PROGRESS IN COLOR, COLORANTS AND COATINGS

available online @ www.pccc.icrc.ac.ir

*Prog. Color Colorants Coat. 9 (2016), 195-206*

Ministry of Science, Research and Technology
Institute for Color Science and Technology

# QSRR Study of Organic Dyes by Multiple Linear Regression Method Based on Genetic Algorithm (GA–MLR)

**D. Beiknejad*, M. J. Chaichi and M. H. Fatemi**

Faculty of Chemistry, University of Mazandaran, P. O. Box: 47416-95447, Babolsar, Iran.

ARTICLE INFO

Article history:
Received: 08 Jun 2016
Final Revised: 02 Jul 2016
Accepted: 05 Jul 2016
Available online: 05 Jul 2016

Keywords:
Disperse dyes
GA-MLR
Paper chromatography
Quantitative structure-retention relationship (QSRR)

## ABSTRACT

*Quantitative structure-retention relationships (QSRRs) are used to correlate paper chromatographic retention factors of disperse dyes with theoretical molecular descriptors. A data set of 23 compounds with known $R_F$ values was used. The genetic algorithm-multiple linear regression analysis (GA-MLR) with three selected theoretical descriptors was obtained. The stability and predictability of the model was validated by use of leave-one-out (LOO), leave-many-out (LMO) cross-validation, external validation, Y-randomization and applicability domain (AD) analysis. The GA-MLR revealed a statistically meaningful model showing the dependence of the RF value on sum of topological distances between N and Br atoms (T(N..Br)), global topological charge index (JGT) and R autocorrelation of lag 5 / unweighted (R5u_A) of the compounds. Prog. Color Colorants Coat. 9 (2016), 195-206© Institute for Color Science and Technology.*



## 1. Introduction

The rapid increase in the manufacture and consumption of synthetic dyes during the past century has been quite phenomenal. There are more than 10,000 commercially available dyes and it is estimated that more than 7 × $10^5$ tonnes of dyestuffs are produced per year [1]. At the present time nearly all industrial fields such as textiles, paper, rubber, plastics, dyestuffs, leather, ink, cosmetics, food, biomedicine, paint and varnishes are users of synthetic dyes [2-4]. They are also often added to a product to influence purchasing behaviors, but do not improve the product itself. The disposal of these dyes is an environmental concern because they are toxic to living organisms, they last for a long time in the ecosystem, cause allergies and skin or eye irritation, and are suspected human carcinogens [5-7]. On the other hand they are resistant against light and microbial attach. Therefore, as persistent organic compounds, synthetic dyes are not readily degradable and are typically not easily removable from water by conventional treatment methods.

Disperse dyes are extensively used in the textile industry for dying polyester, cellulose acetate and triacetate, polyamide, polylactide and acrylic fiber. More than 50% of disperse dyes are simple azo compounds, about 25% are anthraquinones and the rest are methine, nitro or naphthoquinone dyes [8]. The modeling of disperse dyes behavior has been studied by several authors. Oprea et al. proposed a model using

*Corresponding author: beiknejad@yahoo.com

quantitative structure-activity relationship method correlating disperse azo dyes affinity to cellulose fiber with variations in the chemical structure. Based on their study, long linear structures and substituents that maintain linearity are favored [9]. Hatch and Magee proposed a discriminant model for allergic contact dermatitis to anthraquinone disperse dyes. By using quantum chemical descriptors, they have shown that electrophilic reactivity of the anthraquinone disperse dyes is most unlikely and that some form of activation by electron transfer or photopromotion is responsible for protein reactivity and sensitization [10].

Although modern chromatographic methods along with modern data acquisition systems are well developed in chemical analysis [11, 12], a considerable section of the official methods of analysis still relies on paper chromatography (PC) [13]. Strips of paper saturated with color forming reagents permit rapid, inexpensive urinalysis [14]. On the other hand, methods based on paper chromatography are still being used in different areas including phytochemistry [15], radioanalytical chemistry [16, 17], and food chemistry [18].

