Saudi Journal of Biological Sciences 26 (2019) 1101–1106

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

CORAL: Building up QSAR models for the chromosome aberration test

Andrey A. Toropov, Alla P. Toropova *, Giuseppa Raitano, Emilio Benfenati

Department of Environmental Health Science, Laboratory of Environmental Chemistry and Toxicology, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milano, Italy

ARTICLE INFO

Article history:
Received 3 November 2017
Revised 23 April 2018
Accepted 6 May 2018
Available online 9 May 2018

Keywords:
QSAR
Chromosome aberration
SMILES
Monte Carlo method
CORAL software

ABSTRACT

A high level of chromosomal aberrations in peripheral blood lymphocytes may be an early marker of cancer risk, but data on risk of specific cancers and types of chromosomal aberrations are limited. Consequently, the development of predictive models for chromosomal aberrations test is important task. Majority of models for chromosomal aberrations test are so-called knowledge-based rules system. The CORAL software (http://www.insilico.eu/coral, abbreviation of "CORrelation And Logic") is an alternative for knowledge-based rules system. In contrast to knowledge-based rules system, the CORAL software gives possibility to estimate the influence upon the predictive potential of a model of different molecular alerts as well as different splits into the training set and validation set. This possibility is not available for the approaches based on the knowledge-based rules system. Quantitative Structure–Activity Relationships (QSAR) for chromosome aberration test are established for five random splits into the training, calibration, and validation sets. The QSAR approach is based on representation of the molecular structure by simplified molecular input-line entry system (SMILES) without data on physicochemical and/or biochemical parameters. In spite of this limitation, the statistical quality of these models is quite good.

1. Introduction

There are large diversity of biochemical endpoints which should be available for development of medicinal biochemistry at least via computational models (Tenorio-Borroto et al., 2014; González-Díaz et al.,2013a,b; Prado-Prado et al., 2013; Duardo-Sanchez and Gonzalez-Diaz, 2013; Tenorio-Borroto et al., 2012; Riera-Fernández et al., 2012; González-Díaz et al., 2007). Mutagenicity and carcinogenicity are interrelated factors which can catastrophically impact human health (Toropova and Toropov, 2014). The necessity to assess risk of applying of various substances in the above aspect is vital necessity (Gollapudi et al., 2013).

There are increase of the number of publications (2012–2017) dedicated to chromosome aberration assay according to PubMed. Importance of systematization of available data and definition of effective strategy for diagnostics and treatment of different cases of breast cancer accompanied by the chromosome aberration is noted by many authors (Grade et al., 2015; Ben-David et al., 2016; Hosein et al., 2010; Rennstam et al., 2003; Watters et al., 2003; Vulto-van Silfhout et al., 2013; Brookmire et al., 2013; Sun et al., 2015; Castro et al., 2006; Boffetta et al., 2007). The European REACH legislation (Registration, Evaluation, Authorization and Restriction of Chemicals) encourages to improve the safety of chemical substances, increase the research efforts and promote scientific innovation, including the use of alternative approaches to evaluate substances support (REACH, 2006). Among the in vitro tests required to identify mutagenic compounds, bacterial reverse mutation assay (Ames test) (Ames, 1979) and chromosome aberration test are frequently used in the first stages of the assessment for mutagenicity.

It is to be noted that in spite of high influence of REACH legislation there are negative tendencies caused by REACH: (1) The registration process is very expensive, due to the high degree of experimental and administrative work required; and (2) At social level, REACH raises the ethical problem caused by the huge amount of animal testing necessary to meet the requirements of REACH Gozalbes and Vicente de Julián-Ortiz, 2018.

More than 25 years ago, the OECD recognized the need to protect animals in general and, in particular, those used in experimental work. The progress in OECD on the harmonization of chemicals control, especially the agreement on Mutual Acceptance of Data (MAD), has greatly contributed to reduce the number of animals used in testing by avoiding duplicative testing.

* Corresponding author at: Laboratory of Environmental Chemistry and Toxicology, IRCCS – Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milano, Italy.
E-mail address: alla.toropova@marionegri.it (A.P. Toropova).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.sjbs.2018.05.013
1319-562X/

All OECD Test Guidelines (TGs) are available at the OECD website (http://www.oecd.org/env/ehs/testing/oecdguidelinesforthetestingofchemicals.htm).

