# Constants of chain transmission in the radical polymerization as a mathematical function of the molecular structure of monomers and regulators, which are presented by SMILES

Valentin O. Kudyshkin[1], Andrey A. Toropov[2,*], Sayyora Sh. Rashidova[1]

[1] *Institute of Polymer Chemistry and Physics, Academy of Sciences of the Republic of Uzbekistan, Kodyri street 7b, 100128, Tashkent, Uzbekistan*

[2] *Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milano, Italy*

[*)] Corresponding author
Andrey A. Toropov
Laboratory of Environmental Chemistry and Toxicology,
Department of Environmental Health Science,
Istituto di Ricerche Farmacologiche Mario Negri IRCCS
Via Mario Negri 2, 20156 Milano, Italy
Tel: +39 02 3901 4595
Fax: +39 02 3901 4735
Email: andrey.toropov@marionegri.it

**Abstract**

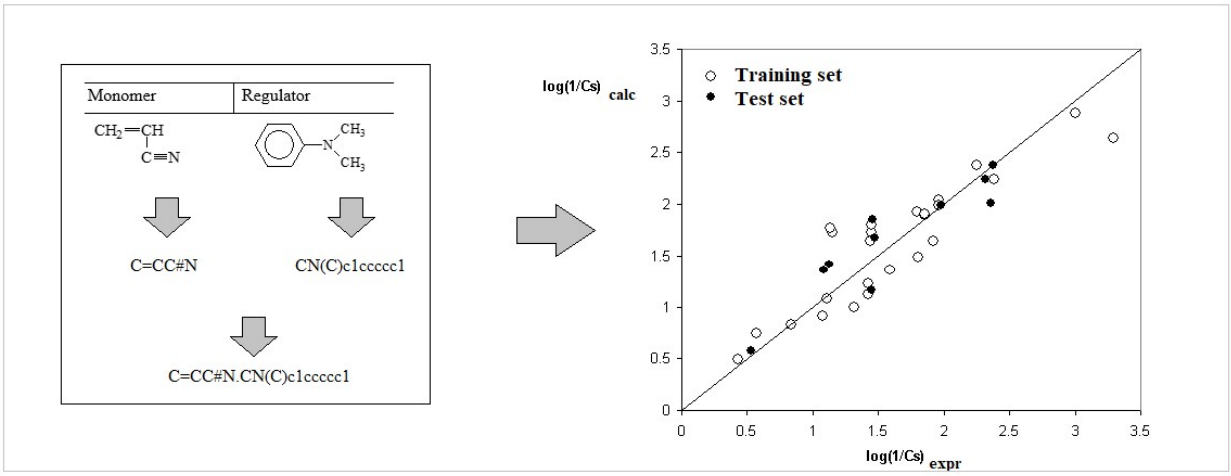
The applying of radical polymerization techniques is an important economic and theoretical task of chemical technology. Hence, the elaboration of the radical polymerization processes has both theoretical and practical significance. The ability of the CORAL software to build up robust models for transfer constants in radical polymerization is demonstrated. The Monte Carlo technique is applied in the CORAL software as the principle of development of the models. For molecular fragments represented by simplified molecular input-line entry system (SMILES), correlation weights were selected that maximize the correlation coefficient between the logarithm of the chain transfer constants in radical polymerization log (1/Cs) and the descriptor calculated from the correlation weights for the training set (n = 25). This correlation was preserved for the external validation set in three optimization samples by the Monte Carlo method.

**Keywords:** Mathematical modelling; radical polymerization; SMILES; monomer; regulator; chain transfer constant

**GRAPHICAL ABSTRACT**

**Constants of chain transmission , log(1/Cs) = F ( monomer SMILES, regulator SMILES )**

**log(1/Cs) = F ( C=CC#N.CN(C)c1ccccc1 )**



## Introduction

The radical polymerization techniques are used to develop block copolymers, for optimization technological processes, as well as for the improvement of target transport of drugs in an organism [1].

