Here special possibilities of 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.

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

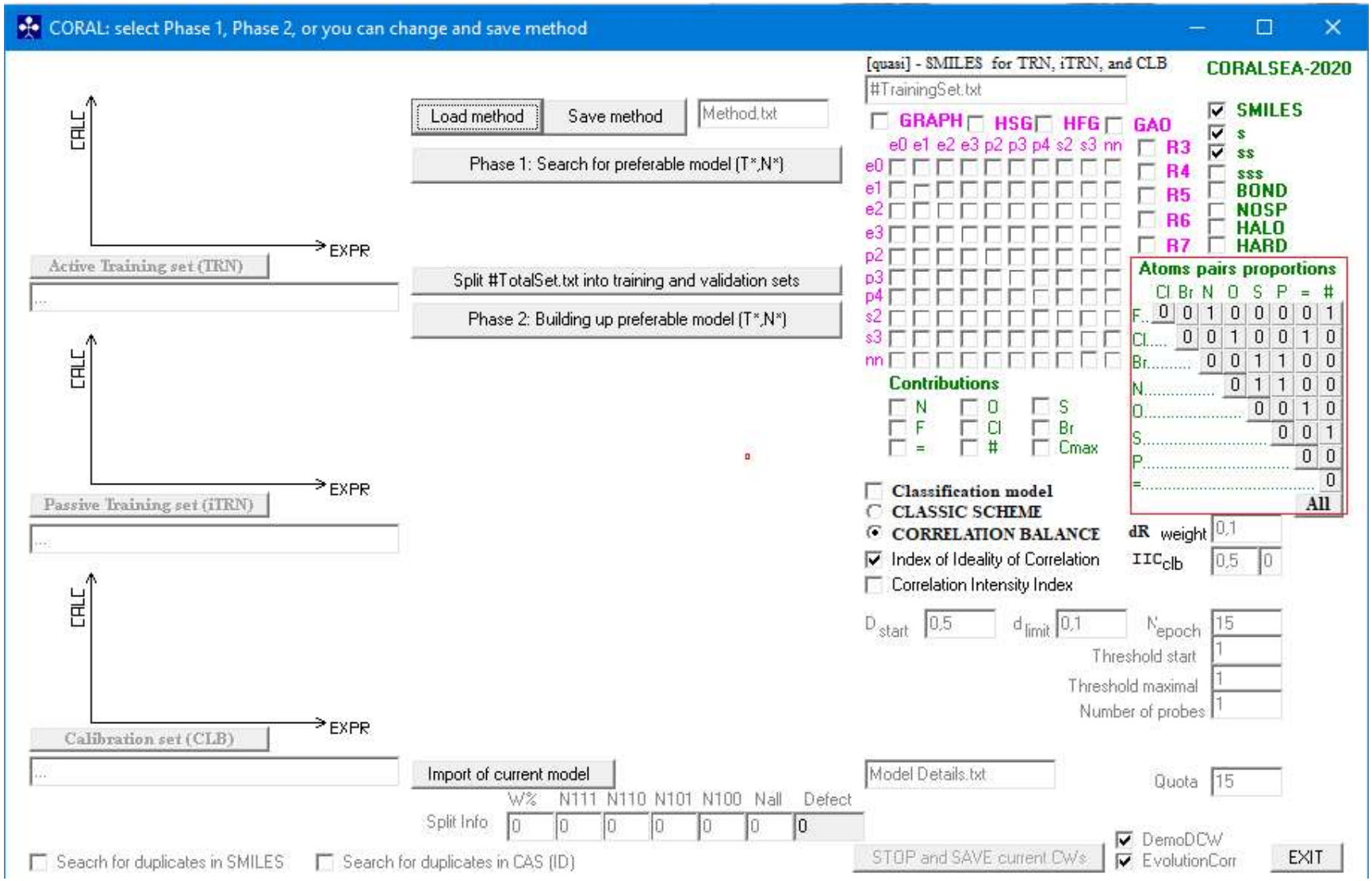
It is to be noted, that method.txt files for different versions are not interchangeable i.e. method.txt used for CORALSEA-2016 cannot be used for CORALSEA-2017, CORALSEA-19, and CORALSEA-2020; as well as method for CORALSEA-2017 cannot be applied for CORALSEA-2019 and CORALSEA-2020. The user should modify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

