Nano World Journal

http://dx.doi.org/10.17756/nwj.2016-034

Research Article

Open Access

Quasi-QSPR to Predict Proteins Behavior Under Various Concentrations of Drug Using Nanoconductometric Assay

Nicola Bragazzi[1], Andrey A. Toropov[2], Alla P. Toropova[2], Eugenia Pechkova[1] and Claudio Nicolini[1,3*]

[1]Laboratories of Biophysics and Nanotechnology, Department of Experimental Medicine, University of Genova, Genova, Italy
[2]IRCCS, Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy
[3]Nanoworld Institute Labs, Fondazione ELBA Nicolini, Pradalunga, Bergamo, Italy

***Correspondence to:**
Professor Claudio Nicolini
President, Nanoworld Institute Fondazione
EL.B.A. Nicolini (FEN), Largo Redaelli 7
Pradalunga, Bergamo 24020, Italy
Tel/Fax: +39 035767215
E-mail: president@fondazioneelba-nicolini.org

Received: November 22, 2016
Accepted: December 01, 2016
Published: December 03, 2016

Citation: Bragazzi N, Toropov AA, Toropova AP, Pechkova E, Nicolini C. 2016. Quasi-QSPR to Predict Proteins Behavior Under Various Concentrations of Drug Using Nanoconductometric Assay. *NanoWorld J* 2(4): 71-77.

Copyright: © 2016 Bragazzi et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY) (http://creativecommons.org/licenses/by/4.0/) which permits commercial use, including reproduction, adaptation, and distribution of the article provided the original author and source are credited.

Published by United Scientific Group

Abstract

Conductometric monitoring of drug-gene and drug-protein interactions is of fundamental importance in functional proteomics. Here, we model our previously obtained findings and characterizations of an important antiblastic used in neuro-oncology (Temozolomide), interacting with selected proteins that represent predictive biomarkers of the rate survival of the patients, of the outcome of chemotherapy and resistance to drug itself (namely, BRIP1 and MLH1) acquired with Nucleic Acid Programmable Protein Arrays (NAPPA)-based nanoconductometric sensor. Quasi-SMILES which are analogies of the traditional SMILES (simplified molecular input-line entry systems) used to represent molecular structure are suggested as a tool to represent complex substances which are acting under different conditions (dose, different peptides). By means of the optimal descriptors quasi-QSPR for conductance and frequency are established. Statistical quality of these models is satisfactory.

Keywords

SMILES, Quasi-SMILES, Optimal descriptor, Monte Carlo method, CORAL software, NAPPA, Nanoconductometer, QCM_D

Introduction

Conductometric monitoring of drug-gene and drug-protein interactions is of fundamental importance in the broad field of functional proteomics [1-8]. Here, we model our previously obtained findings and characterizations of an important antiblastic used in neuro-oncology (Temozolomide), interacting with selected proteins that represent predictive biomarkers of the rate survival of the patients, of the outcome of chemotherapy and resistance to drug itself (namely, BRIP1 and MLH1) acquired with Nucleic Acid Programmable Protein Arrays (NAPPA)-based nanoconductometric sensor [9-11].

The classic quantitative structure–property relationships (QSPRs) are a tool to predict physicochemical endpoints related to various substances which are represented by their molecular structure via correlations of different endpoints with various molecular descriptors calculated with the molecular graph [12-19] or SMILES [20-24].

However, during last decade the considerable number of substances where the molecular structure cannot be involved in the QSPR research have been recognized as an important component of the everyday life [25-27]. These substances are different polymers, proteins, and nanomaterials. In the cases of the

above-mentioned substances quasi-QSPR can be involved as a tool to predict their physicochemical endpoints.

Results

The above-mentioned eclectic data for the case of the frequency in Hz (F), and conductance in mS (G) are represented by Tables 1 and 2. Quasi-SMILES are a tool to represent new paradigm and Endpoint = F (Eclectic data), where Eclectic data is all conditions (controlled or observed) which have influence on results of experiments. The correlation weights of elements of quasi-SMILES give possibility to establish influence of various conditions for results of experiments. Indeed, quasi-SMILES are not traditional SMILES. Tables 1 and 2 contain description of codes utilized to build up quasi-SMILES. This approach is actually need in the case small data related to complex experiments.

The quasi-SMILES used as the basis of building up the models are represented together with numerical data on the endpoints (G and F) in Table 3.

Table 1: Codes for various dose.