In the first case, the genotoxic potential of a target compound is determined by the detection of the renewed functional capability to synthesize the essential amino acid of an auxotrophic histidine-dependent strain of *S. typhimurium*. At the presence of that mutagen, the revertant bacteria can grow up on a medium without histidine (OECD, 2008a). In vitro chromosome aberration assay is used to identify agents that cause structural aberrations in mammalian cells. As for the Ames test, the target compounds are examined with and without metabolizing system since often the interaction with genetic material occurs after metabolic activation. After incubation with the chemical target at intended intervals, the cells are arrested in metaphase and analyzed microscopically looking for chromosomal aberrations.

Many human genetic diseases are caused by chromosome mutations and there is evidence that they are also involved in the alterations of oncogenes and tumor suppressor genes of somatic cells in humans and experimental animals (OECD, 2008b). Chromosomal aberrations in peripheral blood lymphocytes have been used for decades for the surveillance of healthy individuals exposed to known or potential mutagens and carcinogens (Boffetta et al., 2007; Carrano and Natarajan, 1988). In addition, chromosome aberrations are typical features of neoplastic cells, and for certain cancers specific chromosome abnormalities are commonly present (Yunis, 1983).

Although specific chromosome aberrations detected in neoplasms are generated during carcinogenesis, it has been hypothesized that the frequency of chromosomal aberrations represents a marker of susceptibility to cancer, based on the concept that genetic damage in peripheral blood lymphocytes reflects similar damage in different target cells undergoing carcinogenesis (Carrano and Natarajan, 1988; Umbuzeiro et al., 2016). Moreover, the chromosome aberration test is an important parameter of a substance also from the point of view of drug discovery (Nigam, 2009), cosmetics, and food industry (https://www.fda.gov/downloads/Drugs/Guidances/ucm074931.pdf).

Due to their publicly and high quality availability, Ames test data have been used to develop several QSAR models that, during the last years, showed good performance predicting mutagenic activity (Claxton et al., 2010). In the case of the chromosomal aberration endpoint, the predictive models are few. This is probably due to the complexity of mechanism of its induction and the lower availability of high-quality experimental data. In addition, there are different models of the chromosome aberration test which involve topological indices together with physicochemical and biochemical parameters to build up a model (Votano, 2005; Jacobson-Kram and Contrera, 2007; Serra et al., 2003; Mohr et al., 2010; Rosenkranz, 2004; Rothfuss et al., 2006; Estrada and Molina, 2006).

However, often, the involving of physicochemical and biochemical parameters is unavailable. Consequently, the using solely molecular structures without additional data is an attractive alternative for building up a model of chromosome aberration test. The CORAL (CORrealtion And Logic) software allows building up models of this kind. The aim of this study is the estimation of models for chromosome aberration test which are built up using the CORAL software (Toropova and Toropov, 2014).

2. Method

2.1. Data

Experimental data for this work were taken from the Genotoxicity OASIS Database (http://oasis-lmc.org/products/databases/rat-liver-metabolism-extended.aspx) and the Toxicity Japan MHLW (http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp) that include data for chromosomal aberrations determined by in vitro test using Chinese hamster lung (CHL) and ovary (CHO) cells, with and without metabolic activation (metabolic system S9).

After removing duplicates we collected a set of 477 organic compounds: 223 are classified as active and 254 are classified as inactive in chromosomal aberrations test. For each compound, CAS number, simplified molecular input-line entry system (SMILES) and experimental data expressed as active (+1) or inactive (−1) are represented. Finally, SMILES have been normalized by the VEGA platform (www.vega-qsar.eu/). These compounds were randomly split into the training (80%), calibration (10%), and validation (10%) sets (five splits are examined).

The CORAL software is developed with taking into account the following hypothesis: QSAR model is a random event (Toropov et al., 2013). In other words, the same approach that is used to build up a QSAR model gives quite different models for different splits into the training set and validation set. Thus, lucky splits (good statistical quality) and unlucky splits (poor statistical quality) take place for any total set that is used for the QSAR analysis. Consequently, in order to check up an approach really, one should examine a group of different distributions of available data into the training set (visible during building up a model) and the validation set (invisible during building up a model). This experiment confirms that there are lucky and unlucky splits, especially if large number of different splits are examined.