PC is a powerful separation technique when considering cost, portability, flexibility in chromatographic system selection, speed, and extendibility in simultaneous parallel separations. In PC, excellent results are achievable with simple and inexpensive equipment. Advanced liquid chromatographic methods such as HPLC, suffer from limitations such as consumption of materials and time and the number of prior steps are often required to obtain the species of interest from the sample matrix. In this respect, PC is more satisfactory than HPLC. However, it is not possible to use PC in quantitative analysis, to the separation of complex mixtures and as a preparative technique. Furthermore, PC is the most analog method among chromatographic techniques. Like analog photography, in which the image is recorded as an analog on the film, in PC the separation is conducted on the strip or piece of paper. For efficient visualization of analytes, chemical reagents have been in use. Hazards associated with the use of particular visualization reagents, instability of visualization reagent and chromatogram after chemical or thermal treatment along with catalytic fading [19, 20] and hydrolysis of dyes [21], have stressed the need for the development of new methods in this field.

Quantitative structure-retention relationship (QSRR) method was envisaged to reduce the disadvantages associated with above said chromatographic procedures. Although the application of QSRR in chromatography is rooted in early investigations conducted by Martin [22], this discipline is still in continuous development. While most developed QSRR models concern more sophisticated separation techniques such as gas chromatography (GC) [23], high performance liquid chromatography (HPLC) [24] supercritical fluid chromatography (SFC) [25] and capillary electrophoresis (CE) [26], the development of QSRR techniques in planar chromatographic methods especially paper chromatography have been progressed rather slowly because of the presence of a considerable time gap between the prosperity periods of paper chromatography (1960s) and quantitative structure-retention relationships (after 1992).

There are limited studies that have used QSRR models to study planar chromatographic behaviors of chemicals. It is, thus, important to understand the correlation of paper chromatographic retention behavior of disperse dyes with their structural features represented as molecular descriptors. Inspired by the pioneering work of Levy [27] and in the continuation of our earlier works [23-26, 28, 29] we have revisited the work of Levy [27] to see if we can further develop a significant QSRR model using GA-MLR procedure with statistically significant set of parameters.

## 2. Experimental

### 2.1. Data set

The $R_F$ (retention factor) values of 23 disperse dyes as data set was chosen from the literature [30]. Please refer to Supplementary files section for details of the paper chromatographic conditions used for separation of dyes. The $R_F$ values were converted into the $R_M$ values using Bate-Smith and Westall equation [31]:

$$R_M = \log [(1/R_F) - 1)] \quad (1)$$

Different types of disperse dyes belonging to six classes including nitrodiphenylamine, azo, disazo, methine, naphthalimide and anthraquinone are presented in data set (Table 1). This data set was randomly divided into two groups: a training set consisting of 18 molecules that was used for the model generation and a test set of 5 compounds which was used for the evaluation of the QSRR model.

