Caspian J. Chem. 3(2014) 23-34

Caspian Journal of Chemistry

http://caschemistry.journals.umz.ac.ir

In-silico prediction of Cellular Responses to Polymeric Biomaterials from Their Molecular Descriptors

Mohammad H. Fatemi*, Ameneh Kerdarshad, Elham Gholami Rostami

Chemometrics Laboratory, Faculty of Chemistry, University of Mazandaran, Babolsar, Iran

Corresponding author: E-mail: mhfatemi@umz.ac.ir

Tel: +98 112 5242931 Fax: +98 112 5342350

Received 28 October 2013| Received in revised form 1 February 2014| Accepted 22 February 2014

Abstract:

In this work quantitative structure activity relationship (QSAR) methodology was applied for modeling and prediction of cellular response to polymers that have been designed for tissue engineering. After calculation and screening of molecular descriptors, linear and nonlinear models were developed by using multiple linear regressions (MLR) and artificial neural network (ANN) methods. The root mean square error (RMSE) of these models were $RMSE_{MLR}$=12.6 and $RMSE_{ANN}$=10.6. Robustness and reliability of the developed MLR and ANN models were evaluated by using the leave-one-out and leave many out cross-validation methods, which produces the statistics of Q^2_{MLR}=0.74 and Q^2_{ANN}=0.81. Moreover, the chemical applicability domains of these models were determined via leverage approach. The results of these tests indicate the suitability of developed models. Comparison of statistical parameters of MLR and ANN models indicate the suitability of non-linear over linear model. The results of this study revealed the high applicability of QSAR approach in prediction of cellular response to the polymeric biomaterials.

Key words: Quantitative structure–activity relationship, artificial neural network, cellular response, polymeric biomaterial, molecular descriptors.

®2014 Published by University of Mazandaran. All rights reserved.

1. Introduction

The term of "biomaterial" describes a material intended for use in a medical device or implant. These materials are one of the cornerstones of tissue engineering that have been used for years, but recently their degree of perfection has increased significantly [1]. Biomaterials made today are routinely information rich and incorporate biologically active components derived from nature [2]. One of the most important biomaterials are polymeric compounds. These compounds can be used with various shapes, at reasonable cost and with desirable mechanical and physical properties

[3]. A major recent development in this field is the design of biomaterials for tissue engineering matrices to achieve specific biologic effects on cells, which are new ways to analyze the human body's response to materials [4-9].

The complexity of living cells and their interactions with biomaterials has been a conceptual as well as a practical barrier to the use of advanced discovery tools in designing of new biomaterials [10]. In this way, computational methods that can predict the cellular response to implanted biomaterials would be invaluable in the design of new medical devices, whose functions depend on controlling cell–material interactions at the device surface [11]. Among theoretical methods, quantitative structure activity relationships (QSAR) approaches have been successfully established to predict the properties/activities of chemicals from their structural features. In this method the molecular structural descriptors of chemicals are quantitatively correlated to their biological activities [12-15]. These studies consist of two main stages; first, the chemical compounds are translated into a computer readable form, then the quantitative correlation between chemical structural features (molecular descriptors) and their activities can be established by using some feature mapping techniques such as: multiple linear regression (MLR), artificial neural network (ANN) and support vector machine (SVM) [16].

There are some reports about QSAR prediction of bio-responses to polymeric biomaterials [17-26]. For example, J.R. Smith et al. proposed a surrogate (semi empirical) model for prediction of protein adsorption onto the surfaces of biodegradable polymers that have been designed for tissue engineering applications by using an artificial neural network [19].The mean value of the coefficient of multiple correlation (R^2) and the average root-mean-square (relative) error in prediction for the validation data sets of their model was 0.54±0.12 and 38%, respectively. A.V. Gubskaya et al. were used low energy conformations derived descriptors from molecular dynamics simulations for 45 representatives of polyarylates in an improved QSAR model instead of simplistic two-dimensional representations of polymer structures [5].The significance of the newly developed 3D model is that it allows high accuracy prediction of fibrinogen adsorption without the need to experimentally-derived descriptors and it has better predictive quality than the original 2D surrogate model due to utilization of realistic polymer representations. In another work, a subset of 79 polymers that taken from a representative sub-library of 2000 polymethacrylates was used by V. Kholodovych et al. to build QSAR-based polynomial neural network models, which can used to predict cell attachment, cell growth, and fibrinogen adsorption on polymer surface [7].The resulting models gives good prediction statistics and allows to use on much larger set of poly methacrylates. Moreover, J. Ghoshet al. was examined the capabilities of QSAR modeling to predict specific biological responses of thin films of different poly methacrylates [26]. Based on the experimental results of these polymers, separate models were built for homo-, co-, and ter-polymers in the library, which gives good correlation between experimental and predicted values of specific biological responses. Also, Kholodovych et al. was used QSAR methodology to predict cell responses to a diverse set of 62 tyrosine derived biodegradable polymers, by using partial least squares (PLS) techniques [27].Their established five principle components (PCs) model gives the statistical parameters of R^2=0.62 and Q^2=0.56. The main aim of the present work is developing of some QSAR models for the above data set by computing new molecular

structural descriptors and using linear and non-linear feature mapping techniques to developing new QSAR models.