2.2. Optimal descriptor

The optimal descriptor used in this work is calculated as the following:

$$DCW(T^*, N^*) = \sum CW(S_k) + \sum CW(SS_k) + CW(HARD) \quad (1)$$

Simplified molecular input-line entry system (SMILES) (Weininger, 1988) is used to represent the molecular structure via SMILES attributes. In this work, two local SMILES attributes (S_k and SS_k) and one global SMILES attribute (HARD) are involved to build up predictive models.

The S_k are SMILES atoms, i.e. one symbol from SMILES or two symbols which cannot be examined separately, e.g. 'Cl', 'Br', etc. The SS_k are combines of two SMILES atoms. The HARD is global SMILES attribute, which reflects presence (absence) of Nitrogen, Oxygen, Sulphur, Phosphorus, Chlorine, Fluorine, Bromine, Iodine, double and triple covalent bonds (Toropov et al., 2013; Toropova et al., 2011; Toropov et al., 2012a). Table 1 contains example of definition for S_k, SS_k, and HARD. The T is threshold, i.e. integer to discriminate all SMILES attributes into two classes (i) rare, i.e. the number of the given attribute in the training set is less than threshold; and (ii) not rare, i.e. the number of given attribute in the training set is larger (or at least equal) than threshold. The N is the number of epochs of the Monte Carlo optimization of the target function (Toropov et al., 2013). The T = T∗ and N = N∗ are values of the parameters which give the best statistics for the calibration set. So-called semi-correlation (Toropov et al., 2012b; Toropova and Toropov, 2017) has been used to build up predictive models for chromosomal aberrations test. Fig. 1 elucidates the interrelations between semi-correlation and binary classification model. Fig. 2 contains an example of the model for chromosome aberration test.

2.3. Statistical criteria

In order to build up classification model i.e. separation of two classes (i) active (1); and (ii) inactive (−1) (Toropova and

Table 1
Examples of the Sk, SSk, and HARD for molecular structure represented by the following SMILES O = [N+]([O−])c1ccc(cc1)Cl.

S_k												SS_k
O.....												
=.....												O...=.....
[.....												[...=.....
N.....												[...N.....
+.....												N...+.....
[.....												[...+.....
(.....												[...(.....
[.....												[...(.....
O.....												[...O.....
−.....												O...−.....
[.....												[...−.....
(.....												[...(.....
c.....												c...(.....
1.....												c...1.....
c.....												c...1.....
c.....												c...c.....
c.....												c...c.....
(.....												c...(.....
c.....												c...(.....
c.....												c...c.....
1.....												c...1.....
(.....												1...(.....
Cl.....												Cl..(.....
		=	#	@	N	O	S	P	F	Cl	Br	I
HARD	$	1	0	0	1	1	0	0	0	1	0	0

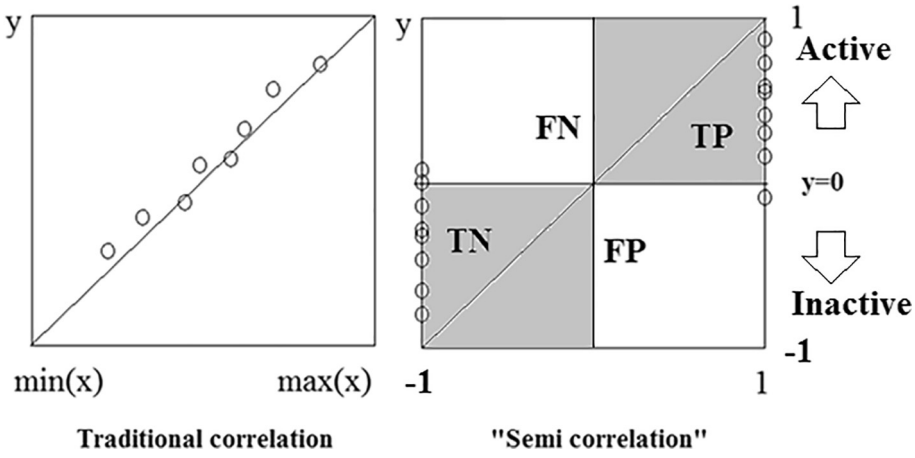


Fig. 1. Interpretations for traditional correlation and semi-correlation.