**Table 1:** The chemical structures of the data set molecules.

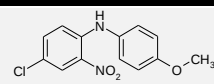
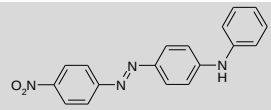
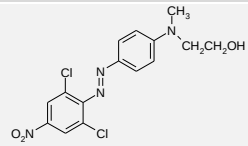
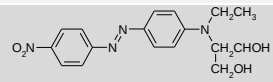
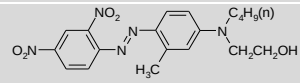
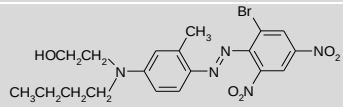
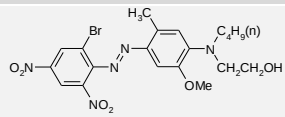
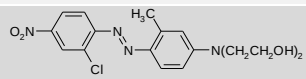
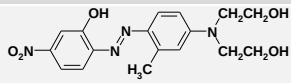
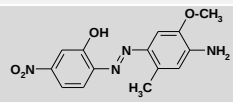
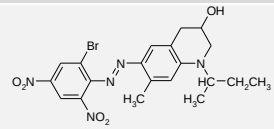
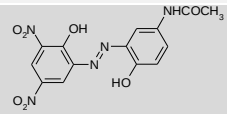
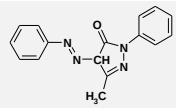
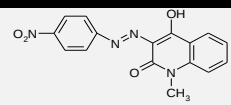
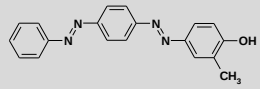
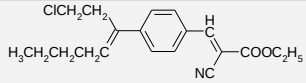
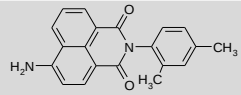
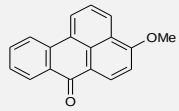
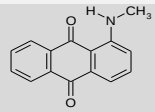
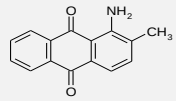
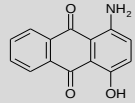
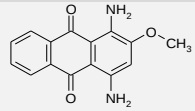
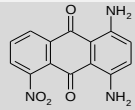
| No. | Commercial name | Structure | C.I. No.[a] | CAS Registry No. | Class[b] |
|---|---|---|---|---|---|
| 1 | Celliton fast yellow RR (Disperse orange 15) |  | 10350 | 6373-69-9 | I |
| 2 | Dispersol fast orange A (Disperse orange 1) |  | 11080 | 2581-69-3 | II |
| 3 | Celliton fast brown 3R (Disperse orange 5) |  | 11100 | 6232-56-0 | II |
| 4 | Artisil fast scarlet GP (Disperse red 2) |  | 11118 | 3769-58-2 | II |
| 5 | Celliton violet R (Disperse violet 13) |  | 11195 | 6374-02-3 | II |
| 6 | Celliton discharge violet B (Disperse violet 24) |  | 11200 | 6374-03-4 | II |
| 7 | Celliton discharge blue 3R |  | 11205 | | II |
| 8 | Celliton fast rubine 3B (Disperse red 21) |  | 11215 | 3769-57-1 | II |
| 9 | Celliton discharge rubine BBF (Disperse Red 16) |  | 11225 | 6253-14-1 | II |
| 10 | Cibacetscarlet G (Disperse Red 31) |  | 11250 | 6253-14-1 | II |
| 11 | Celliton discharge blue RRF (Disperse violet 7) |  | 11410 | 6486-13-1 | II |
| 12 | Celliton fast yellow G |  | 11835 | 3738-04-3 | II |
| 13 | Sudan yellow 3G (Disperse yellow 16) |  | 12700 | 4314-14-1 | II |

**Table 1:** Continued.

| No. | Commercial name | Structure | C.I. No. | CAS Registry No. | Class[(a)] |
|---|---|---|---|---|---|
| 14 | Celliton yellow 3GN (Disperse yellow 10) |  | 12795 | 8805-71-8 | II |
| 15 | Celliton fast yellow 5R (Disperse yellow 7) |  | 26090 | 6300-37-4 | III |
| 16 | Celliton fast yellow 7G "F" (Disperse yellow 31) |  | 48000 | 4361-84-6 | IV |
| 17 | Celliton brilliant yellow FFA-CF (Disperse yellow 11) |  | 56200 | 2478-20-8 | V |
| 18 | Duranol brilliant yellow 6G (Disperse yellow 13) |  | 58900 | 3688-79-7 | VI |
| 19 | Celliton pink R (Disperse red 9) |  | 60505 | 82-38-2 | VI |
| 20 | Celliton orange R (Disperse orange 11) |  | 60700 | 82-28-0 | VI |
| 21 | Duranol red 2B (Disperse red 15) |  | 60710 | 116-85-8 | VI |
| 22 | Duranol red X3B (Disperse red 11) |  | 62015 | 2872-48-2 | VI |
| 23 | Celliton fast violet B (Disperse violet 8) |  | 62030 | 82-33-7 | VI |

a) Color Index Number

b) I: nitrodiphenylamine, II: azo, III: disazo, IV: methine, V: naphthalimide, VI: anthraquinone

## 2.2. Molecular Modeling

All structures were drawn with the Hyperchem software (Ver. 7) [32] and pre-optimized with the molecular mechanics force field (MM+). The final geometries were obtained by re-optimization of the MM+ optimized structures by applying AM1 semi-empirical method in Hyperchem program [33, 29]. All calculations were carried out at the restricted Hartree-Fock level with no configuration interaction. The molecular structures were optimized using the Polak-Ribiere algorithm until the RMS gradient was 0.01 kcal $mol^{-1}$. The Hyperchem output files were used by the DRAGON program (Ver. 3) to calculate descriptors (1497 descriptors) [34]. Statistical investigation of the data and multivariate data analysis has been performed mainly by the QSARINS software (Ver. 2.2) [35].