2. Methodology

2.1. Data Set

The values of response of fetal rat lung fibroblasts (FRLF) to 62 polymeric substrates were reported by Kholodovych et al. that were selected as data set (Table 1) [27].

They have used combinatorial chemistry techniques to prepare a series of structurally related polyarylates derived from monomers consisting of a tyrosine-derived diphenol and a diacid (Fig. 1).

Table. 1: List of 62 polyarylates used to build the QSAR model, together with corresponding experimental and ANN and MLR predicted values of FRLF NMA

No.	Diphenol	Diacid	FRLF NMA (%TCPS)	FRLF NMA_{ANN}	FRLF NMA_{MLR}
1	DTB	Adipate	32.29	31.55	42.23
2	DTB	Diglycolate	75.89	73.22	82.43
3	DTB	Dioxaoctanedioate	82.81	73.73	63.26
4	DTB	Methyl adipate	35.12	33.25	41.22
5	DTB	Sebacate	58.05	57.89	49.23
6	DTB	Suberate	76.34	76.47	64.21
7[a]	DTB	Succinate	75.82	79.53	73.26
8	DTBn	Adipate	52.48	52.54	31.93
9	DTBn	Diglycolate	73.77	73.84	75.94
10	DTBn	Dioxaoctanedioate	69.93	69.63	84.89
11	DTBn	Glutarate	71.49	70.54	64.14
12	DTBn	Methyl adipate	32.01	33.42	45.60
13	DTBn	Sebacate	66.53	60.24	60.12
14	DTBn	Suberate	67.24	72.73	68.95
15	DTBn	Succinate	77.77	78.14	82.59
16	DTD	Adipate	2.00	1.02	7.48
17	DTD	Diglycolate	67.65	73.15	61.16
18[a]	DTD	Dioxaoctanedioate	66.31	56.47	48.43
19	DTD	Glutarate	18.83	16.68	24.54
20	DTD	Methyl adipate	20.83	20.74	16.70
21	DTD	Sebacate	7.92	12.78	18.35
22[b]	DTD	Suberate	31.48	12.55	19.03
23	DTE	Adipate	75.66	73.54	59.38
24	DTE	Diglycolate	82	81.49	80.15
25	DTE	Dioxaoctanedioate	77.69	77.79	86.47
26	DTE	Glutarate	78.47	78.98	82.64
27	DTE	Methyl adipate	38.55	39.15	45.34
28[b]	DTE	Sebacate	68.4	84.87	71.74
29	DTE	Suberate	69.51	73.21	84.23
30	DTE	Succinate	97.59	97.26	86.32
31	DTH	Adipate	16.48	16.38	30.98
32	DTH	Diglycolate	69.5	69.51	70.17
33	DTH	Dioxaoctanedioate	59.09	57.60	67.66
34[b]	DTH	Glutarate	52.8	48.16	52.95
35	DTH	Methyl adipate	25.48	26.01	20.11
36	DTH	Sebacate	50.62	51.59	49.09
37	DTH	Suberate	63.64	63.47	51.34
38[*]	DTH	Succinate	30.18	-	-

39	DTiP	Adipate	62.36	63.34	58.09
40	DTiP	Diglycolate	81.44	80.92	88.59
41[b]	DTiP	Dioxaoctanedioate	70.89	78.39	93.38
42	DTiP	Glutarate	79.79	79.84	73.86
43	DTiP	Methyl adipate	85.01	85.20	65.23
44	DTiP	Sebacate	78.30	78.34	76.07
45	DTiP	Suberate	70.44	71.08	79.56
46	DTiP	Succinate	77.07	74.50	87.33
47[b]	DTM	Adipate	86.22	95.47	71.02
48	DTM	Diglycolate	88.33	91.53	93.70
49[*]	DTM	Dioxaoctanedioate	84.61	-	-
50	DTM	Glutarate	94.60	95.10	96.26
51	DTM	Methyl adipate	78.00	77.86	81.87
52	DTM	Sebacate	87.85	87.41	85.57
53[a]	DTM	Suberate	80.71	91.13	104.61
54	DTM	Succinate	114.66	114.86	97.90
55	DTO	Adipate	4.80	4.55	9.95
56	DTO	Diglycolate	66.69	66.93	70.12
57	DTO	Dioxaoctanedioate	71.67	70.62	59.33
58	DTO	Glutarate	43.18	32.80	37.49
59[a]	DTO	Methyl adipate	40.88	53.87	19.66
60[a]	DTO	Sebacate	17.58	31.36	34.11
61	DTO	Suberate	47.38	46.09	41.72
62	DTO	Succinate	25.57	41.83	44.53