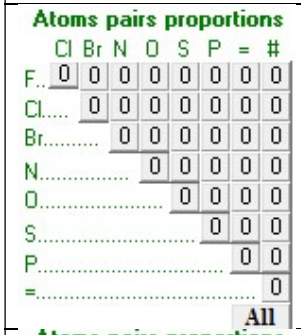
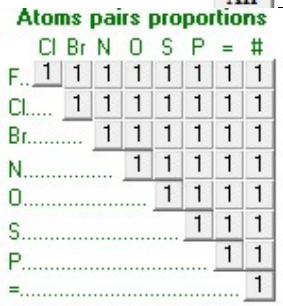
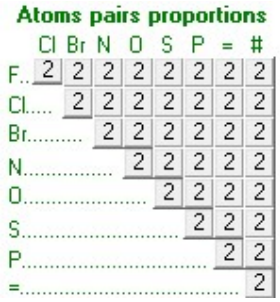
The example of the CORAL model represented in the folder CORALSEA-2020 related to data on anti-HIV activity from article:
Structural Chemistry (2020) 31,1441–1448 (for details please see pdf in the folder)

FYI:
You can send questions, suggestions, remarks using contact addresses:
andrey.toropov@marionegri.it,
alla.toropova@marionegri.it,
aatoropov@yahoo.com,
allatoropova@ymail.com

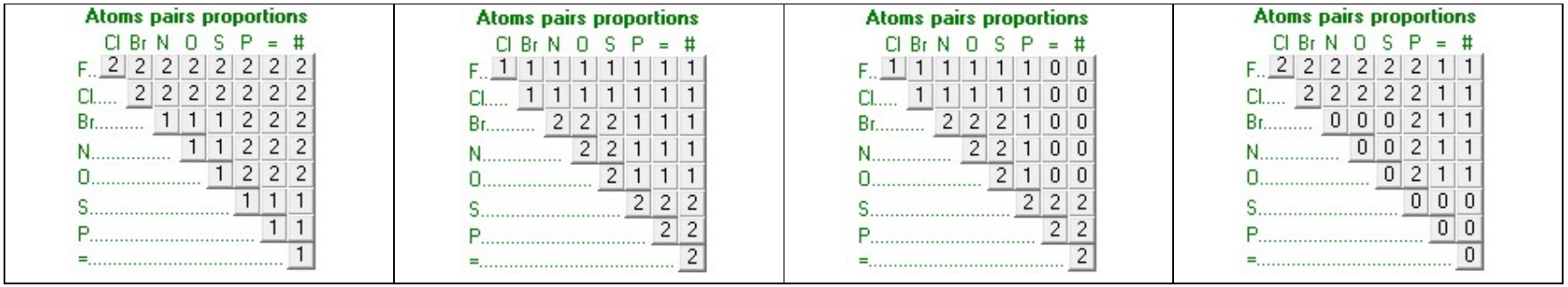
1. Self-organized atoms pairs proportions vector.