## 2.3. Validation paremeters

The quality of GA-MLR model determined from a number of metrics including correlation coefficient ($R^2_{train}$), adjusted $R^2$ ($R^2_{adj}$), root-mean-squared error for the training set ($RMSE_{train}$), leave-one-out (LOO)

cross-validated correlation coefficient ($Q^2_{LOO}$), and leave-many-out (LMO) cross-validated correlation coefficient ($Q^2_{LMO}$), y-scrambling test ($R^2_{Yscr}$ and $Q^2_{Yscr}$), random response procedure ($R^2_{Yrnd}$ and $Q^2_{Yrnd}$), random descriptor procedure ($R^2_{Xrnd}$ and $Q^2_{Xrnd}$), external validation ($R^2_{ext}$), lack-of-fit **(**LOF**)** test, standard error of estimate (s) and variance ratio to judge the overall significance of the regression coefficients (F).

## 3. Results and discussions

There are 23 molecules in the data set and according to a rule of thumb, the models with maximum number of variables of 3 were investigated [26]. The optimum number of descriptors to be included in the model was determined by plotting $Q^2_{LOO}$ vs. lack of fitting (LOF) [37]. This plot helps in the selection of the models with the best compromise between high predictability (high $Q^2_{LOO}$) and small dimension (low LOF) (as, for example, the model which contains 3 descriptors i.e. *T(N..Br)*, *JGT* and *R5u_A* in the Figure 1).

Figure 2 shows the variation of the main model parameters ($Q^2$ and $R^2$) with number of descriptors. It is clear from this figure that in three numbers of factors we have maximum $Q^2$ and $R^2$ values.

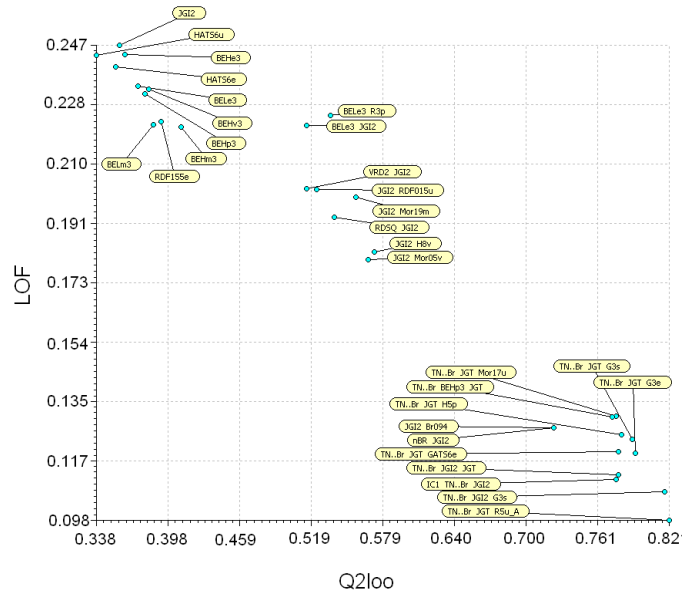



**Figure 1:** Plot of $Q^2$LOO vs. lack of fitting.

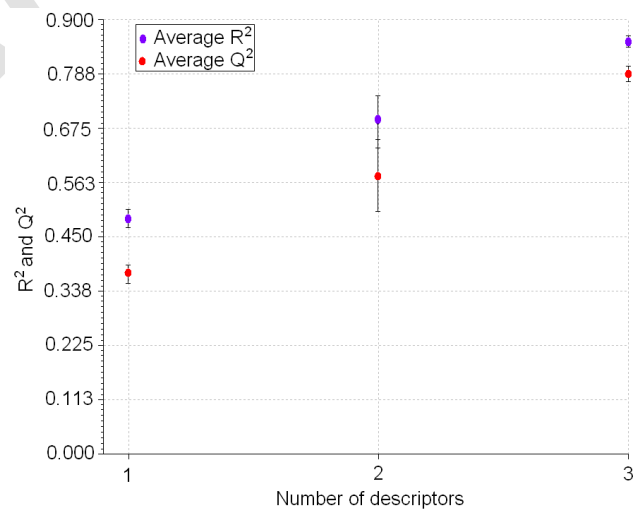



**Figure 2:** Effect of the number of descriptors on the main model parameters.

GA-MLR analyses were carried out, and the following model was generated:

$R_M$= 3.4754+ 0.0672 *T(N..Br)*-4.9555 *JGT* - 0.6058 *R5u_A* (2)

$N_{train}$= 18, $R^2_{train}$= 0.8793, $R^2_{adj}$= 0.8534, $RMSE_{train}$= 0.2090, s= 0.2370, *F*= 33.9865

$Q^2_{LOO}$=0.8213, $Q^2_{LMO}$= 0.7337, $R^2_{ext}$= 0.8653

The corresponding experimental and predicted values of the $R_M$ for all molecules studied in this work are listed in Table 2.

Table 3 represents the selected descriptors and their chemical meanings. The correlation matrix among these three descriptors is shown in Table 4. As shown in Table 4, the inter-correlation of the descriptors used in the GA-MLR model was low (below 0.53) which is in conformity with a statistically significant model. Figure 3 shows the predicted vs. experimental $R_M$ values plots using the Eq. (2). As can be observed, the ability of models to describe $R_M$ data was satisfactory.

**Table 2:** Data set and corresponding experimental and predicted values of $R_M$.

| No[a] | T(N..Br) | JGT | R5u_A | $R_F$ | $R_{M(exp)}$ | $R_{M(pred)}$ | Residual |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0.567 | 1.153 | 0.55 | -0.087 | -0.033 | 0.054 |
| 2 | 0 | 0.412 | 0.718 | 0.07 | 1.123 | 0.999 | -0.124 |
| 3* | 0 | 0.606 | 1.157 | 0.56 | -0.105 | -0.229 | -0.124 |
| 4 | 0 | 0.520 | 1.289 | 0.65 | -0.269 | 0.118 | 0.387 |
| 5 | 0 | 0.529 | 1.314 | 0.24 | 0.501 | 0.058 | -0.443 |
| 6* | 24 | 0.623 | 1.131 | 0.11 | 0.908 | 1.316 | 0.408 |
| 7 | 24 | 0.656 | 1.208 | 0.07 | 1.123 | 1.106 | -0.017 |
| 8 | 0 | 0.579 | 1.222 | 0.62 | -0.213 | -0.134 | 0.079 |
| 9 | 0 | 0.579 | 1.300 | 0.63 | -0.231 | -0.181 | 0.050 |
| 10* | 0 | 0.601 | 0.874 | 0.31 | 0.347 | -0.032 | -0.379 |
| 11 | 24 | 0.733 | 1.381 | 0.20 | 0.602 | 0.619 | 0.017 |
| 12 | 0 | 0.688 | 0.675 | 0.62 | -0.213 | -0.343 | -0.130 |
| 13 | 0 | 0.455 | 0.706 | 0.20 | 0.602 | 0.793 | 0.191 |
| 14 | 0 | 0.612 | 1.121 | 0.82 | -0.659 | -0.236 | 0.423 |
| 15 | 0 | 0.417 | 0.716 | 0.09 | 1.005 | 0.975 | -0.030 |
| 16 | 0 | 0.470 | 1.304 | 0.26 | 0.454 | 0.356 | -0.098 |
| 17 | 0 | 0.694 | 1.319 | 0.86 | -0.788 | -0.763 | 0.025 |
| 18 | 0 | 0.533 | 1.047 | 0.25 | 0.477 | 0.200 | -0.277 |
| 19 | 0 | 0.569 | 1.290 | 0.68 | -0.327 | -0.126 | 0.201 |
| 20* | 0 | 0.643 | 1.247 | 0.42 | 0.140 | -0.467 | -0.607 |
| 21 | 0 | 0.639 | 1.355 | 0.72 | -0.410 | -0.512 | -0.102 |
| 22 | 0 | 0.676 | 1.412 | 0.77 | -0.525 | -0.730 | -0.205 |
| 23* | 0 | 0.705 | 1.584 | 0.82 | -0.659 | -0.978 | -0.319 |

a) The numbers marked by an asterisk are test set.