Dose	Code[k]
1	A
2	B
5	C
10	D
20	E
50	F
100	G
200	H

Table 2: Codes for the presence (absence) of two proteins.

Proteins	Code[k]
MLH1	1
BRIP1	2
MLH1 and BRIP1	3

According to OECD principles the traditional QSPR should have defined domain of applicability. For the case of the quasi-SMILES the domain of applicability is defined according to the following principles.

The measure of statistical quality of attributes A_k which are involved to build up model can be estimated as the following:

$$\text{defect}(A_k) = \begin{cases} \dfrac{|P_{TRN}(A_k) - P_{CLB}(A_k)|}{N_{TRN}(A_k) + N_{CLB}(A_k)}, & \text{if } N_{CLB}(A_K) > 0 \\ 1, \text{ otherwise} \end{cases} \quad \ldots\ldots\ldots\ldots (1)$$

where the $P_{TRN}(A_k)$ is the probability of presence of the SA_k in SMILES of the training set, i.e.

$$P_{TRN}(A_k) = N_{TRN}(A_k) / N_{TRN}$$

The $P_{CLB}(A_k)$ is the probability of presence of the A_k in SMILES of the calibration set, i.e.

$$P_{CLB}(A_k) = N_{CLB}(A_k) / N_{CLB}$$

Table 3: quasi-SMILES used to represent available eclectic data which have impact on endpoints (G and F).

No.	Quasi -SMILES	G (mS)	F (Hz)
1	A1	0.29	17076
2	B1	0.2604	13548
3	C1	0.2247	14604
4	D1	0.2245	14472
5	E1	0.198	15288
6	F1	0.169	14028
7	G1	0.1563	13920
8	H1	0.0566	9480
9	A2	0.12	8352
10	B2	0.11	11016
11	C2	0.11	11616
12	D2	0.11	11352
13	E2	0.11	10740
14	F2	0.1	9600
15	G2	0.09	11016
16	H2	0.08	9600
17	A3	0.3957	10920
18	B3	0.3458	10668
19	C3	0.3468	10044
20	D3	0.2871	10212
21	E3	0.1601	7992
22	F3	0.0777	6360
23	G3	0.046	5772
24	H3	0.0462	5664

The $N_{TRN}(A_k)$ is the number (frequency) of SMILES which contain A_k in the training set;

The N_{TRN} is the total number of SMILES in the training set;

The $N_{CLB}(A_k)$ is the number (frequency) of SMILES which contain A_k in the calibration set;

The N_{CLB} is the total number of SMILES in the calibration set.

The defect(A_k) = 1, if A_k absents in the calibration set.

The logic: If the probability of an attribute be in the training set is equal to the probability of the attribute in the calibration set it is the ideal situation and the defect is zero. However, this situation is not typical, i.e. the difference between the probability of an attribute in the training set and the probability of the attribute in the calibration set is not zero. Under such circumstances, the frequency of an attribute in the training set and in the calibration set also should be taken into account: if these are small then the defect of the attribute must be larger. Finally, if A_k is absent in the calibration set, the defect(A_k) is maximal. Thus, the measure calculated with Equation 9 can be used for estimation of the statistical significance of A_k (Table 3) involved in building up model.

The criterion definition of domain of applicability for a quasi-SMILES. Having the numerical data on conductance one can estimate reliability of the model for a representation of protein behaviors under different concentration of drug(s) by a quasi-SMILES (Table 2): the basic hypothesis is "the probability of the quasi-SMILES to be in the domain of applicability is inversely proportional of sum of A_k-defects

$$\text{Defect-quasi-SMILES} = \sum \text{defect}(A_k) \quad \text{.....(2)}$$

If the Defect-quasi-SMILES calculated with Equation 2 is equal to zero this is an ideal situation. However, in praxis, the ideal situation is rare. Consequently, one should define some limitation for the Defect-quasi-SMILES value. The possible selection for the limit is the following:

$$\text{Defect-quasi-SMILES} < 2 * \overline{\text{Defect-quasi-SMILES}} \quad \text{.....(3)}$$

where $\overline{\text{Defect-quasi-SMILES}}$ is average of the Defect-quasi-SMILES for the training set.

The inequality 3 should be classified as a semi-qualitative criterion, because the large value of the Defect-quasi-SMILES is not the guarantee, the prediction for substance represented by the quasi-SMILES will be poor, and vice versa, the small value of the Defect-quasi-SMILES is not the guarantee that the prediction will be good. However, "probabilistic" meaning of this criterion is quite transparent.