Toropov, 2017), the following statistical criteria have been used: sensitivity, specificity, accuracy, and Mattews correlation coefficient (MCC).

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

In these equations TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively, in a confusion matrix.

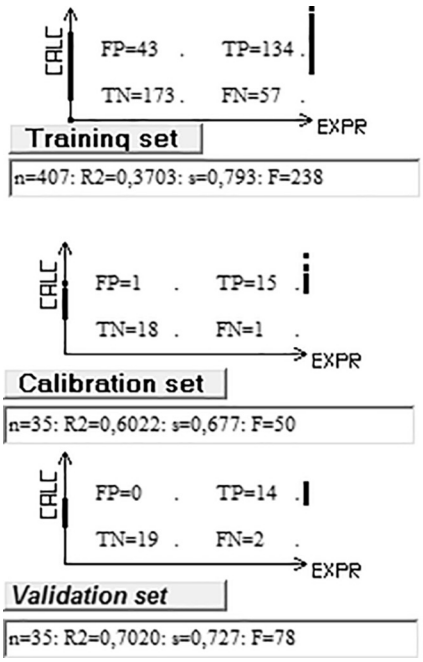
The MCC coefficient is used in machine learning as a balanced measure of the quality of binary classifications and it is useful even if the classes are of very different sizes (Dao et al., 2011).

A model is good if MCC → 1 (in praxis, the MCC should be larger than 0.6).

2.4. Domain of applicability

Domain of applicability is important component of a QSAR analyses. Diversity of QSAR approaches cause the diversity of conceptions for domain of applicability. A collection of conceptions of domain of applicability is available in literature (Gadaleta et al., 2016): (i) Chemical-physical domain; (ii) Structural domain; (iii) Response domain; and (iv) Integrated methods.

However, in the case of the CORAL models, the statistical defects of SMILES calculated according to distribution of available data into the training, invisible training, calibration, and validation sets are the basis to define domain of applicability. The defect of



Training set

TP= 134 TN= 173 FP= 43 FN= 57 N= 407
Sensitivity= 0.7016
Specificity= 0.8009
Accuracy = 0.7543
MCC = 0.5059

Calibration set

TP= 15 TN= 18 FP= 1 FN= 1 N= 35
Sensitivity= 0.9375
Specificity= 0.9474
Accuracy = 0.9429
MCC = 0.8849

Validation set

TP= 14 TN= 19 FP= 0 FN= 2 N= 35
Sensitivity= 0.8750
Specificity= 1.0000
Accuracy = 0.9429
MCC = 0.8898

Fig. 2. Graphical representation of semi correlations for split 2 ("lucky split") and statistical characteristics of this model for chromosome aberration test. TP = true positive; TN = true negative; FP = false positive; and FN = false negative.

SMILES attribute is defined via the difference of the probability of the attribute in the training set and probability of the attribute in the calibration set. The SMILES-defect is the summation of these defects of attributes. If a SMILES is characterized by the SMILES-defect which is lower than the doubled average defect over compounds of the training set, the SMILES falls into the domain of applicability, otherwise the SMILES is out of the domain of applicability (Toropova and Toropov, 2017):

$$Attribute\_Defect = \frac{|P(A) - P'(A)|}{N(A) + N'(A)} \quad (6)$$

The P(A) and P′(A) are probabilities of attribute A in the training and calibration sets, respectively. The N(A) and N′(A) are frequencies of A in the training and calibration sets, respectively.

$$SMILES\_Defect = \sum_{k=1}^{NA} Attribute\_Defect[k] \quad (7)$$

The NA is the number of attributes in a SMILES.

$$Domain\ Applicability = \begin{cases} YES,\ if\ SMILES\_Defect < 2^{*}\overline{SMILES\_Defect} \\ NO,\ if\ SMILES\_Defect > 2^{*}\overline{SMILES\_Defect} \end{cases} \quad (8)$$

The $\overline{SMILES\_Defect}$ is average *SMILES_defect* over training set.