Quantitative structure-property/activity relationships (QSPRs/QSARs) is a tool to predict the physicochemical behaviour of different substances including polymer systems [2,3]. The CORAL software (http://ww.insilico.eu/coral) is a tool to build up different QSPR/QSAR models [4-6].

It is known that the properties of polymers are determined not only by the chemical nature of the monomer units, but also by molecular weight characteristics, and often the optimal properties are manifested in a certain range of molecular weights (MM). Therefore, the important problem is the good regulation of the molecular characteristics of polymers in the synthesis process, which, within the framework of classical radical polymerization, can be carried out using chain transfer reactions to a specially introduced agent - regulator. Quantitatively, the activity of chain transfer agents is described by the chain transfer constant (Cs), which depends on the chemical structure of the monomer and the chain transfer agent. The determination of the chain transfer constant requires a sufficiently long experiment. In this regard, it is of interest to use new theoretical approaches to model the chain transfer constants based on information on the chemical structure of the reagents. In this regard, the use of the OSPR method-modeling for predicting the kinetic constants of polymerization based on information on the chemical structure of monomers and regulators presents a certain prospect [1, 2]. Earlier, we attempted to model the chain transfer constant using the optimization of the correlation weights of local graph invariants [2]. Computer experiments were conducted for 19 and 7 pairs of monomer - regulator for the training and control samples, respectively.

The aim of the present study is the estimation of the ability of optimal descriptors [2,4-6] calculated with simplified molecular input-line entry system (SMILES) [7-9] to be a tool for prediction of log(1/Cs).

## 2. Method

### *2.1 Data*

Numerical data on Cs take in the literature [10, 11]. The 35 systems of “monomer-regulator” were taken from the above sources [10, 11]. These systems were randomly distributed to the training set (n=25) and test set (n=10).

### *2.2 Optimal descriptor*

The optimal descriptors is a class of the molecular descriptors calculated with correlation weights of various molecular features extracted from SMILES. The optimal descriptor used here is calculated as the following

$$DCW = \sum W(S_k) \quad (1)$$

The $S_k$ is a fragment of SMILES. The $W(S_k)$ is the correlation weight for the $S_k$. The numerical data for correlation weights are calculated by the Monte Carlo technique. These should provide a maximal value of the determination coefficient between the experimental and predicted log(1/Cs) for the training set:

$$\log(1/Cs) = C_0 + C_1 * DCW \quad (2)$$

The statistical quality of the model calculated with Eq. 2 for the training set should be checked up with monomers from the test set.

Monomers and regulators are listed in Table 1. SMILES are prepared by ChemSketch software [12].

**3. Results and Discussion**

Table 2 contains the statistical characteristics of the model for three runs of the Monte Carlo optimization. One can see, the statistical quality of the model is good and reproducing for the above three runs. Table 3 contains numerical data on correlation weights used to calculations with Eq. 1. Table 4 contains an example of the *DCW* calculation with correlation weights obtained in ~~the~~ run 1.

The model is the following:

$$\log(1/Cs) = 0.8489 + 0.1178 * DCW \quad (3)$$

n=25, r2=0.8094, s=0.290, F=98 (training set)
n=10, r2=0.8509, s=0.250, F=46 (test set)

Table 5 contains the numerical data on the model. Figure 1 represents the model graphically. The predictive potential of the model confirmed by the high value of the concordance correlation coefficient that is equal to 0.895 [13] as well as by the high value of the index of ideality of correlation that is equal to 0.829 [14].

Previous researches dedicated to building up QSPR suitable for polymer objects were based on the representation of monomers via molecular graphs [15] whereas here SMILES has been

applied as an alternative of the graph. Taking into account the prevalence SMILES as a representation of the molecular structure, one can see the approach described here has defined advantages.

## 4. Conclusions

The optimal descriptors calculated according to the described scheme can be used to model the log (1 / Cs) chain transfer constants in radical copolymerization. The practicability of representation of the topological structure of monomers together with the topological structure of regulators in order to build up predictive models for polymer objects is confirmed. Last, but not least, SMILES can be an attractive alternative to a graph in the QSPR analysis of polymer objects.