The quasi-QSPR model for the conductance (G) (Tables 4 and 5) is the following

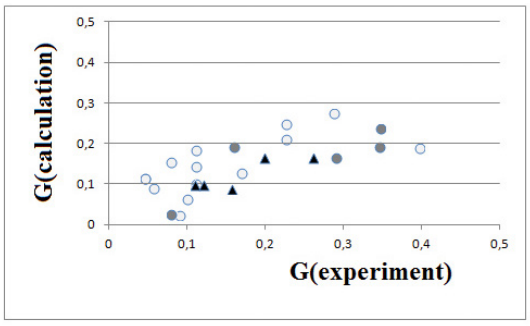
$$G(\text{mS}) = 0.06570\ (\pm 0.00754) + 0.0540839\ (\pm 0.004) * \text{DCW}(2,24) \quad \text{.....(4)}$$

n=14, r^2=0.46, s=0.075, F=10 (training set)

n=5, r^2=0.64, s=0.118 (calibration set)

n=5, r^2=0.72, s=0.061 (validation set)

Figure 1 contain the graphical representation of the model calculated with Equation 4.



Training set (○) Calibration set (●) Validation set (▲)

Figure 1: Graphical representation of G model calculated with Equation 4.

Table 4: The correlation weights for quasi-QSPR model for conductance (G).

Code[k]	Correlation weight(Code[k])
1	1.84537
2	0.64666
3	2.31175
A	0.0
B	0.0
C	0.82925
D	1.54966
E	0.0
F	-0.66670
G	-1.39892
H	-1.40334

Table 5: Distribution of available data into the training, calibration and validation sets together with experimental data on the conductance (G).

ID	Set	Quasi-SMILES	DCW (2,24)	Expr	Calc	Expr-Calc
3	train	C1	2.67463	0.2250	0.2104	0.0146
4	train	D1	3.39503	0.2250	0.2493	-0.0243
6	train	F1	1.17867	0.1690	0.1295	0.0395
8	train	H1	0.44203	0.0570	0.0896	-0.0326
11	train	C2	1.47591	0.1100	0.1455	-0.0355
12	train	D2	2.19631	0.1100	0.1845	-0.0745
13	train	E2	0.64666	0.1100	0.1007	0.0093
14	train	F2	-0.02005	0.1000	0.0646	0.0354
15	train	G2	-0.75227	0.0900	0.0250	0.0650
17	train	A3	2.31175	0.3960	0.1907	0.2053
20	train	D3	3.86141	0.2870	0.2745	0.0125
22	train	F3	1.64505	0.0780	0.1547	-0.0767
23	train	G3	0.91283	0.0460	0.1151	-0.0691
24	train	H3	0.90841	0.0460	0.1148	-0.0688
1	calib	A1	1.84537	0.2900	0.1655	0.1245
16	calib	H2	-0.75668	0.0800	0.0248	0.0552
18	calib	B3	2.31175	0.3460	0.1907	0.1553
19	calib	C3	3.14101	0.3470	0.2356	0.1114
21	calib	E3	2.31175	0.1600	0.1907	-0.0307
External validation set						
		Quasi-SMILES	**DCW (2,24)**	**Expr**	**Calc**	**Applicability**
2	Valid	B1	1.84537	0.2600	0.1655	YES
5	Valid	E1	1.84537	0.1980	0.1655	YES
7	Valid	G1	0.44645	0.1560	0.0899	No
9	Valid	A2	0.64666	0.1200	0.1007	YES
10	Valid	B2	0.64666	0.1100	0.1007	YES

The quasi-QSPR model for the frequency (F) (Tables 6 and 7) is the following

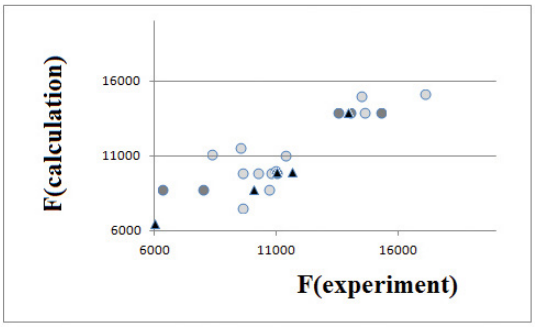
$$F(\text{Hz}) = 8243{,}12\ (\pm 284{,}29) + 2439{,}09\ (\pm 188{,}39) * \text{DCM}(2,21) \quad \text{...(5)}$$

n=13, r^2=0,66, s=1713, F=21 (training set)

n=6, r^2=0,89, s=1404 (calibration set)

n=5, r^2=0,72, s=0,061 (validation set)

Figure 2 contain the graphical representation of the model calculated with Equation 5.