3. Results and discussion

Table 2 contains the statistical characteristics for models of chromosome aberration test built up with the CORAL software. Table 3 contains the statistical characteristics for models suggested in the literature. One can see that the CORAL models are satisfactory and comparable with the analogical models from the literature. Results for the training set are in the range of 0.67–0.76 for sensitivity. Better results have been always obtained for specificity, with values reaching 0.83. The values for accuracy are of course between those of sensitivity and specificity, within a very sharp

Table 2
The statistical quality of models for chromosome aberration test.

Split	Set	n	Sensitivity	Specificity	Accuracy	MCC
1	Training	399	0.7592	0.7981	0.7794	0.5578
	Calibration	39	0.8333	0.8667	0.8462	0.6868
	Validation	39	0.8750	0.8387	0.8462	0.6244
2	Training	407	0.7016	0.8009	0.7543	0.5059
	Calibration	35	0.9375	0.9471	0.9429	0.8849
	Validation	35	0.8750	1.000	0.9429	0.8898
3	Training	380	0.7348	0.7889	0.7632	0.5248
	Calibration	49	0.9333	0.8235	0.8571	0.7097
	Validation	48	0.8148	1.000	0.8958	0.8112
4	Training	398	0.7513	0.7707	0.7613	0.5221
	Calibration	40	0.9412	0.9565	0.9500	0.8977
	Validation	39	1.000	0.6923	0.7949	0.6574
5	Training	399	0.6742	0.8326	0.7619	0.5156
	Calibration	39	0.7600	1.000	0.8462	0.7294
	Validation	39	0.8500	0.9474	0.8974	0.7995

Table 3
The statistical quality of models for chromosome aberration test suggested in the literature.

Reference	Set	n	Sensitivity	Specificity	Accuracy
Multicase methodology Rothfuss et al. (2006)	Training	537	0.528[a]	0.75[a]	0.649[a]
	Internal Validation	53	0.568[a]	0.717[a]	0.651[a]
Machine learning Rothfuss et al. (2006)	Training	521	0.751[b]	0.768[b]	0.76[b]
	Validation	58	0.708[b]	0.714[b]	0.716[b]
Rosenkranz (2004)	Dataset in 9 cross-validation folds	190	0.54	0.70	0.62
(KNN) Serra et al.(2003)	Training	346	0.693	0.861	0.812
	Validation	37	0.727	0.923	0.865
(SVM) Serra et al.(2003)	Training	308	0.989	1	0.997
	Cross-validation	38	0.727	1	0.921
	Validation	37	0.727	0.885	0.838
Estrada and Molina (2006)	Training	216	0.849	0.869	0.86
	Validation	156	0.818	0.829	0.828

[a] Mean value of 10 indipendent validations.
[b] Values represent mean ± standard deviation of 20 indipendent validations.

range, between 0.76 and 78. Indeed, the split 5, which has the lowest sensitivity value, has the highest specificity value, while split 1, with the highest sensitivity value, has a relatively low specificity value. As it often happens with CORAL, highest statistical parameters have been obtained with the calibration set. Better results for specificity are observed on the validation set, with values in the range between 0.81 and 1.0.

The basic hypothesis for the CORAL software is "the good statistical quality of a model for calibration set should be accompanied by the good statistical quality of the model for external validation set". According this conception the best CORAL model observes for split #4 (MCC = 0.8977). However, for other splits the MCC is quite satisfactory with values larger than 0.6.

The fluctuations of the different splits are due to the relatively limited number of chemicals. In these circumstances, only a few substances, which are false positives or false negatives in one or the other split, have high impact on the statistical values. Anyhow, the five splits provide a realistic scenario of the possible expected results in different cases. The general picture of the data indicate that the values are always good, for all criteria examined here.

The statistical parameters of other models published in the literature are quite similar to those we obtained. The best published model (Rothfuss et al., 2006) gave sensitivity for the training set of 0.75. The CORAL-model gives similar quality (0.76). The specificity of the CORAL-model is higher. The model by Rosenkranz (2004) has low statistical quality, quite similar to the model developed by Rothfuss et al. (2006) through Multicase methodology. In addition, the CORAL shows better predictive potential for the validation set than the model by Estrada Estrada and Molina (2006). The Support Vector Machine described by Serra et al. (2003) gives prediction poorer that the CORAL. Thus, the CORAL software gives useful predictions for examined endpoint.