**Acknowledgment**

AAT expresses gratitude the project LIFE-CONCERT contract (LIFE17 GIE/IT/000461) for the support.

**Disclosure statement**

The authors reported no potential conflict of interest.

**References**

[1] S. Oliver, L. Zhao, A.J. Gormley, R. Chapman, C. Boyer, Living in the Fast Lane - High Throughput Controlled/Living Radical Polymerization, Macromolecules, 52(1) (2019) 3-23. DOI: 10.1021/acs.macromol.8b01864

[2] A.P. Toropova, A.A. Toropov, V.O. Kudyshkin, R. Rallo, Prediction of the Q-e parameters from structures of transfer chain agents, J. Polym. Res. 22 (7) (2015) 128. DOI: 10.1007/s10965-015-0778-3

[3] T. Zhu, W. Chen, H. Cheng, Y. Wang, R.P. Singh, Prediction of polydimethylsiloxane-water partition coefficients based on the pp-LFER and QSAR models, Ecotoxicol. Environ. Saf. 182 (2019) 109374. DOI: 10.1016/j.ecoenv.2019.109374

[4] S. Ahmadi, A. Akbari, Prediction of the adsorption coefficients of some aromatic compounds on multi-wall carbon nanotubes by the Monte Carlo method, SAR QSAR Environ. Res. 29(11) (2018) 895-909. DOI: 10.1080/1062936X.2018.1526821

[5] Manisha, S. Chauhan, P. Kumar, A. Kumar, Development of prediction model for fructose-1,6- bisphosphatase inhibitors using the Monte Carlo method, SAR QSAR Environ. Res. 30(3) (2019) 145-159. DOI: 10.1080/1062936X.2019.1568299

[6] P.G.R. Achary, A.P. Toropova, A.A. Toropov, Combinations of graph invariants and attributes of simplified molecular input-line entry system (SMILES) to build up models for sweetness, Food Res. Int. 122 (2019) 40-46. https://doi.org/10.1016/j.foodres.2019.03.067

[7] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31-36. https://doi.org/10.1021/ci00057a005

[8] D. Weininger, A. Weininger, J.L. Weininger, SMILES. 2. Algorithm for generation of unique SMILES notation, J. Chem. Inf. Comput. Sci. 29 (1989) 97-101. https://doi.org/10.1021/ci00062a008

[9] D. Weininger, Smiles. 3. Depict. Graphical depiction of chemical structures, J. Chem. Inf. Comput. Sci. 30 (1990) 237-243. https://doi.org/10.1021/ci00067a005

[10] Reference book on polymer chemistry (in Russian) https://elib.pstu.ru/vufind/Record/RUPSTUbooks177494, Accessed October 1, 2019.

[11] V.O. Kudyshkin, N.R. Vokhidova, N.I. Bozorov, O.E. Sidorenko, I.N. Ruban, N.L. Voropaeva, S.Sh. Rashidova, Use of allylbenzene and allyl phenyl ether as chain-transfer agents in radical polymerization, Russ. J. Appl. Chem. 77 (2004) 994-997. https://doi.org/10.1023/B:RJAC.0000044130.01483.f7