Training set (○) Calibration set (●) Validation set (▲)

Figure 2: Graphical representation of F model calculated with Equation 5.

Method

Nanogravimetry makes use of functionalized piezoelectric Quartz Crystals (QC), which vary their resonance frequency

Table 6: The correlation weights for quasi-QSPR model for frequency (F).

Code[k]	Correlation weight(Code[k])
1	2,32213
2	0,68485
3	0,22742
A	0,51170
B	0,0
C	0,0
D	0,46432
E	0,0
F	0,0
G	0,0
H	-0,95951

Table 7: Distribution of available data into the training, calibration and validation sets together with experimental data on the frequency (F).

ID	Set	Quasi-SMILES	DCW (2,21)	Expr	Calc	Expr-Calc
1	train	A1	2,83382	17076,	15155,0628	1920,9372
3	train	C1	2,32213	14604,	13906,9882	697,0118
4	train	D1	2,78644	14472,	15039,4942	-567,4942
8	train	H1	1,36261	9480,	11566,6488	-2086,6488
9	train	A2	1,19655	8352,	11161,6109	-2809,6109
12	train	D2	1,14917	11352,	11046,0423	305,9577
13	train	E2	0,68485	10740,	9913,5362	826,4638
14	train	F2	0,68485	9600,	9913,5362	-313,5362
16	train	H2	-0,27466	9600,	7573,1969	2026,8031
17	train	A3	0,73912	10920,	10045,9004	874,0996
18	train	B3	0,22742	10668,	8797,8257	1870,1743
20	train	D3	0,69174	10212,	9930,3317	281,6683
23	train	G3	0,22742	5772,	8797,8257	-3025,8257
2	calib	B1	2,32213	13548,	13906,9882	-358,9882
5	calib	E1	2,32213	15288,	13906,9882	1381,0118
6	calib	F1	2,32213	14028,	13906,9882	121,0118
15	calib	G2	0,68485	11016,	9913,5362	1102,4638
21	calib	E3	0,22742	7992,	8797,8257	-805,8257
22	calib	F3	0,22742	6360,	8797,8257	-2437,8257
External validation set						
ID		**Quasi-SMILES**	**DCW (2,21)**	**Expr**	**Calc**	**Applicability**
7	Valid	G1	2,32213	13920,	13906,9882	YES
10	Valid	B2	0,68485	11016,	9913,5362	YES
11	Valid	C2	0,68485	11616,	9913,5362	YES
19	Valid	C3	0,22742	10044,	8797,8257	YES
24	Valid	H3	-0,73209	5664,	6457,4863	No

(f) when a mass (m) is adsorbed to or desorbed from their surface [28-30]. This is well described by the well-known Sauerbrey's equation: 0 Δ = – f / f m / A 1 ρ where, f_0 is the fundamental frequency, A is the surface area covered by the adsorbed molecule and ρ and 1 are the quartz density and thickness, respectively. Quartz resonators response strictly depends on the biophysical properties of the analyte, such as the viscoelastic coefficient. The dissipation factor (D) of the crystal's oscillation is correlated with the softness of the studied material and its measurement can be computed by taking into account the bandwidth of the conductance curve 2Γ, according to the following equation:

$$D = 2\Gamma / f$$

where, f is the peak frequency value. In our analysis, we introduced also a "normalized D factor", DN, that we defined as the ratio between the halfwidth half-maximum (Γ) and the half value of the maximum value of the conductance (G_{max}) of the measured conductance curves [4, 5]:

$$DN = 2\Gamma / G_{max}$$

DN is more strictly related to the curve shape, reflecting the conductance variation [4, 5].