4. Conclusions

The suggested models are built up according to OECD principles. The statistical quality of the models is comparable with similar models suggested in the literature. The semi-correlation is special category used in the CORAL software to build up the binary classifications, in form Yes/No, Active/Inactive. Factually, the approach (semi-correlations) has no analogies. However there are successful attempts to use the approach as a tool of SAR analysis (Toropov et al., 2012b,c; Toropova and Toropov, 2017). The principle "QSAR is a random event" is confirmed for the case of the semi-correlations developed for different splits into the training and validation sets (Table 2). In other words, the predictive potential of the semi-correlations takes place for all splits, but there are dispersion of statistical characteristics for different splits: there are lucky splits (e.g. #2) and there are unlucky splits (e.g. #1).

Acknowledgments

Authors thank the project LIFE-COMBASE contract (LIFE15 ENV/ES/000416) for financial support.

Conflict of interest

The author confirms that this article content has no conflicts of interest.

References

Ames, B.N., 1979. Identifying environmental chemicals causing mutations and cancer. Science 204, 587–593.
Ben-David, U., Ha, G., Khadka, P., Jin, X., Wong, B., Franke, L., Golub, T.R., 2016. The landscape of chromosomal aberrations in breast cancer mouse models reveals driver-specific routes to tumorigenesis. Nat. Commun. 7, 12156–12160.
Boffetta, P., van der Hel, O., Norppa, H., Fabianova, E., Fucic, A., Gundy, S., Lazutka, J., Cebulska-Wasilewska, A., Puskailerova, D., Znaor, A., Kelecsenyi, Z., Kurtinaitis, J., Rachtan, J., Forni, A., Vermeulen, R., Bonassi, S., 2007. Chromosomal aberrations and cancer risk: results of a cohort study from Central Europe. Am. J. Epidemiol. 165 (1), 36–43.
Brookmire, L., Chen, J.J., Levy, D.D., 2013. Evaluation of the highest concentrations used in the in vitro chromosome aberrations assay. Environ. Mol. Mutagen. 54, 36–43.
Carrano, A.V., Natarajan, A.T., 1988. International considerations for population monitoring using cytogenetic techniques. Commission for protection against environmental mutagens and carcinogens. Mutat. Res. 204, 379–406.
Castro, M.M.A., Onsten, T.G.H., Moreira, J.C.F., de Almeida, R.M.C., 2006. Chromosome aberrations in solid tumors have a stochastic nature. Mutat. Res-Fund. Mol. M 600, 150–164.
Claxton, L.D., Umbuzeiro, G.de A., DeMarini, D.M., 2010. The Salmonella mutagenicity assay: the stethoscope of genetic toxicology for the 21st century. Environ. Health. Perspect. 118, 1515–1522.
Dao, P., Wang, K., Collins, C., Ester, M., Lapuk, A., Sahinalp, S.C., 2011. Optimally discriminative subnetwork markers predict response to chemotherapy. Bioinformatics 27, 205–213.
Duardo-Sanchez, A., Gonzalez-Diaz, H., 2013. Legal issues for chem-bioinformatics models. Front. Biosci. 5, 361–374.
Estrada, E., Molina, E., 2006. Automatic extraction of structural alerts for predicting chromosome aberrations of organic compounds. J. Mol. Graph. Model. 25 (3), 275–288.
Gadaleta, D., Mangiatordi, G.F., Catto, M., Carotti, A., Nicolotti, O., 2016. Applicability domain for QSAR models: where theory meets reality. Int. J. Quant. Struct.-Property Relationships 1(1), 45–63.
Gollapudi, B.B., Johnson, G.E., Hernandez, L.G., Pottenger, L.H., Dearfield, K.L., Jeffrey, A.M., Julien, E., Kim, J.H., Lovell, D.P., MacGregor, J.T., Moore, M.M., van Benthem, J., White, P.A., Zeiger, E., Thybaud, V., 2013. Quantitative approaches for assessing dose-response relationships in genetictoxicology studies. Environ. Mol. Mutagen. 54, 8–18.
González-Díaz, H., Arrasate, S., Gómez-SanJuan, A., Sotomayor, N., Lete, E., Besada-Porto, L., Ruso, J.M., 2013a. General theory for multiple input-output perturbations in complex molecular systems. 1. Linear QSPR electronegativity