In the above table [a] and [b] represent the internal and external test sets, respectively and *refers to the outliers.

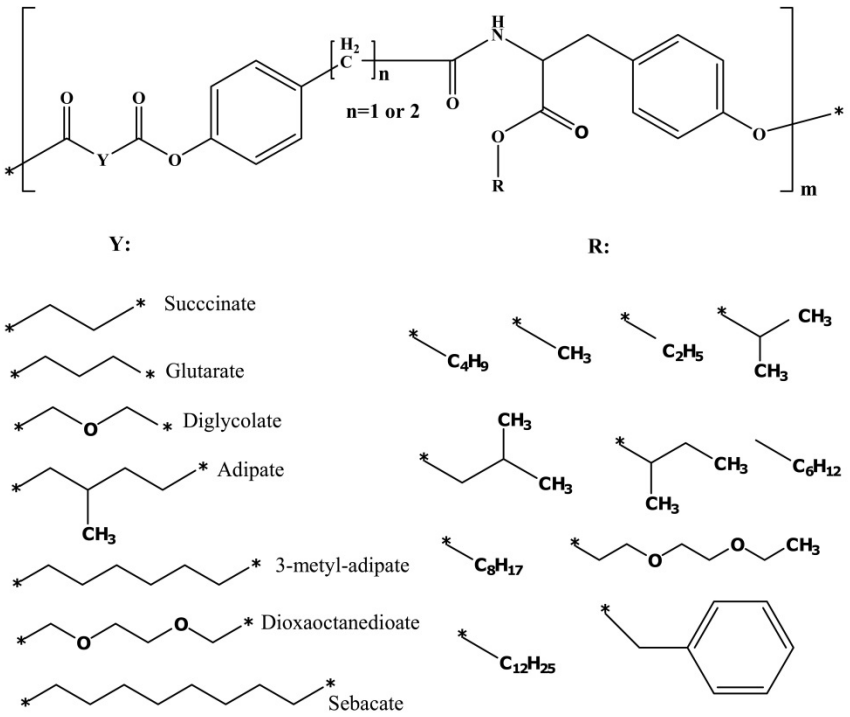

Figure 1. Library of 62 polyarylates obtained from 14 tyrosine-derived diphenols and eight diacids. Polymers are strictly alternating copolymers consisting of a diacid (DA) and a diphenol (DP) component varied at Y and R, respectively. The number of methyl groups in the DP component is also variable.

In the combinatorial approach, copolymers are synthesized from a set of x structural variations of 'A' and y structural variations of 'B'. The 'A' monomer template in the polyarylate library is the desaminotyrosyl–tyrosine alkylesters (DTR) diphenol while the 'B' monomer template is a dicarboxylic acid (Fig. 1) [28-29]. Cellular response to each polyarylate sample was then quantified as 'normalized metabolic activity' (NMA), which was its average measured metabolic activity given as a percentage of the average measured value for the separate tissue culture polystyrene (TCPS) wells [27].

Compounds in data set were sorted according to their FRLF NMA values, then the training (50), internal (5) and external (5) test sets were chosen from this list by desired distances from each other. The training set and internal test set participated in the developing of the ANN model and adjusting its parameters, while the external test set was used to evaluate the prediction power of the obtained model. In the case of MLR model, the training set was used to model generation and the internal and external test sets were considered as a test set, which was used in evaluation of MLR model.

2.2 Descriptors; generation and screening

Molecular descriptors are the simple mathematical representation of a molecule and are used to encode significant structural features of molecules. In order to calculate descriptors, the repeating structural unit of each co-polymer was drawn by Hyperchem program (ver. 7) [30]. Then, the optimization of these structures was done with the semi empirical AM1 method. The obtained Hyperchem output files were used by Dragon program [31]to calculate molecular descriptors.