The QCM_D instrument was developed by Elbatech (Elbatech srl, Marciana-LI, Italy). The quartz was connected to an RF gain-phase detector (Analog Devices, Inc., Norwood, MA, USA) and was driven by a precision DDS (Analog Devices, Inc., Norwood, MA, USA) around its resonance frequency, thus acquiring a conductance versus frequency curve ("conductance curve") which shows a typical Gaussian behaviour. The conductance curve peak was at the actual resonance frequency while the shape of the curve indicated how the viscoelastic effects of the surrounding layers affected the oscillation. The QCM_D software, QCMAgic-Q5.3.256 (Elbatech srl, Marciana-LI, Italy) allows to acquire the conductance curve or the frequency and dissipation factor variation versus time. In order to have a stable control of the temperature, the experiments were conducted in a temperature chamber. Microarrays were produced on standard nanogravimetry quartz used as highly sensitive transducers. The QC expressing proteins consisted of 9.5 MHz, AT-cut quartz crystal of 14 mm blank diameter and 7.5 mm electrode diameter, produced by ICM (Oklahoma City, USA). The electrode material was 100 Å Cr and 1000 Å Au and the quartz was [4, 5]. The NAPPA-QC arrays were printed with 100 spots per QC. Quartzes gold surfaces were coated with cysteamine to allow the immobilization of the NAPPA printing mix. Briefly, quartzes were washed three times with ethanol, dried with Argon and incubated over night at 4 °C with 2 mM cysteamine. Quartzes were then washed three times with ethanol to remove any unbound cysteamine and dried with Argon. Plasmids DNA coding for GST tagged proteins were transformed into *E. coli* and DNA were purified using the NucleoPrepII anion exchange resin (Macherey Nagel). NAPPA printing mix was prepared with 1.4 μg uL^{-1} DNA, 3.75 μg uL^{-1} BSA (Sigma-Aldrich), 5 mM BS3 (Pierce, Rockford, IL, USA) and 66.5 μg polyclonal capture GST antibody (GE Healthcares). Negative controls, named master mix (hereinafter abbreviated as "MM"), were obtained replacing DNA for water in the printing mix. Samples were incubated at room temperature for 1 h with agitation and then printed on the cysteamine-coated gold quartz using the Qarray II from Genetix. In order to enhance the sensitivity, each quartz was printed with 100 identical features of 300 microns diameter each, spaced by 350 microns center-to-center. The human cDNAs immobilized on the NAPPA-QC were: MLH1 (mutL homolog 1) and BRIP1 (BRCA1 interacting protein C-terminal helicase 1). Gene expression was performed immediately before the assay, following the protocol described by Spera et al. [4]. Briefly, IVTT was performed using HeLa lysate mix (1-Step Human Coupled IVTT Kit, Thermo Fisher Scientific Inc.), prepared according

to the manufacturers' instructions. The quartz, connected to the nanogravimeter inside the incubator, was incubated for 10 min at 30 °C with 40 µL of HeLa lysate mix for proteins synthesis and then, the temperature was decreased to 15 °C for a period of 5 min to facilitate the proteins binding on the capture antibody (anti-GST). After the protein expression and capture, the quartz was removed from the instrument and washed at room temperature, in 500 mM NaCl PBS for 3 times. The protocol described above was followed identically for both negative control QC (the one with only MM, i.e., all the NAPPA chemistry except the cDNA) and protein displaying QC. After protein expression, capture and washing the QCs were used for the interaction studies QC displaying the expressed protein was spotted with 40 µl of drug solutions in PBS at increasing concentrations at 22 °C. Reproducibility of the experiments was assessed computing the Coefficient of Variation (CV, or σ*), using the following equation:

$$\sigma^* = \sigma / \mu$$

where, σ is the standard deviation and µ is the mean. We also tested the possibility to analyze drug-protein interactions in QC displaying multiple proteins. For this aim, we co-printed cDNA for BRIP1&MLH1 on a single QC. We analyzed the interaction response to TMZ on both NAPPA-expressed QCs. We analyzed the interaction between BRIP1, MLH1 and TMZ drug solutions at different concentrations to analyze the binding kinetics after protein expression and capture the expressing QC was spotted, in sequence, with 40 µL of increasing Temozolomide solutions of concentration: 1, 2, 5, 10, 20, 50, 100 and 200 µg mL^{-1}. As negative control we analyzed the interaction between BRIP1/FANCJ, a helicase initially linked to breast cancer [31] and to Fanconi anemia and TMZ, while MLH1, which is a protein involved in DNA mismatches repair, is known to interact with TMZ.

The basic idea of the quasi-QSPR is replacing of the classic paradigm

Endpoint = F(Molecular structure)

by the new paradigm

Endpoint = F (Eclectic data).

It is to be noted the molecular structure (sometimes fragments of the molecular structure) can be involved in building up a predictive model. In this case one faced with a hybrid paradigm

Endpoint = F (Molecular structure and Eclectic data).

Figure 3 shows the generalized illustration for the situations where above new paradigms can be used to solve the practical tasks.