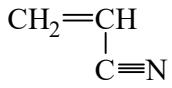
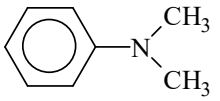
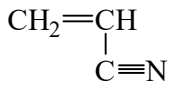
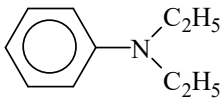
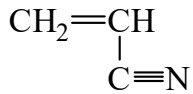
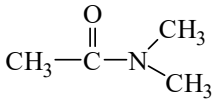
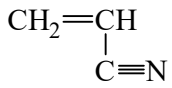
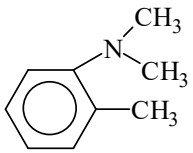
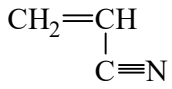
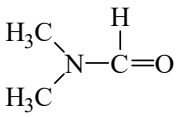
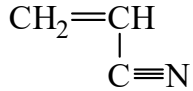
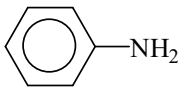
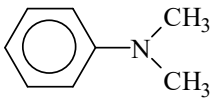
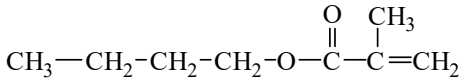
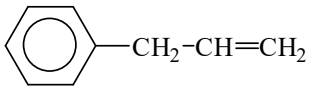
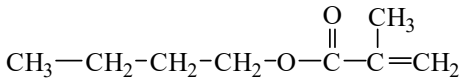
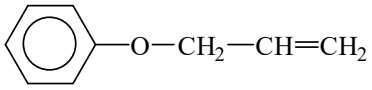
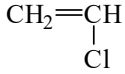
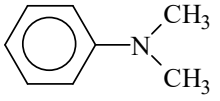
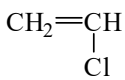
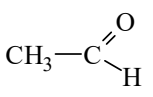
[12] ACD/ChemSketch Freeware, version 11.00 - Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2007.

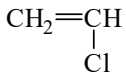
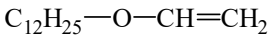
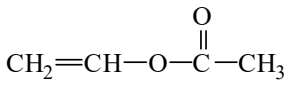
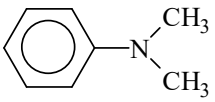
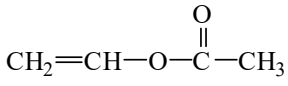
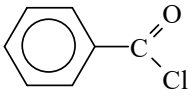
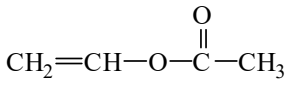
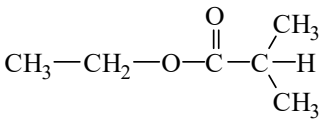
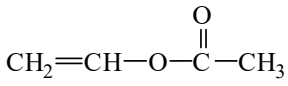
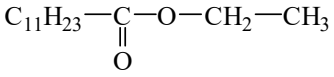
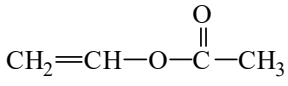
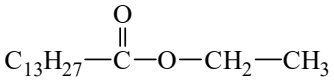
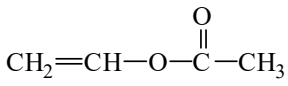
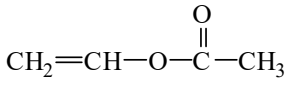
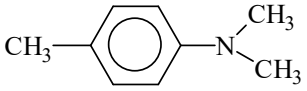
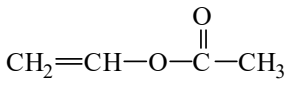
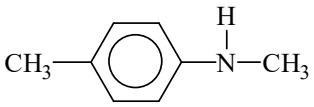
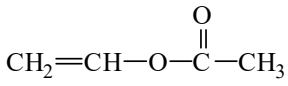
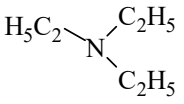
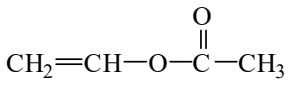
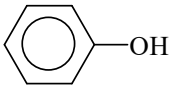
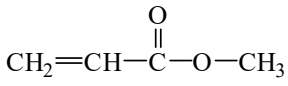
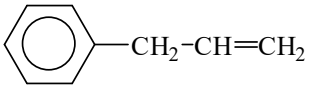
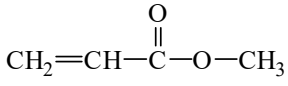
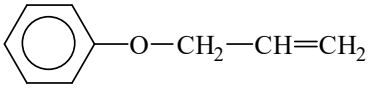
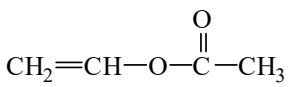
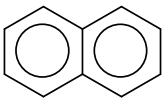
[13] L. I-Kuei Lin, A concordance correlation coefficient to evaluate reproducibility, Biometrics 45 (1989) 255-268. https://www.jstor.org/stable/2532051