After calculation of the molecular descriptors, those that stayed constant for all molecules were eliminated and pairs of variables with a correlation coefficient greater than 0.9 were classified as inter-correlated and one in each correlated pair was deleted. The remaining 187 descriptors were used to generate the QSAR models.

In the next step, the method of stepwise multiple linear regression was used for selection of the most relevant descriptors and MLR model construction [32]. To find the optimum number of descriptors, the influence of the number of the descriptors on the correlation coefficients (R^2) of MLR models were calculated and monitored in Fig. 2.

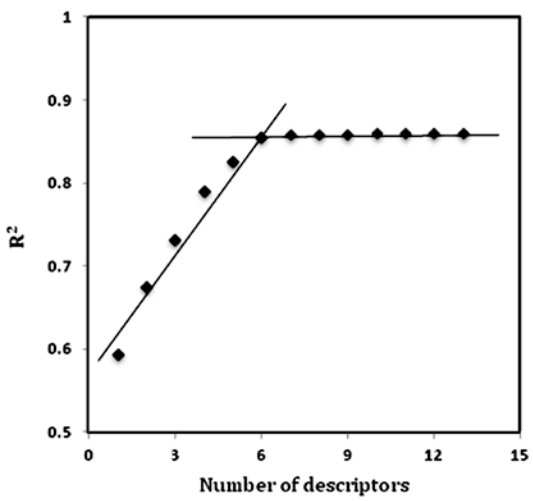

Figure 2. The variations of R^2 against the number of descriptors in the MLR model.

Table 2. Specifications of multiple linear regression model

Descriptor	Notation	Coefficient	*SE*	P-value
Moran autocorrelation –lag7 / weighted by atomic Sanderson electronegativities	MATS7e	200.130	±91.489	0.000
3D-MoRSE– signal 08 / unweighted	Mor08u	-206.385	±34.382	0.021
H autocorrelation of lag 0 / weighted by atomic masses	H0m	5.883	±2.462	0.000
Radial Distribution Function – 3.0 /weighted by atomic masses	RDF030m	147.196	±17.701	0.001
Ramification index	Ram	-6.994	±1.865	0.008
Eigenvector coefficient sum from adjacency matrix	VEA1	11.330	±4.043	0.000
Constant		-92.062	±19.659	0.034

As it is shown in this Figure, after addition of six descriptors no significant improvement in the R^2 was observed, therefore to prevent the over parameterization six descriptors were eventually selected.

The specifications of the MLR model that was developed by these six descriptors were shown in Table 2. Moreover the correlation matrix between these descriptors is shown in Table 3. As can be seen in this table there is not any high correlation between selected descriptors.

2.3 Diversity test

One of the most critical aspects in splitting of the data set is to warrant enough molecular diversity for it. In this study, diversity analysis was performed on the data set to make sure that the structures of the training, internal and external test sets can represent those of the whole ones [33].

Table 3. Correlation matrix between selected descriptors

	VEA1	Ram	RDF030m	H0m	Mor08u	MATS7e
VEA1	1					
Ram	0.514	1				
RDF030m	0.047	-0.130	1			
H0m	-0.079	0.202	0.013	1		
Mor08u	-0.164	-0.197	0.202	-0.574	1	
MATS7e	-0.056	0.356	0.046	-0.271	0.345	1

In this way, the mean distance of one chemical to the remaining ones was computed from descriptor space matrix as follows:

$$d_{ij} = \sqrt{\sum_{k=1}^{m}\left(x_{ij} - x_{jk}\right)^2} \quad (1)$$

$$\bar{d}_i = \frac{\sum_{i=1}^{n} d_{ij}}{n-1} \quad i = 1, 2, \ldots, n \quad (2)$$

where d_{ij} is a distance score for two different compounds and x_{ik} and x_{jk} are compounds descriptors. Then the calculated mean distances were normalized within the interval 0 to1, and plotted against the values of dependent variable (see Fig. 3). Inspection to this figure indicates that the structures of the compounds are diverse for both training and test sets. It warrants model stability and that the test set is suitable to assess the predictive performance of the developed models.