models in physical, organic, and medicinal chemistry. Curr. Top. Med. Chem. 13, 1713–1741.
González-Díaz, H., Arrasate, S., Sotomayor, N., Lete, E., Munteanu, C.R., Pazos, A., Besada-Porto, L., Ruso, J.M., 2013b. MIANN models in medicinal, physical and organic chemistry. Curr. Top. Med. Chem. 13, 619–641.
González-Díaz, H., Vilar, S., Santana, L., Uriarte, E., 2007. Medicinal chemistry and bioinformatics–current trends in drugs discovery with networks topological indices. Curr. Top. Med. Chem. 7, 1015–1029.
Gozalbes, R., Vicente de Julián-Ortiz, J. 2018. Applications of chemoinformatics in predictive toxicology for regulatory purposes, especially in the context of the EU REACH legislation. Int. J. Quant. Struct.-Property Relationships 3(1), 1–24.
Grade, M., Difilippantonio, M.J., Camps, J., 2015. Patterns of chromosomal aberrations in solid tumors. Recent Results Cancer. Res. 200, 115–142.
Hosein, A.N., Wu, M., Arcand, S.L., Lavallée, S., Hébert, J., Tonin, P.N., Basik, M., 2010. Breast carcinoma-associated fibroblasts rarely contain p53 mutations or chromosomal aberrations. Cancer. Res. 70 (14), 5770–5777.
Jacobson-Kram, D., Contrera, J.F., 2007. Genetic toxicity assessment: employing the best science for human safety evaluation part I: Early screening for potential human mutagens. Toxicol. Sci. 96 (1), 16–20.
Mohr, J., Jain, B., Sutter, A., Laak, A.T., Steger-Hartmann, T., Heinrich, N., Obermayer, K., 2010. A maximum common subgraph kernel method for predicting the chromosome aberration test. J. Chem. Inf. Model. 50 (10), 1821–1838.
Nigam, P.K., 2009. Adverse reactions to cosmetics and methods of testing. Indian. J. Dermatol. Venereol. Leprol. 75, 10–19.
OECD, 2008a. Organization for Economic Co-operation and Development guideline for testing of the chemicals, n.471, bacterial reverse mutation test. Commission Regulation (EC) No 440. http://www.oecd.org/chemicalsafety/assessmentofchemicals.
OECD, 2008b. Organization for Economic Co-operation and Development guideline for testing of the chemicals, n.473, In Vitro Mammalian Chromosome Aberration Test. Commission Regulation (EC) No 440. http://www.oecd.org/chemicalsafety/risk-assessment/1948434.pdf.
Prado-Prado, F., Garcia-Mera, X., Rodriguez-Borges, J.E., Concu, R., Perez-Montoto, L. G., Gonzalez-Diaz, H., Duardo-Sanchez, A., 2013. Patents of bio-active compounds based on computer-aided drug discovery techniques. Front. Biosci. 5, 399–407.
REACH, 2006. Regulation (EC) No. 1907/2006 of the European Parliament and of the Council, of December 18, 2006, concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93.
Rennstam, K., Ahlstedt-Soini, M., Baldetorp, B., Bendahl, P.-O., Borg, Å., Karhu, R., Tanner, M., Tirkkonen, M., Isola, J., 2003. Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. Cancer Res. 63 (24), 8861–8868.
Riera-Fernández, P., Martín-Romalde, R., Prado-Prado, F.J., Escobar, M., Munteanu, C.R., Concu, R., Duardo-Sanchez, A., González-Díaz, H., 2012. From QSAR models of drugs to complex networks: state-of-art review and introduction of new Markov-spectral moments indices. Curr. Top. Med. Chem. 12, 927–960.
Rosenkranz, H.S., 2004. SAR modeling of genotoxic phenomena: the consequence on predictive performance of deviation from a unity ratio of genotoxicants/non-genotoxicants. Mutat. Res. 559, 67–71.
Rothfuss, A., Steger-Hartmann, T., Heinrich, N., Wichard, J., 2006. Computational prediction of the chromosome-damaging potential of chemicals. Chem. Res. Toxicol. 19 (10), 1313–1319.