**Table 3:** Descriptors used in the regressed GA-MLR model.

| Symbol | Definition | Class |
|---|---|---|
| *T(N..Br)* | Sum of topological distances between N..Br | Topological |
| *JGT* | Global topological charge index | Galvez topological charge indices |
| *R5u_A* | R autocorrelation of lag 5 /unweighted | GETAWAY |

**Table 4:** Correlation coefficient matrix for Eq. (2).

| No. | T(N..Br) | JGT | R5u_A |
|---|---|---|---|
| *T(N..Br)* | 1 | - | - |
| *JGT* | 0.365 | 1 | - |
| *R5u_A* | 0.136 | 0.530 | 1 |

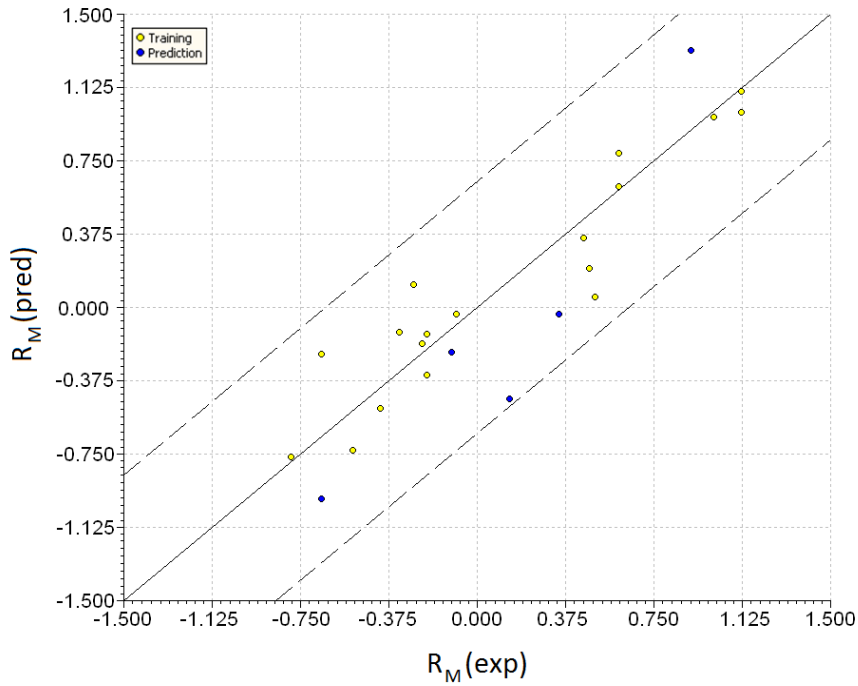



**Figure 3:** The linear relation between experimental and predicted $R_M$ values, the dotted lines indicate the 3σ interval.

The applicability domain (AD) of the model was evaluated by plotting standardized residuals vs. leverage (Hat) values (The Williams plot). As can be seen from Figure 4, all chemicals have cross validated standardized residuals lower than three standard deviation units (3σ) and leverages lower than the warning h* value of 0.667. Thus there is no outlier and structurally influential chemicals in the developed model.

Also the Y-scrambling test was applied to assess the presence of chance correlation in the developed model [37]. As can be seen from Figure 5, there is a significant difference in the quality ($R^2$ and $Q^2$ values) of the original model and that of model obtained with random responses (for 2000 iterations). Additional tests, including the scatter plots of leave-many-out (LMO) models vs. the original model, random response, and random descriptor procedures indicated adequate statistical quality of the model (see Supplementary files Figures S1–S4) [37]. $K_{xy}$ is correlation between the block of the modeling descriptors and the response.