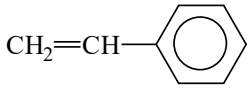
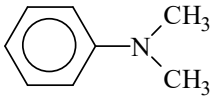
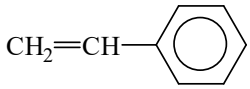
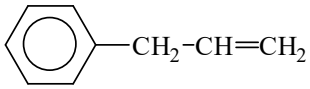
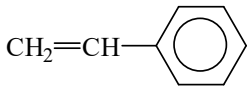
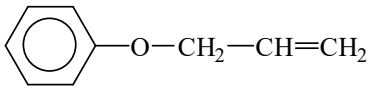
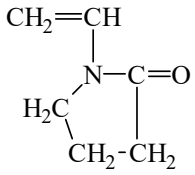
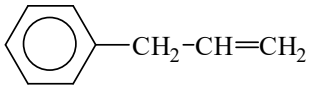
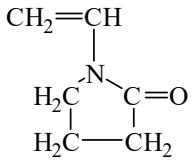
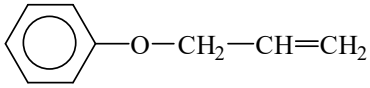
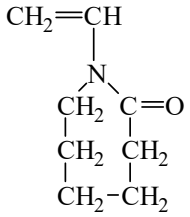
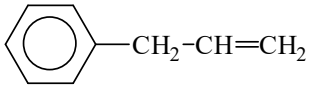
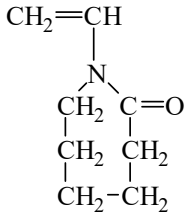
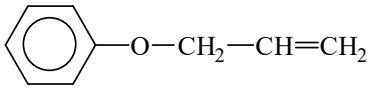
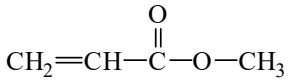
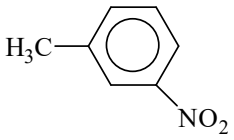
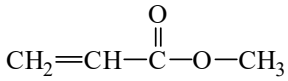
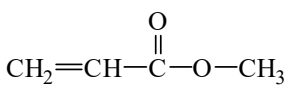
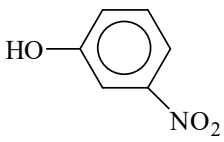
[14] A.A.Toropov, A.P. Toropova, The index of ideality of correlation: A criterion of predictive potential of QSPR/QSAR models? Mutat. Res. Genet. Toxicol. Environ. Mutagen. 819 (2017) 31-37. DOI: 10.1016/j.mrgentox.2017.05.008

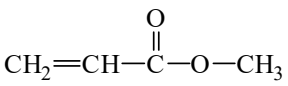
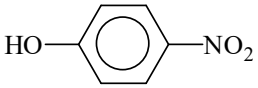
[15] A.A. Toropov, V.O. Kudyshkin, N.L. Voropaeva, I.N. Ruban, S.Sh. Rashidova, QSPR modeling of the reactivity parameters of monomers in radical copolymerizations, J. Struct. Chem. 45 (2004) 945-950. https://doi.org/10.1007/s10947-005-0084-8