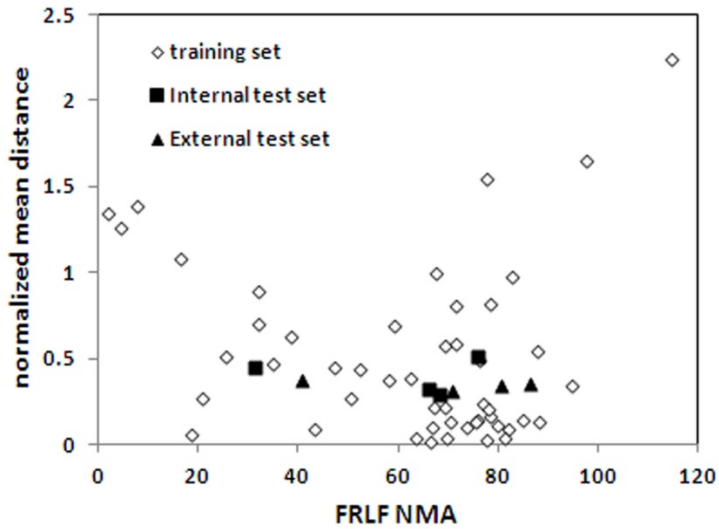

Figure 3. Scatter plot of normalized mean distance of samples versus experimental data.

2.4 Nonlinear model

Artificial neural networks are biologically inspired computer programs designed to simulate the way in which the human brain processes information [34]. ANNs are rapidly becoming the method of choice for structure-activity and structure-property relationship studies due to their capability to mapping complex patterns. In this work, in order to check any nonlinear relationships between structural descriptors and experimental values of FRLF NMA, multilayer perceptron neural network (MLP) [35] was applied by using STATISTICA (release. 7.1) software [36]. This network consisted of a layer of input units, a layer of hidden units and a layer of output unit. The node in each layer is connected to the nodes of the next layer by weights. The values of weights were optimized by Levenberg-Marquardt (LM) algorithm during training epochs. Then the trained network was used to calculate the FRLF NMA values of external test set as well as training and internal test sets.

3. Results and discussion

3.1. Linear modeling

In this work, quantitative relationships between FRLF NMA of some polymeric biomaterials and their structural descriptors were investigated by using linear and non-linear approaches. The best

MLR equation with DRAGON calculated molecular descriptors using the FRLF NMA as the dependent variable was derived by SPSS (ver. 17) software [37].

The specifications of the selected MLR model are shown in Table 2. As can be seen in this table six descriptors appeared in the models, which can encode different structural features of molecule. Among them MATS7e, is the lag 7 of Moran autocorrelation type descriptors, that was weighted by atomic Sanderson electronegativities and can encode electro-topological aspects of a molecule.H0m is belonged to GETAWAY descriptors series, which have shown great potential as powerful variables in QSAR modeling of different biological activities because they can encode information about molecular shape, size and atom distribution [38].

RDF030m obtained from radial basis functions centered on different interatomic distances (from 0.5A to 15.5A). These descriptors are based on the distance distribution in the geometrical representation of a molecule and constitute a radial distribution function code (RDF code) that shows certain characteristics in common with the 3D-MORSE (molecule representation of structures based on electron diffraction) code. Formally, the radial distribution function of an ensemble of atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius R.Mor08u is a 3D-MoRSE type that can be represent descriptor.

It is a representation of the 3D structure of a molecule and encodes structural features such as mass and amount of branching. The last descriptors are Ram and VEA1descriptors from topological series, which obtained from molecular graph (usually H-depleted) matrices and can account the molecular symmetry in terms of atom topological uniqueness.

These six descriptors can encode features of molecular structures such as size, shape and charge distribution that can affect the biological activities of the chemicals. Further explanations and meaning of these molecular descriptors, and their calculation procedures, can be found in the Handbook of molecular descriptors [39].The calculated FRLF NMA values by this model were shown in Table 1. The root mean square errors (RMSE) of this calculation were 12.6 and 15.9 for training and test sets, respectively.

3.2 Nonlinear modeling

In order to investigate any nonlinear relationship between selected molecular descriptors and FRLF NMA, a three-layer network with an exponential transfer function in the hidden layer and a logistic transfer function in the output layer was designed. The number of neurons in the hidden layer was optimized based on the minimizing of mean square errors. Then, the network was trained by Levenberg–Marquardt algorithm. In the next step, the developed 6-5-1 network was used to predict the cellular response for external test set as well as training and internal test sets.

The calculated values of FRLF NMA are shown in Table 1. The RMSE values of this estimation were 10.6, 12.6 and 10.7 for training, internal and external test sets, respectively. The plot of the predicted values of cellular response against the experimental values is shown in Fig 4.

Inspection to this figure indicates good correlation between experimental and calculated values of FRLF NMA (R^2_{train}=0.98, R^2_{int}=0.86 and R^2_{ext}=0.96). The residuals of predicted values are plotted against the experimental values in Fig 5. Random propagation of residuals over zero line indicates that there is no systematic error in developed model.