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



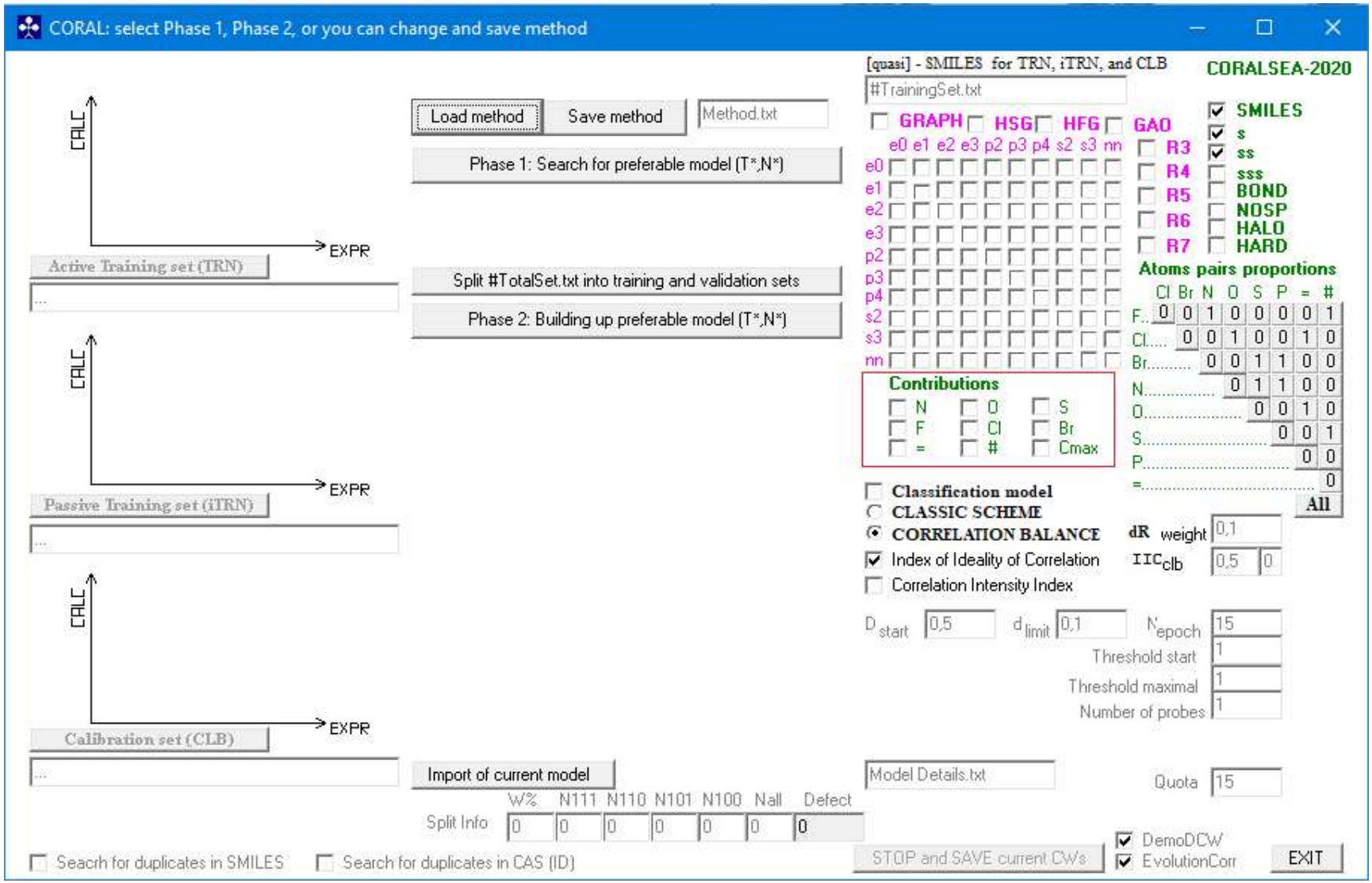
Status	Comment
	The button “all” gives possibility to switch status of all buttons from the square into the following status: 0, 1, and 2. Status 0 means that APP vector is not involved in the modeling process.
	Status 1 means that pair-combinations (in format Yes-Yes) of indicated eight chemical elements 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). 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
	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). 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

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

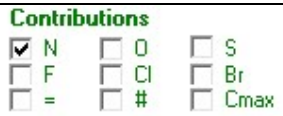
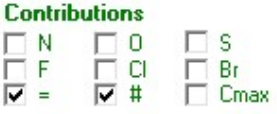


If the corresponding pair is absent or in other words, one of the chemical elements (as well as a double or triple bond) is absent the component of APP vector will not be involved for the modelling process. Thus, the self-organization of the vector is the selection of pairs or APP which indeed have a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

2. Molecular features contributions.
The molecular contributions indicated by red frame.



Some examples explain the above options are the following.