Table 1 Numerical data on Cs taken in the literature [10,11]

| ID | Monomer | Regulator | Cs |
|---|---|---|---|
| 1 |  |  | 0.0708 |
| 2 |  |  | 0.0359 |
| 3 |  |  | 0.000505 |
| 4 |  |  | 0.0334 |
| 5 |  |  | 0.001 |
| 6 |  |  | 0.0044 |
| 7 |  |  | 0.038 |
| 8 |  |  | 0.073 |
| 9 |  |  | 0.035 |
| 10 |  |  | 0.27 |
| 11 |  |  | 0.011 |

| | | | |
|---|---|---|---|
| 12 |  |  | 0.0156 |
| 13 |  |  | 0.026 |
| 14 |  |  | 0.0366 |
| 15 |  |  | 0.016 |
| 16 |  |  | 0.0105 |
| 17 |  |  | 0.014 |
| 18 |  |  | 0.075 |
| 19 |  |  | 0.038 |
| 20 |  |  | 0.083 |
| 21 |  |  | 0.036 |
| 22 |  |  | 0.012 |
| 23 |  |  | 0.014 |
| 24 |  |  | 0.011 |
| 25 |  |  | 0.1457 |

| | | | |
|---|---|---|---|
| 26* |  |  | 0.0053 |
| 27 |  |  | 0.084 |
| 28 |  |  | 0.048 |
| 29 |  |  | 0.078 |
| 30 |  |  | 0.036 |
| 31 |  |  | 0.37 |
| 32 |  |  | 0.295 |
| 33 |  |  | 0.00412 |
| 34 |  |  | 0.00486 |
| 35 |  |  | 0.00562 |

| | | | |
|---|---|---|---|
| 36 |  |  | 0.00426 |

*) outlier

Table 2 The statistical quality of models calculated with Eq. 2

| | Training set, n=25 | | | Test set, n=10 | | |
|---|---|---|---|---|---|---|
| **Run** | $r^2$ | *s* | *F* | $r^2$ | *s* | *F* |
| **1** | 0.8094 | 0.290 | 98 | 0.8509 | 0.250 | 46 |
| **2** | 0.8064 | 0.292 | 96 | 0.8560 | 0.246 | 48 |
| **3** | 0.8103 | 0.290 | 98 | 0.8484 | 0.252 | 45 |
| **Average** | 0.8087 | 0.291 | 97 | 0.8518 | 0.249 | 46 |

Table 3 Correlation weights for three runs of the Monte Carlo optimization

| Promoters for log(1/Cs) | $S_k$ | $W(S_k)$ run 1 | $W(S_k)$ run 2 | $W(S_k)$ run 3 |
|---|---|---|---|---|
| increase | | | | |
| | # | 5.8018 | 6.6609 | 5.2256 |
| | + | 0.3449 | 1.1372 | 1.3606 |
| | - | 0.3911 | 0.9293 | 0.7507 |
| | . | 4.7479 | 2.9346 | 2.7729 |
| | = | 4.1671 | 4.6296 | 3.9259 |
| | N | 0.8395 | 0.9611 | 0.9045 |
| | O | 0.7282 | 0.8522 | 0.7168 |
| | [ | 0.6045 | 0.3190 | 0.2137 |
| | c | 0.4581 | 0.5024 | 0.4555 |
| decrease | | | | |
| | ( | -0.8366 | -0.8778 | -0.8215 |
| | 1 | -4.0206 | -4.2513 | -3.9007 |
| | 2 | -3.9925 | -4.1779 | -3.8288 |
| | Cl | -2.0521 | -1.7040 | -2.0528 |
| | C | -0.3957 | -0.4225 | -0.3778 |

Table 4 An example of the *DCW* calculation (run 1)

SMILES = "C=CC#N.CN(C)c1ccccc1"

*DCW*=7.4515977

| $S_k$ | $W(S_k)$ |
|---|---|
| C | -0.3957 |
| = | 4.1671 |
| C | -0.3957 |
| C | -0.3957 |
| # | 5.8018 |
| N | 0.8395 |
| . | 4.7479 |
| C | -0.3957 |
| N | 0.8395 |
| ( | -0.8366 |
| C | -0.3958 |
| ( | -0.8366 |
| c | 0.4581 |
| 1 | -4.0206 |
| c | 0.4581 |
| c | 0.4581 |
| c | 0.4581 |
| c | 0.4581 |
| c | 0.4581 |
| 1 | 4.0206 |