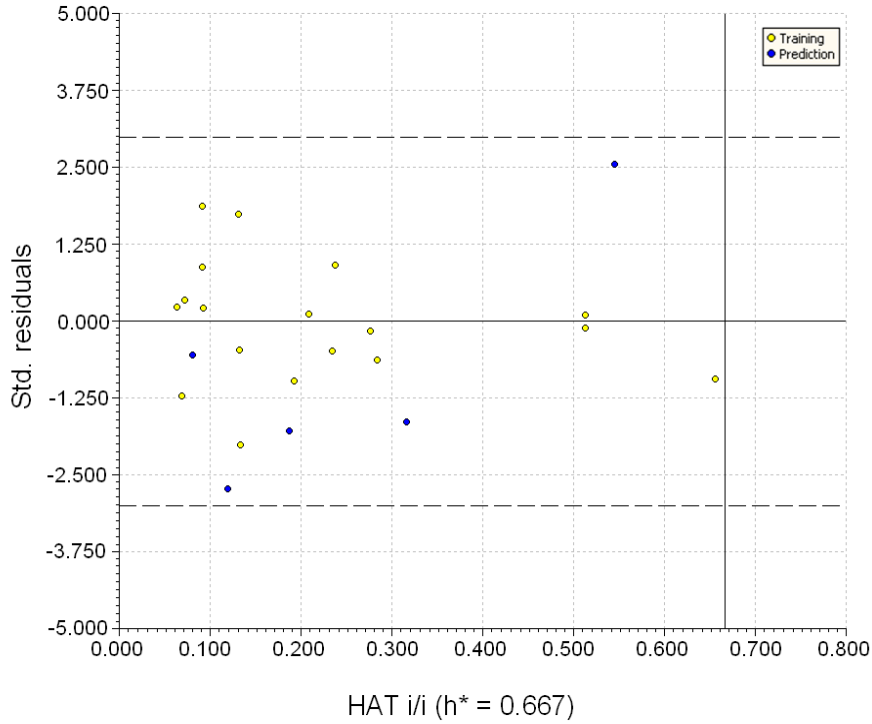



**Figure 4:** The Williams plot of the training and test sets (standardized residuals =3σ).

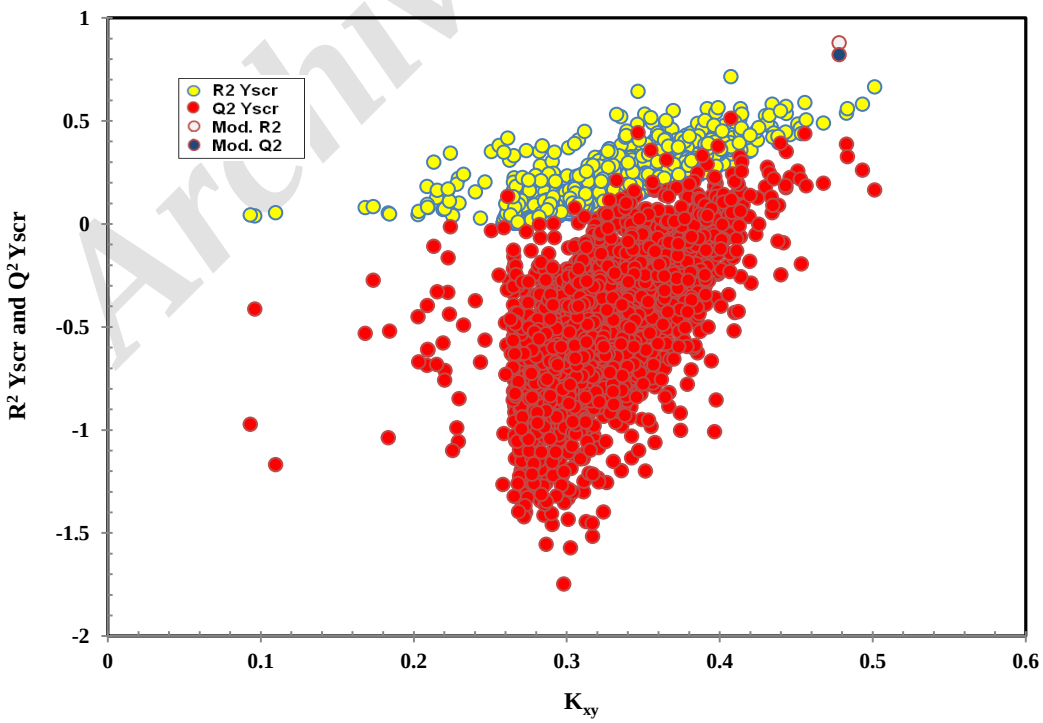



**Figure 5:** The plot of Y-scrambled models compared to the original model. Kxy is correlation between the block of the modeling descriptors and the response.