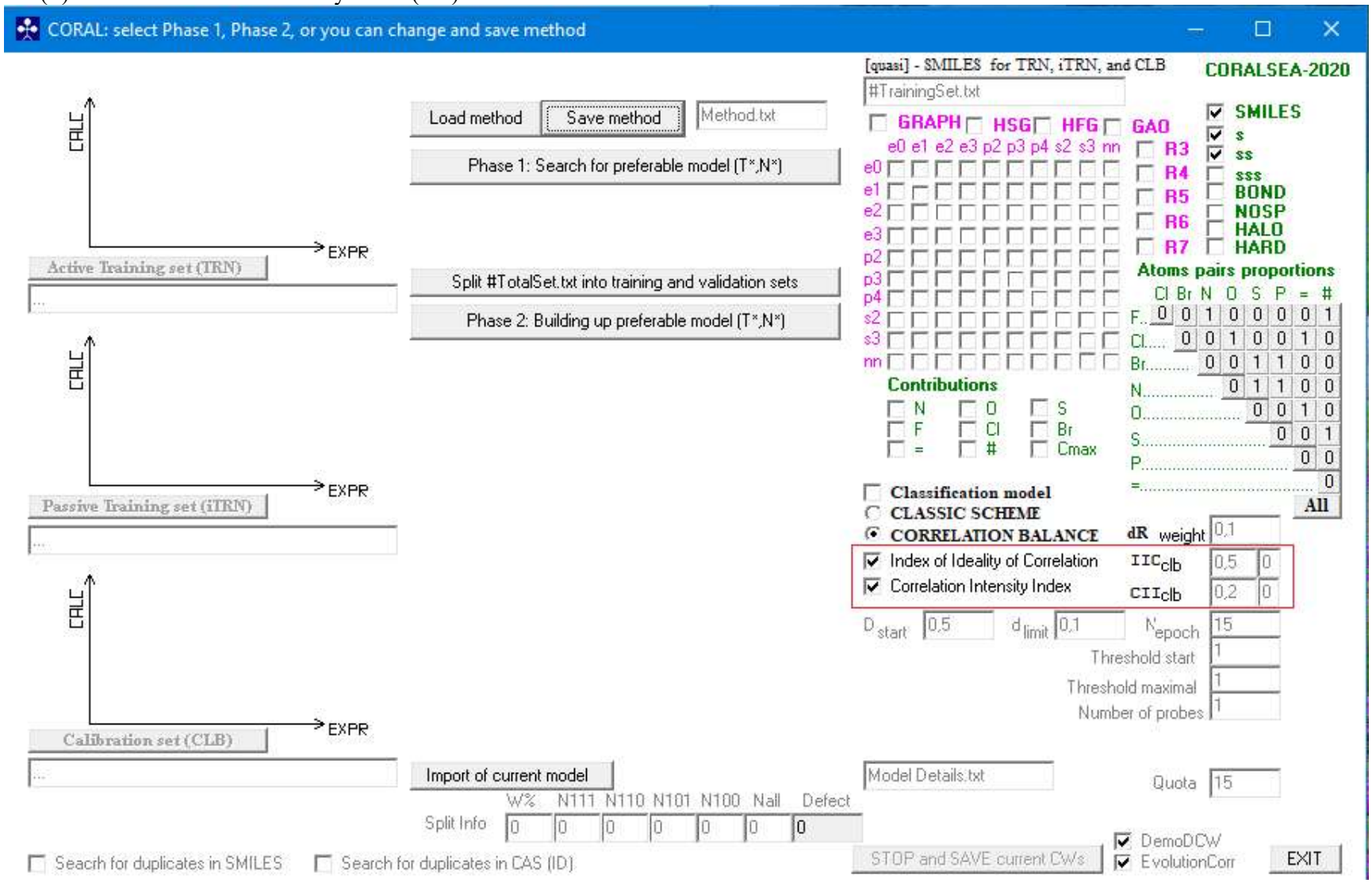
Status	Comment
	The number of chemical elements (N, O, S, F, Cl, Br); the number of double and triple bonds (=,#); and the number of cycles (rings) in the molecule (SMILES) are not involved for building up model.
	The total number of nitrogen atoms will be applied to develop model via correlation weights 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
	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
	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
	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.