Table 5 Experimental and calculated with Eq. 3 log(1/Cs)

| ID | SMILES | *DCW* | log(1/Cs) experiment | log(1/Cs) predicted |
|---|---|---|---|---|
| **Training set** | | | | |
| 1 | C=CC#N.CN(C)c1ccccc1 | 7.4516 | 1.150 | 1.727 |
| 2 | C=CC#N.CN(C)c1ccccc1 | 7.4516 | 1.445 | 1.727 |
| 3 | C=CC#N.CN(C)C(C)=O | 15.1746 | 3.297 | 2.636 |
| 5 | C=CC#N.CN(C)C=O | 17.2435 | 3.000 | 2.880 |
| 7 | C=CC(=O)OCCCC.CN(C)c1ccccc1 | 3.1777 | 1.420 | 1.223 |
| 8 | C=CC(=O)OCCCC.C=CCc1ccccc1 | 7.7828 | 1.137 | 1.766 |
| 10 | C=CCl.CN(C)c1ccccc1 | -0.8460 | 0.569 | 0.749 |
| 11 | C=CCl.CC=O | 10.1753 | 1.959 | 2.048 |
| 12 | C=CCl.C=COCCCCCCCCCCCC | 5.4266 | 1.807 | 1.488 |
| 13 | CC(=O)OC=C.CN(C)c1ccccc1 | 4.3649 | 1.585 | 1.363 |
| 14 | CC(=O)OC=C.ClC(=O)c1ccccc1 | 6.7643 | 1.437 | 1.646 |
| 15 | CC(=O)OC=C.CC(C)(C)C(=O)OCC | 9.1163 | 1.796 | 1.923 |
| 17 | CC(=O)OC=C.O=C(CCCCCCCCCCCCC)OCC | 8.9011 | 1.854 | 1.897 |
| 19 | CC(=O)OC=C.Cc1ccc(cc1)N(C)C | 2.2960 | 1.420 | 1.119 |
| 21 | CC(=O)OC=C.CCN(CC)CC | 8.0744 | 1.444 | 1.800 |
| 22 | CC(=O)OC=C.Oc1ccccc1 | 6.7182 | 1.921 | 1.640 |
| 23 | C=CC(=O)OC.C=CCc1ccccc1 | 8.9699 | 1.854 | 1.906 |
| 24 | C=CC(=O)OC.C=CCOc1ccccc1 | 9.6981 | 1.959 | 1.991 |
| 25 | CC(=O)OC=C.c1cccc2ccccc12 | -0.1624 | 0.837 | 0.830 |
| 27 | C=Cc1ccccc1.C=CCc1ccccc1 | 0.5187 | 1.076 | 0.910 |
| 28 | C=Cc1ccccc1.C=CCOc1ccccc1 | 1.2469 | 1.319 | 0.996 |
| 29 | O=C1CCCN1C=C.C=CCc1ccccc1 | 1.9217 | 1.108 | 1.075 |
| 31 | CCN1CCCCCC1=O.C=CCc1ccccc1 | -3.0368 | 0.432 | 0.491 |
| 33 | C=CC(=O)OC.Cc1cc(ccc1)[N+](=O)[O-] | 11.8650 | 2.385 | 2.247 |
| 35 | C=CC(=O)OC.O=[N+]([O-])c1cccc(O)c1 | 12.9889 | 2.250 | 2.379 |
| **Test set** | | | | |
| 4 | C=CC#N.Cc1ccccc1N(C)C | 7.0559 | 1.476 | 1.680 |