In practical aspect, the process of build up a model can be defined as the following:

1. Selection of eclectic data (impacts which can have influence to endpoint of interest);

2. The representation of these data by means of SMILES like lines, which can be named as "quasi-SMILES";

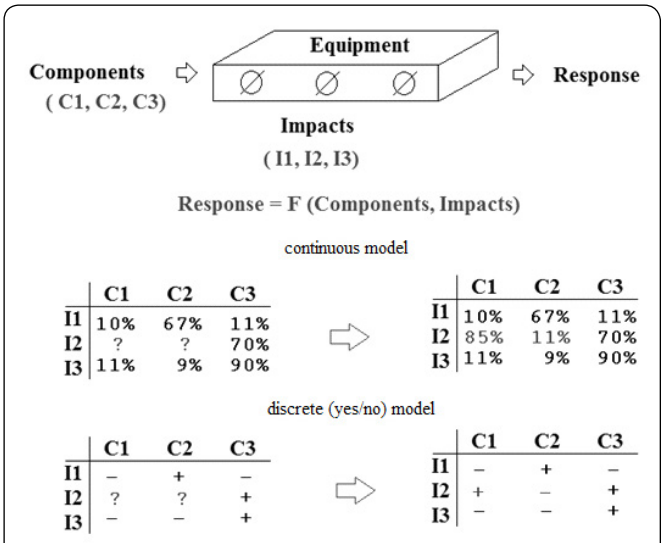


Figure 3: Graphical representation of situations which can be represented by the quasi-SMILES.

3. Calculation by the Monte Carlo method correlation weights for various impacts which give maximal correlation coefficicnet between descriptor and endpoint for (a) training set; and (b) calibration set; the descriptor is calculated by formula:

DCW(Threshold, N) = Σ Correlation weight(Code[k])

where Code[k] is the representation for k-th impact;

4. Calculation by the least square method the model

Endpoint = C0 + C1* DCW(Threshold,N) .....(6)

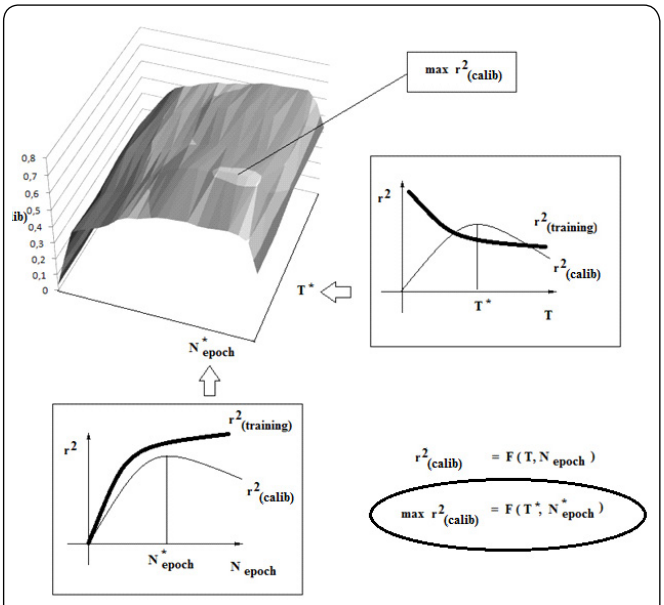


Figure 4: Graphical representation of selection for the preferable threshold and the number of epochs of the Monte Carlo optimization.

5. Estimation of the model with data distributed into external validation set. Data on the training set and calibration set are "visible" during building up the model, whereas data on the validation set are "invisible" during building up the model.

The threshold is parameter in order to define rare (noise) impacts. These should be removed from the modeling process. The N is number of epoch of the optimization. The number of epochs of the Monte Carlo optimization will be too large the statistical quality of the model for the training set will be increase but the statistical quality of the model for calibration set will be step by step decrease. Thus, one should define the number N which produce maximal statistical quality for the calibration set. Figure 4 shows the graphical interpretation of the selection preferable Threshold (T*) and the N (N*).

Conclusions

Thus, described approach gives satisfactory prediction for both endpoints, Unfortunately, the statistical quality and the domain of applicability of the model are dependent upon the distribution into the visible training and calibration sets and invisible calibration set, however, in the case of available large datasets this influence will be decreased, the main idea: prevalence of elements of quasi-SMILES in training and in validation sets be as identic as possible. This approach has been tested in several research works, by and with authors from various countries [32-38].