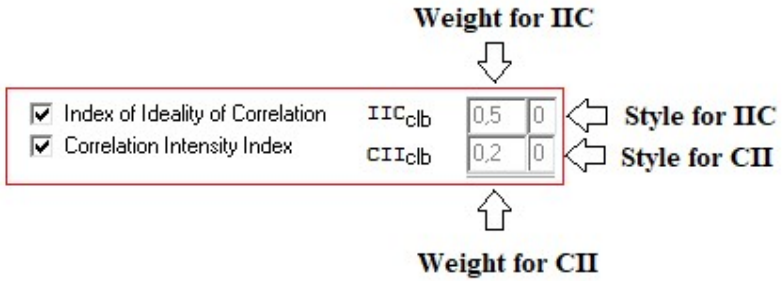
The new possibilities of the Monte Carlo optimization are based on two new criteria of predictive potential (Table 1).

(i) The index of ideality of correlation (IIC);

(ii) The correlation intensity index (CII).



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
☐ Index of Ideality of Correlation ☐ Correlation Intensity Index	The optimization without applying *IIC* and *CII*, i.e. target function is TF_1
☑ Index of Ideality of Correlation IIC_{clb} 0,5 0 ☐ Correlation Intensity Index	Optimization with TF_2, *WeightForIIC*=0.5
☑ Index of Ideality of Correlation IIC_{itrn} 0,5 1 ☐ Correlation Intensity Index	Optimization with TF_3, *WeightForIIC*=0.5
☑ Index of Ideality of Correlation $IIC_{itrn,clb}$ 0,5 2 ☐ Correlation Intensity Index	Optimization with TF_4, *WeightForIIC*=0.5
☐ Index of Ideality of Correlation ☑ Correlation Intensity Index CII_{clb} 0,2 0	Optimization with TF_5, *WeightForCII*=0.2
☐ Index of Ideality of Correlation ☑ Correlation Intensity Index CII_{itrn} 0,2 1	Optimization with TF_6, *WeightForCII*=0.2
☐ Index of Ideality of Correlation ☑ Correlation Intensity Index $CII_{itrn,clb}$ 0,2 2	Optimization with TF_7, *WeightForCII*=0.2
☑ Index of Ideality of Correlation IIC_{itrn} 0,3 1 ☑ Correlation Intensity Index $CII_{itrn,clb}$ 0,5 2	Optimization with target function $TF = TF_3+TF_7$; *WeightsForIIC*=0.3; and *WeightsForCII*=0.5.