| | | | | |
|---|---|---|---|---|
| 6 | C=CC#N.Nc1ccccc1 | 9.9162 | 2.357 | 2.017 |
| 9 | C=CC(=O)OCCCC.C=CCOc1ccccc1 | 8.5110 | 1.456 | 1.851 |
| 16 | CC(=O)OC=C.O=C(CCCCCCCCCCCC)OCC | 9.6926 | 1.979 | 1.991 |
| 18 | CC(=O)OC=C.Cc1ccc(N)cc1 | 4.7606 | 1.125 | 1.410 |
| 20 | CC(=O)OC=C.Cc1ccc(NC)cc1 | 4.3649 | 1.081 | 1.363 |
| 30 | O=C1CCCN1C=C.C=CCOc1ccccc1 | 2.6499 | 1.444 | 1.161 |
| 32 | CCN1CCCCCC1=O.C=CCOc1ccccc1 | -2.3086 | 0.530 | 0.577 |
| 34 | C=CC(=O)OC.Cc1ccc(cc1)[N+](=O)[O-] | 11.8650 | 2.313 | 2.247 |
| 36 | C=CC(=O)OC.O=[N+]([O-])c1ccc(O)cc1 | 12.9889 | 2.371 | 2.379 |

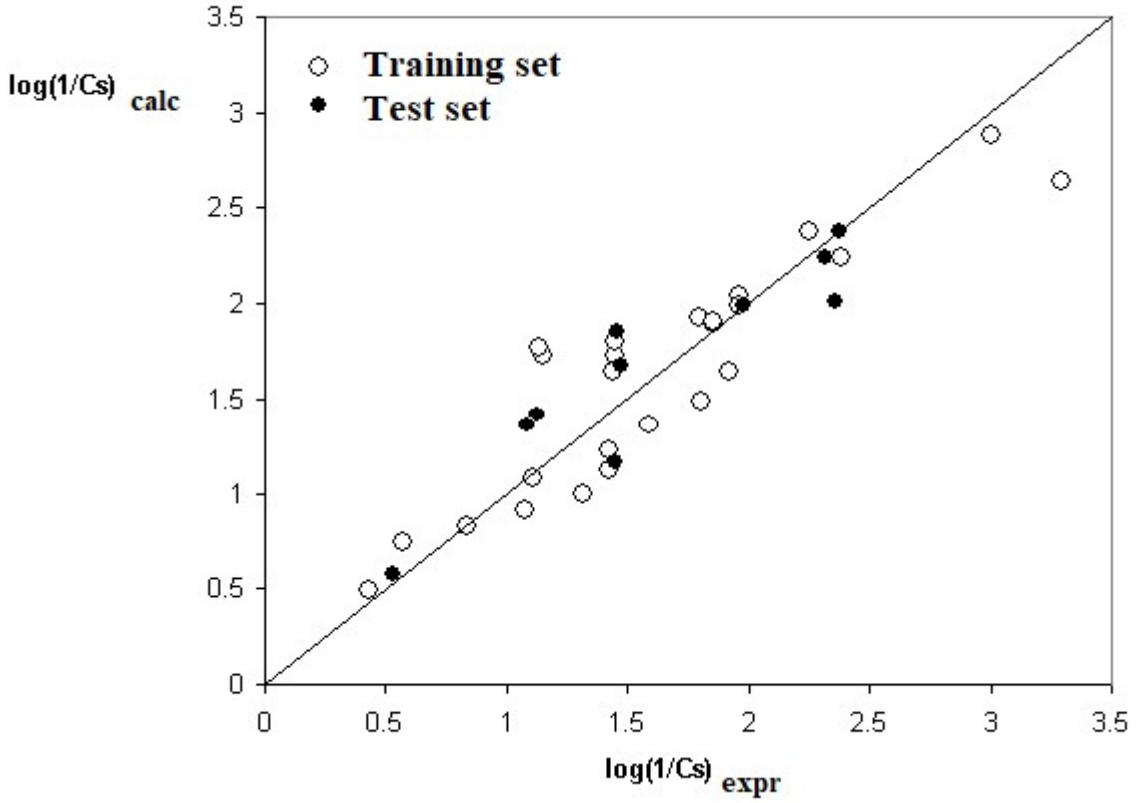



Figure 1 Graphical representation of model calculated with Eq. 3