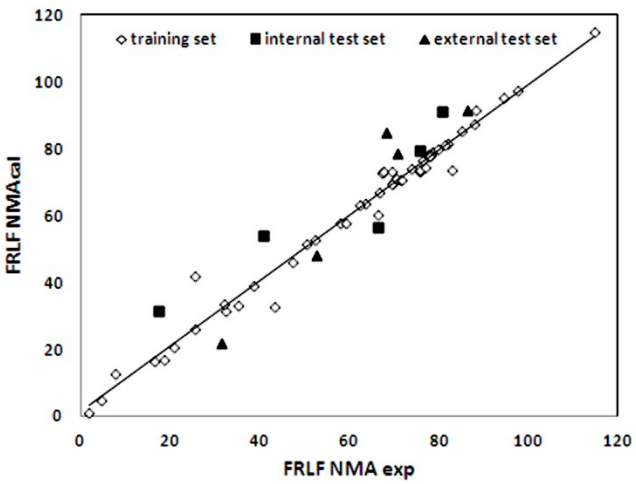

Figure 4. Calculated versus experimental activity values using ANN model.

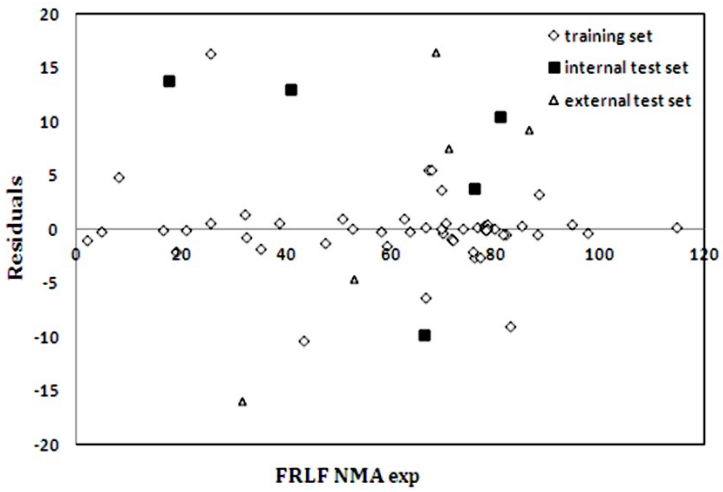

Figure 5. The plot of prediction residuals versus experimental values of FRLF NMA for all molecules in the data set.

Other statistical parameters of ANN and MLR models are shown in Table 4. Comparison between these values revealed that the ANN model produces better statistical results in terms of R^2, RMSE and P value over the MLR model, which strongly suggested a nonlinear relationship between the selected descriptors and FRLF NMA of these compounds. In order to evaluate the robustness and predictive power of models, the leave-six-out cross-validation test was performed and the values of the cross-validation correlation coefficient (Q^2) and standard deviation based on predicted residual sum of square (SPRESS) were calculated from the following equation:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

$$\text{SPRESS} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}} \quad (4)$$

In the above expressions $\bar{y}$is the mean of dependent variable, n is the number of observation and k is the number of independent variable.Q^2indicate the relative predictive ability of a model and SPRESS is the standardized predicted error sum of squares, which is a standard index to measure the accuracy of a modeling method. The obtained Q^2 and

SPRESS were 0.81 and 11.4, respectively, that indicated the robustness of developed ANN model. Descriptors that were used in this model are belonging to topological and electronic categories.

Sensitivity analysis approach is used to rank the general importance of the descriptors that appear in the model.

This approach is called sequential zeroing of weights (SZWs). The SZW method [40] estimates the degradation in output variables of a trained neural network when the weights connecting the ith input variable to the nodes of the hidden layer are set to zero. In this way, the contribution of that variable to the network response is excluded. The SZW method can only be implemented after a neural network is trained. Therefore, SZW results have to be validated based on the assumption that the ANN model is a sufficiently accurate model. According to this method, the measure that is calculated in order to reveal the importance of the ith input variable is the difference between the root mean square error (RMSE) of the complete network predictions and the RMSE obtained when the ith variable is excluded from the trained network (RMSEi), both being calculated on the same data set according to the following equation:

$$\mathrm{Rmdiff_i} = \mathrm{RMSE_i} - \mathrm{RMSE} \qquad (5)$$

Generally, the values of Rmdiffi are greater than zero. The variable with the greater importance is the one that leads to a greater value of Rmdiffi. According to the results of sensitivity analysis on ANN model, the importance order of descriptors was MATS7e > RAM > VEA1 >H0M >MOR08U > RDF030M. Appearing of these descriptors to the model indicate the effects of electronic and steric interaction on cellular responses to chemicals.