Serra, J.R., Thompson, E.D., Jurs, P.C., 2003. Development of binary classification of structural chromosome aberrations for a diverse set of organic compounds from molecular structure. Chem. Res. Toxicol. 16, 153–163.
Sun, Z., Prodduturi, N., Sun, S.Y., Thompson, E.A., Kocher, J.-P.A., 2015. Chromosome X genomic and epigenomic aberrations and clinical implications in breast cancer by base resolution profiling. Epigenomics 7, 1099–1110.
Tenorio-Borroto, E., Ramirez, F.R., Speck-Planche, A., Cordeiro, M.N., Luan, F., Gonzalez-Diaz, H., 2014. QSPR and flow cytometry analysis (QSPR-FCA): review and new findings on parallel study of multiple interactions of chemical compounds with immune cellular and molecular targets. Curr. Drug Metab. 15, 414–428.
Tenorio-Borroto, E., Peñuelas-Rivas, C.G., Vásquez-Chagoyán, J.C., Prado-Pradoa, F.J., García-Mera, X., González-Díaz, H., 2012. Immunotoxicity, flow cytometry, and chemoinformatics: review, bibliometric analysis, and new QSAR model of drug effects over macrophages. Curr. Top. Med. Chem. 12, 1815–1833.
Toropov, A.A., Toropova, A.P., Benfenati, E., Gini, G., Puzyn, T., Leszczynska, D., Leszczynski, J., 2012a. Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria *Escherichia coli*. Chemosphere 89 (9), 1098–1102.
Toropov, A.A., Toropova, A.P., Rasulev, B.F., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2012b. CORAL: Binary classifications (active/inactive) for liver-related adverse effects of drugs. Curr. Drug Saf. 7 (4), 257–261.
Toropov, A.A., Toropova, A.P., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2012c. CORAL: classification model for predictions of anti-sarcoma activity. Curr. Top. Med. Chem. 12 (24), 2741–2744.
Toropov, A.A., Toropova, A.P., Puzyn, T., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2013. QSAR as a random event: modeling of nanoparticles uptake in PaCa2 cancer cells. Chemosphere 92 (1), 31–37.
Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2011. CORAL: quantitative structure-activity relationship models for estimating toxicity of organic compounds in rats. J. Comput. Chem. 32 (12), 2727–2733.
Toropova, A.P., Toropov, A.A., 2014. CORAL software: prediction of carcinogenicity of drugs by means of The Monte Carlo method. Eur. J. Pharm. Sci. 52, 21–25.
Toropova, A.P., Toropov, A.A., 2017. CORAL: binary classifications (active/inactive) for drug-induced liver injury. Toxicol. Lett. 268, 51–57.
Umbuzeiro, G.deA., Heringa, M., Zeiger, E., 2016. In Vitro genotoxicity testing: significance and use in environmental monitoring. Chapter, Part of the series Advances in Biochemical Engineering/Biotechnology, pp. 1–22.
Votano, J.R., 2005. Recent uses of topological indices in the development of in silico ADMET models. Curr. Opin. Drug Discov. Develop. 8, 32–37.
Vulto-van Silfhout, A.T., van Ravenswaaij, C.M.A., Hehir-Kwa, J.Y., Verwiel, E.T.P., Dirks, R., van Vooren, S., Schinzel, A., de Vries, B.B.A., De Leeuw, N., 2013. An update on ECARUCA, the european cytogeneticists association register of unbalanced chromosome aberrations. Eur. J. Med. Gen. 56, 471–474.
Watters, A.D., Going, J.J., Cooke, T.G., Bartlett, J.M.S., 2003. Chromosome 17 aneusomy is associated with poor prognostic factors in invasive breast carcinoma. Breast Cancer. Res. Treat. 77 (2), 109–114.
Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28, 31–36.
Yunis, J.J., 1983. The chromosomal basis of human neoplasia. Science 221, 227–236.