Table 1

Statistical criteria of the predictive potential for QSPR/QSAR models

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

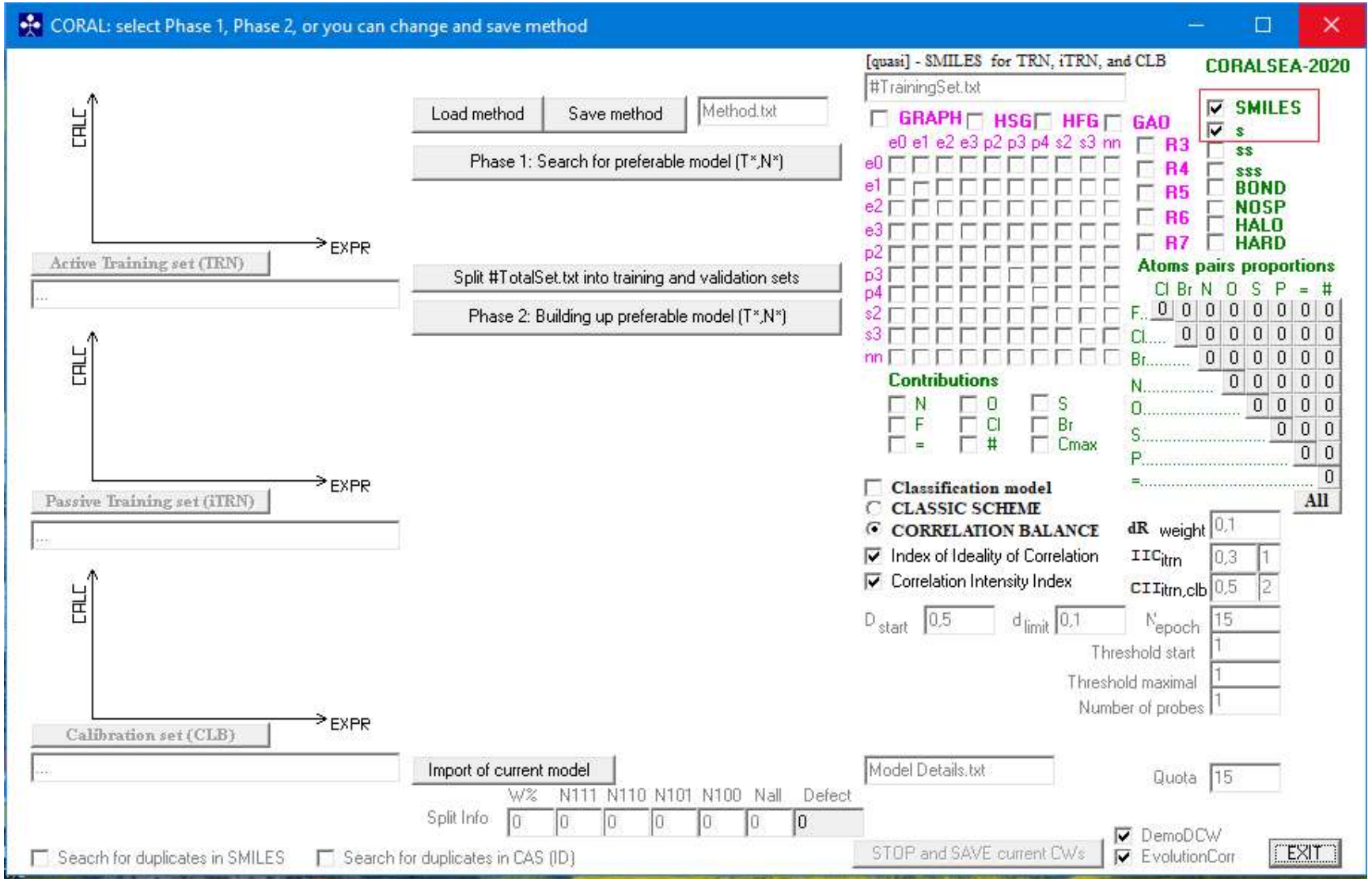
4. New options for design of the quasi-SMILES codes.

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

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

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
…
Size of symbols sequence between square brackets is 10. [123456789A], no more.
The only method with the SMILES-file similar to the above can be the following (indicated by red)



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)(c1ccccn1=O)Cc2ccccc2.[Dark][Low Conc] 3.840

*2 O=S(=O)(c1ccccn1=O)Cc2ccccc2C.[Light][High Conc] 4.060

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.compbiolchem.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