Table 4. Statistical parameters of ANN and MLR models

Model	Training			Internal test			External test			Validation			Q^2	SPRESS
	R^2	RMSE	F	R^2	RMSE	F	R^2	RMSE	F	R^2	RMSE	F		
MLR	0.85	12.6	42.1							0.75	15.9	2.048	0.74	13.4
ANN	0.98	10.6	1231.0	0.86	12.6	19.3	0.96	10.7	75.1				0.81	11.4

3.3 Applicability domain analysis

The domain of applicability is an important concept in quantitative structure activity relationships that allows one to estimate the uncertainty in the prediction of a particular molecule based on how similar it is to the compounds used to build the model [40].

A simple measure of a chemical being too far from the applicability domain of the model is its leverageh_i, which is defined as:

$$\mathrm{h_i} = \mathrm{x_i^T}(\mathrm{X^TX})^{-1}\mathrm{x_i}(\mathrm{i} = 1, \dots, \mathrm{n}) \qquad (6)$$

where x_i is the descriptor row-vector of the query compound and X is the n× k-1 matrix of k model descriptor values for n training set compounds. The superscript T refers to the transpose of the matrix/vector. The warning leverage h^* is calculated as follows:

$$h^*=3p/n \qquad (7)$$

where p is the number of model variable plus one, and n is the number of training compounds. The observation that a chemical has a leverage value greater than the warning leverage indicates that the

chemical falls outside the applicability domain. To visualize the applicability domain of a QSPR model, the standardized residuals versus leverage (Hat diagonal) values was plotted (Fig. 6).As it can be seen from this figure, all predictions were reliable except the number 38 and 49 in training set. The anomalous behavior of these compounds could be due to one of the following reasons: (1) incorrect experimental input data, (2) the descriptors selected do not capture some relevant structural features present in the molecule and absent in the others, (3) its biological mechanism is different from the remaining chemicals.

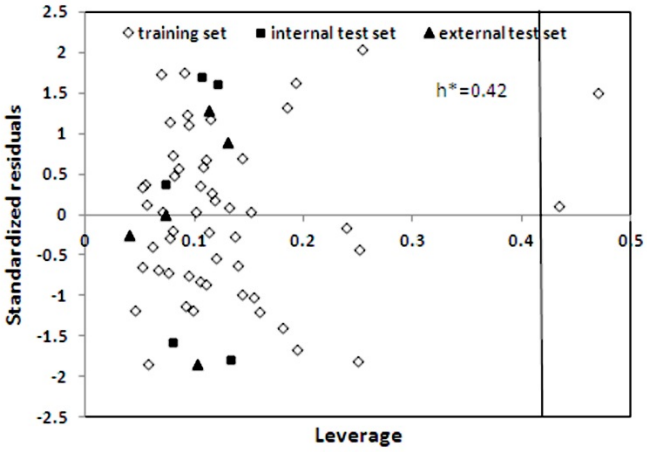

Figure 6. The plot of standardized residuals versus hat values, with a warning leverage of h*=0.42.

4. Conclusion

Multiple linear regression and artificial neural network are used as feature mapping techniques for prediction of FRLF cell response from their molecular structural descriptors. The superiority of ANN over MLR model indicates the dominant of nonlinear relation between selected molecular descriptors of interested polymers and fibroblast cell responses. The results of this study revealed that quantitative structure-activity relationship approach has a high applicability for accurate prediction of cellular response surface to tyrosine derived biodegradable polymers. Moreover the result of this study indicate that computational strategies that have demonstrated success in the rational design of new therapeutics in pharmaceutical discovery can be employed to offer guidance and direction in the design, selection, and optimization of novel bio-relevant materials.

5. References

[1] D. S. Kohane, R. Langer, Pediatr. Res., 63(2008)487.

[2] N. Huebsch, D. J. Mooney, Nature, 462(2009)426.

[3] W. You-Xiong, J. L. Robertson, B. William, J. Spillman, R. O. Claus, Pharm. Res., 21(2004)1362.

[4] L. S. Nair, C. T. Laurencin, Adv. Biochem. Eng. Biotechnol., 102(2006)47.

[5] A. V. Gubskaya, V. Kholodovyc, D. Knight, J. Kohn, W. J. Welsh, Polymer, 48(2007)5788.

[6] M. Yliperttula, B. G. Chung, A. Navaladi, A. Manbachi, A. Urtti, Eur. J. Pharm., Sci., 35(2008)151.

