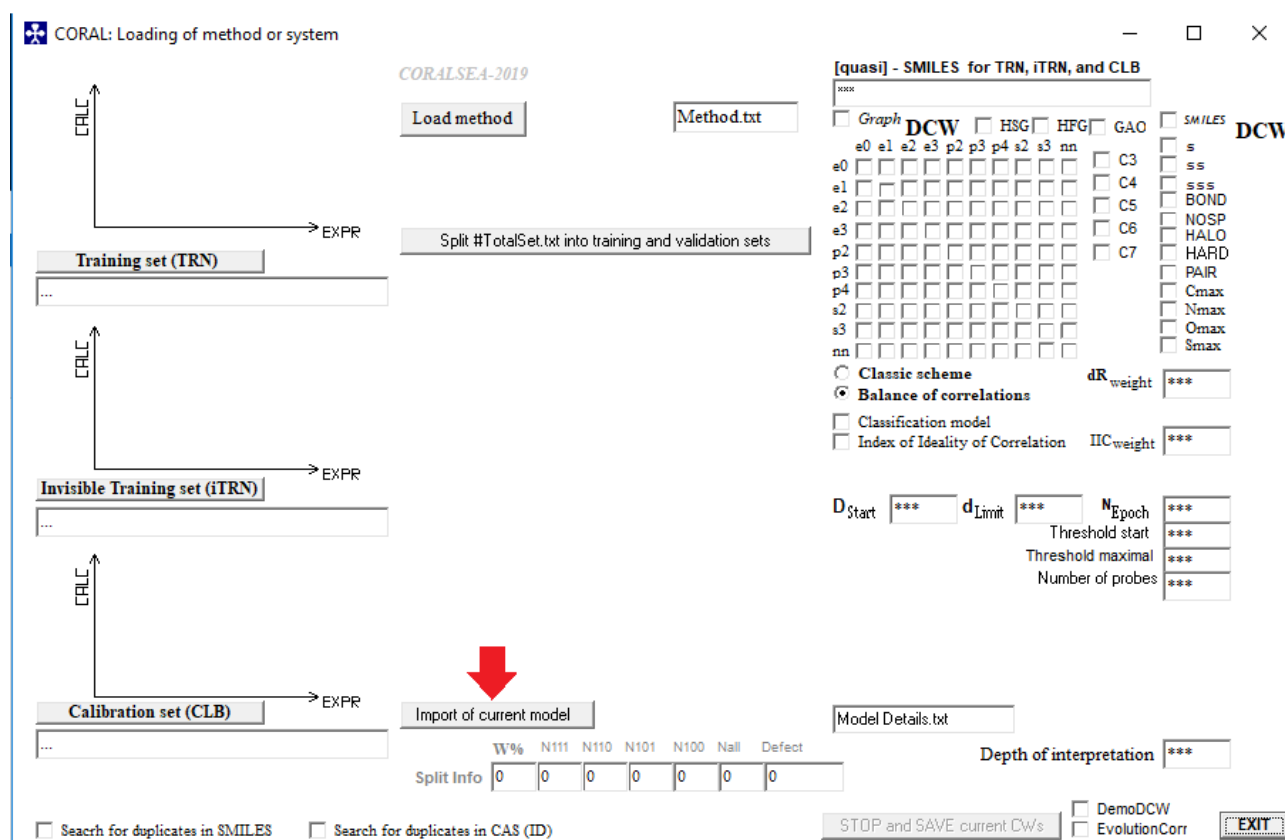


#### 4. Possibility to continue the Monte Carlo optimization

If you have saved your model (i.e. CWs), you can continue optimization process.  
First, run the CORALSEA.exe



Second, click button **Import of current model**





Continue optimization



**FYI:**

[andrey.toropov@marionegri.it](mailto:andrey.toropov@marionegri.it),  
[alla.toropova@marionegri.it](mailto:alla.toropova@marionegri.it),  
[aatoropov@yahoo.com](mailto:aatoropov@yahoo.com),  
[allatoropova@ymail.com](mailto:allatoropova@ymail.com)