References

1. Nicolini C, Bragazzi NL, Pechkova E. 2016. Microarray-based functional nanoproteomics for an industrial approach to cancer: I bioinformatics and miRNAome. *NanoWorld J* 2(1): 1-4. doi: 10.17756/nwj.2016-020
2. Nicolini C, Bragazzi N, Pechkova E. 2016. Microarray-based functional nanoproteomics for an industrial approach to cancer. II mass spectrometry and nanoconductimetry. *NanoWorld J* 1(4): 128-132. doi: 10.17756/nwj.2016-017
3. Nicolini C. 2016. Molecular Bioelectronics. The 19 years of progress, Second edition, World Scientific, New Jersey, USA.
4. Spera R, Festa F, Bragazzi NL, Pechkova E, LaBaer J, et al. 2013. Conductometric monitoring of protein-protein interactions. *J Proteome Res* 12(12): 5535-5547. doi: 10.1021/pr400445v
5. Nicolini C, Bragazzi N, Pechkova E. 2012. Nanoproteomics enabling personalized nanomedicine. *Adv Drug Deliv Rev* 64(13): 1522-1531. doi: 10.1016/j.addr.2012.06.015
6. Nicolini C, Bezerra T, Pechkova E. 2012. Protein nanotechnology for new design and development of biocrystals and biosensors. *Nanomedicine (Lond)* 7(8): 1-4. doi: 10.2217/nnm.12.84
7. Nicolini C, Adami M, Sartore M, Bragazzi NL, Bavastrello V, et al. 2012. Prototypes of newly conceived inorganic and biological sensors for health and environmental applications. *Sensors* 12(12): 17112-17127. doi: 10.3390/s121217112
8. Nicolini C. 2010. Nanogenomics in medicine. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* 2(1): 59-76. doi: 10.1002/wnan.64
9. Nicolini C, LaBaer J. 2010. Functional Proteomics and Nanotechnology-based Microarrays. Pan Stanford Series on Nanobiotechnology, Singapore.
10. Nicolini C, Pechkova E. 2010. Nanoproteomics for nanomedicine. *Nanomedicine (Lond)* 5(5):677-682. doi: 10.2217/nnm.10.46
11. Nicolini C, Pechkova E. 2010. An overview of nanotechnology-based functional proteomics for cancer and cell cycle progression. *Anticancer Res* 30(6): 2073-2080.
12. Afantitis A, Melagraki G, Koutentis PA, Sarimveis H, Kollias G. 2011. Ligand-based virtual screening procedure for the prediction and the identification of novel β-amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks. *Eur J Med Chem* 46(2): 497-508. doi: 10.1016/j.ejmech.2010.11.029
13. Furtula B, Gutman I. 2011. Relation between second and third geometric–arithmetic indices of trees. *J Chemom* 25(2): 87-91. doi: 10.1002/cem.1342
14. García J, Duchowicz PR, Rozas MF, Caram JA, Mirífico MV, et al. 2011. A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases. *J Mol Graphics Model* 31: 10-19. doi: 10.1016/j.jmgm.2011.07.007
15. Garro Martinez JC, Duchowicz PR, Estrada MR, Zamarbide GN, Castro EA. 2011. QSAR study and molecular design of open-chain enaminones as anticonvulsant agents. *Int J Mol Sci* 12(12): 9354-9368. doi: 10.3390/ijms12129354
16. Ibezim E, Duchowicz PR, Ortiz EV, Castro EA. 2012. QSAR on aryl-piperazine derivatives with activity on malaria. *Chemom Intell Lab Syst* 110(1): 81–88. doi: 10.1016/j.chemolab.2011.10.002
17. Toropov AA, Roy K. 2004. QSPR modeling of lipid-water partition coefficient by optimization of correlation weights of local graph invariants. *J Chem Inf Comput Sci* 44(1): 179-186. doi: 10.1021/ci034200g
18. Toropov AA, Toropova AP. 2002. QSAR modeling of toxicity on optimization of correlation weights of Morgan extended connectivity. *J Mol Struct THEOCHEM* 578(1-3): 129-134. doi: 10.1016/S0166-1280(01)00695-9
19. Toropov AA, Toropova AP. 2003. QSPR modeling of alkanes properties based on graph of atomic orbitals. *J Mol Struct THEOCHEM* 637(1-3): 1-10. doi: 10.1016/S0166-1280(02)00492-X
20. Weininger D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 28(1): 31-36. doi: 10.1021/ci00057a005
21. Weininger D, Weininger A, Weininger JL. 1989. SMILES. 2. Algorithm for generation of unique SMILES notation. *J Chem Inf Comput Sci* 29(2): 97-101. doi: 10.1021/ci00062a008
22. Weininger D. 1990. SMILES. 3. DEPICT. Graphical depiction of chemical structures. *J Chem Inf Comput Sci* 30(3): 237-243. doi: 10.1021/ci00067a005
23. Toropov AA, Rasulev BF, Leszczynski J. 2008. QSAR modeling of acute toxicity by balance of correlations. *Bioorg Med Chem* 16(11): 5999-6008. doi: 10.1016/j.bmc.2008.04.055
24. Toropov AA, Toropova AP, Benfenati E, Gini G, Leszczynska D, et al. 2011. SMILES-based QSAR approaches for carcinogenicity and anticancer activity: comparison of correlation weights for identical SMILES attributes. *Anticancer Agents Med Chem* 11(10): 974-982. doi: 10.2174/187152011797927625
25. Toropova AP, Toropov AA, Puzyn T, Benfenati E, Leszczynska D, et al. 2013. Optimal descriptor as a translator of eclectic information into the prediction of thermal conductivity of micro-electro-mechanical systems. *J Math Chem* 51(8): 2230–2237. doi: 10.1007/s10910-013-0211-2
26. Toropova AP, Toropov AA. 2013. Optimal descriptor as a translator of eclectic information into the prediction of membrane damage by means of various TiO_2 nanoparticles. *Chemosphere* 93(10): 2650-2655.
27. Toropov AA, Toropova AP. 2014. Optimal descriptor as a translator of eclectic data into endpoint prediction: Mutagenicity of fullerene as a mathematical function of conditions. *Chemosphere* 104: 262-264. doi: 10.1016/j.chemosphere.2013.10.079
28. Nicolini C, Bragazzi N, Pechkova E. 2016. Quartz crystal microbalance with dissipation factor monitoring (QCM_D) protocol. *Protocol Exchange*. doi: 10.1038/protex.2016.003
29. Nicolini C. 1996. Molecular Bioelectronics. World Scientific Singapore, New York, USA.
30. Nicolini C. 1986. Bioscience at the Physical Science Frontier:

Proceedings of a Foundation Symposium on the 150th Anniversary of Alfred Nobel's Birth. Humana Press, Clifton, New Jersey, USA.

31. Cantor SB, Xie J. 2010. Assessing the link between BACH1/FANCJ and MLH1 in DNA crosslink repair. *Environ Mol Mutagen* 51(6): 500-507. doi: 10.1002/em.20568.
32. Manganelli S, Leone C, Toropov AA, Toropova AP, Benfenati E. 2016 QSAR model for cytotoxicity of silica nanoparticles on human embryonic kidney cells. *Materials Today: Proceedings* 3(3): 847-854. doi: 10.1016/j.matpr.2016.02.018
33. Manganelli S, Leone C, Toropov AA, Toropova AP, Benfenati E. 2016. QSAR model for predicting cell viability of human embryonic kidney cells exposed to SiO_2 nanoparticles. *Chemosphere* 144: 995-1001. doi: 10.1016/j.chemosphere.2015.09.086
34. Toropova AP, Toropov AA, Rallo R, Leszczynska D, Leszczynski J. 2015. Optimal descriptor as a translator of eclectic data into prediction of cytotoxicity for metal oxide nanoparticles under different conditions. *Ecotoxicol Environ Saf* 112: 39-45. doi: 10.1016/j.ecoenv.2014.10.003
35. Toropova AP, Toropov AA, Veselinović AM, Veselinović JB, Benfenati E, et al. 2016. Nano-QSAR: Model of mutagenicity of fullerene as a mathematical function of different conditions. *Ecotoxicol Environ Saf* 124: 32-36. doi: 10.1016/j.ecoenv.2015.09.038
36. Toropov AA, Toropova AP. 2015. Quasi-SMILES and nano-QFAR: United model for mutagenicity of fullerene and MWCNT under different conditions. *Chemosphere* 139: 18-22. doi: 10.1016/j.chemosphere.2015.05.042
37. Toropov AA, Achary PGR, Toropova AP. 2016. Quasi-SMILES and nano-QFPR: The predictive model for zeta potentials of metal oxide nanoparticles. *Chem Phys Lett* 660: 107-110. doi: 10.1016/j.cplett.2016.08.018
38. Toropov AA, Toropova AP. 2015. Quasi-QSAR for mutagenic potential of multi-walled carbon-nanotubes. *Chemosphere* 124: 40-46. doi: 10.1016/j.chemosphere.2014.10.067