[7] V. Kholodovych, A. V. Gubskaya, M. Bohrer, N. Harris, D. Knight, J. Kohn, W. J. Welsh, Polymer, 49(2008)2435.

[8] P. A. Gunatillake, R. Adhikari, Eur. Cell. Mater., 5(2003)1.

[9] L. G. Griffith, Acta Mater., 48(2000)263.

[10] J. Kohn, W. J. Welsh, D. Knight, Biomaterials, 28(2007)4171.

[11] J. Kohn, Nat. Mater., 3(2004)745.

[12] Y. K. Peterson, X. S. Wang, P. J. Casey, A. Tropsha, J. Med. Chem., 52(2009)4210.

[13] L. Holm, K. Frech, B. Dzhambazov, R. Holmdahl, J. Kihlberg, A. Linusson, J. Med. Chem., 50(2007)2049.

[14] S. Oloff, R. B. Mailman, A. Tropsha, Med. Chem., 48(2005)7322.

[15] E. V. Kuz'min, A. G. Artemenko, E. N. Muratov, I. L. Volineckaya, V. A. Makarov, O. B. Riabova, P. Wutzler, M. Schmidtke, J. Med. Chem., 50(2007)4205.

[16] M. H. Fatemi, Anal. Chim. Acta., 556(2006)355.

[17] J. R. Smith, A. Seyda, N. Weber, D. Knight, S Abramson, J. Kohn, Macromol.Rapid.Commun.;25(2004)127.

[18] J. R. Smith, V. Kholodovych, D. Knight, J. Kohn, W. J. Welsh. Polymer, 46(2005)4296.

[19] J. R. Smith, D. Knight, J. Kohn, K. Rasheed, N. Weber, V. Kholodovych, W. J. Welsh, *J.* Chem. Inf. Comput. Sci., 44(2004)1088.

[20] P. M. Kou, N. Pallassana, R. Bowden, B. Cunningham, A. Joy, J. Kohn, J. E. Babensee, Biomaterials, 33(2012)1699.

[21] J. R. Smith, V. Kholodovych, D. Knight, W. J. Welsh, J. Kohn, QSAR Comb. Sci., 24(2005)99.

[22] A. D. Costache, J. Ghosh, D. D. Knight, J. Kohn, Adv. Eng. Mater., 12(2010)B3.

[23] C.G. Simon, Y. Yang, V. Thomas, S. M. Dorsey, A. W. Morgan, Com. Chem. High. T. Scr., 12(2009)544.

[24] A. Peters, D. M. Brey, J. A. Burdick, Tissue Eng. Part B: Reviews, 15(2009)225.

[25] D. Hofmann D, M. Entrialgo-Castaño, K. Kratz, A. Lendlein, Adv. Mater. 21(2009)3237.

[26] P. F. Holmes, M. Bohrer, J. Kohn, Prog. Polym. Sci., 33(2008)787.

[27] V. Kholodovych, J. R. Smith, D. Knight, S. Abramson, J. Kohn, W. J. Welsh, Polymer, 45(2004)7367.

[28] S. Brocchini, K. James, V. Tangpasuthadol , J. Kohn, J. Biomed. Mater. Res. 42(1998)66.

[29] S Brocchini, K. James, V. Tangpasuthadol, J. Kohn, J. Am. Chem. Soc., 119(1997)4553.

[30] HyperChem Release 7.0 for windows, Hypercube, Inc., 2002.

[31] The Dragon Website. http://www.disat.unimib.it/chem.

[32] N. Bingham, J. M. Fry, Regression: Linear models in statistics, Springer: London, UK, 2010.

[33] A. G. Maldonado, J. Doucet, M. Petitjean, B. T. Fan, Mol. Divers., 10(2006)39.

[34] S. A. Kustrin, R. Beresford. J. Pharm. Biomed. Anal., 22(2000)717.

[35] G. Daqi, J. Yan. Pattern Recognit., 38(2005)1469.

[36] The Statistica Website. http://www.statsoft.com/.

[37] SPSS for Windows, Rel. 17.0.0. 2008. Chicago: SPSS Inc.

[38] H. Kusic, B. Rasulev, D. Leszczynsk, J. Leszczynski, N. Koprivana, Chemosphere, 75(2009).1128.

[39] V. Consonni, R. Todeschini, Hand book of Molecular Descriptors. Wiley-Vch, Weinheim, 2000.

[40] L.I. Nord , S.P. Jacobsson, Chemomet. Intell. Lab. Sys. 44 (1998), 153.

[41] A. Tropsha, P. Gramatica, V.K. Gombar, QSAR Comb. Sci., 22(2003)69.