### 3.1. Interpretation of descriptors

The QSRR study revealed that *T(N..Br)* has positive contribution to the $R_M$ value, while *JGT* and *R5u_A* descriptors have negative contribution to the $R_M$ value. However, to the best of our knowledge, no reports have been published on the application of QSRR in paper chromatography. On the other hand, several investigations have been conducted to study the cellulose-dye interactions. Therefore the explanations and mechanisms in cellulose-dye interaction studies have been used to interpret the descriptors appeared in the regressed QSRR model.

Cellulose is formed by long chains of β-glucopyranose units connected one to another at the 1–4 positions. Stationary phase in PC is high-purity fibrous cellulose. The fibers are longer than those in TLC. Partition chromatographic mechanisms operate on cellulose surface even if adsorption effects cannot be excluded [38]. It has been shown that several types of interactions including: electrostatic (electronic) [39-41], specific binding centers due to the supramolecular structure of the cellulose [42], fragment composition [40] and steric effects [39, 43] have influence on the affinity of dyes for cellulose fiber.

In model, the topological descriptor *T(N..Br)* represents the sum of topological distances between nitrogen and bromine in those molecules that contain both of these atoms [44]. *T(N..Br)* is a 2D descriptor and has a positive coefficient, which advocates larger distances between N and Br atoms in a molecule for the higher $R_M$ value.

*JGT* is the global topological charge index which belongs to the Galvez topological charge indices. Topological descriptors help to differentiate the molecules according mostly to their size, degree of branching, flexibility and overall shape [45]. These indices describe charge transfer between pairs of atoms and therefore global charge transfer in a molecule. The GA-MLR analysis revealed a model showing the significant dependence of the $R_M$ value on global topological charge index of the disperse dyes. This descriptor has the negative sign, which indicates that an increase in the global charge transfer in a molecule, hence increasing the polarizability, leads to a decrease in its $R_M$ value [46]. *JGT* can be interpreted to contain a blending of size and electronic effects; which is in accordance with the literature for assessing cellulose-dye interactions [47].

The last descriptor is *R5u-A*. This R-GETAWAY descriptor has been proposed as chemical structure descriptors derived from the molecular influence matrix (MIM). They are calculated from the influence/distance matrix R where the elements of the MIM are combined with those of the geometry matrix [48]. *R5u_A* gives information about the presence of significant substituents in the molecule. Mantle atoms in the molecule result in higher values for *R5u_A* [49].

The hydrogen bonding between peripheral amino groups and parallel chains of cellulose are good candidates to give significant *R5u_A* values (Figure 6). The inspection of the values of this parameter indicate compounds 22 and 23 as presenting the highest *R5u_A*. This can be corroborated nicely with the detrimental presence of the amino groups in compounds 22 and 23.

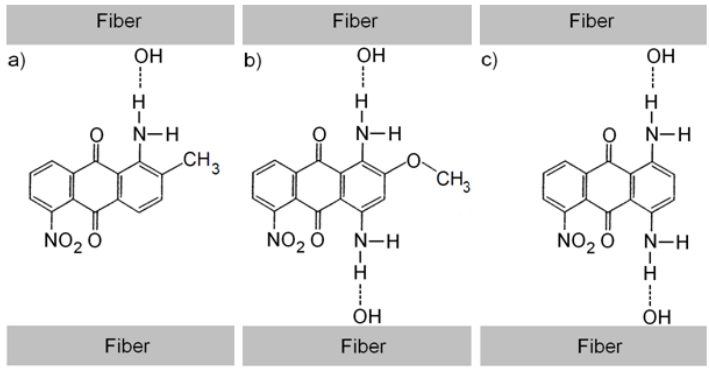



**Figure 6:** The proposed molecular mechanism for the interaction of compounds 20, 22 and 23 via hydrogen bond with cellulose.

The details of molecular mechanisms for the interaction of compounds 20, 22 and 23 with cellulose, are illustrated in Figure 6, where the dye molecule donates a lone pair of electrons to the hydrogen atom of a hydroxyl group of cellulose via one (compound 20) or two (compounds 22 and 23) amino functional groups. The compound 23 has two amino groups in 1 and 4-positions, which causes the compound 23 to have highest *R5u_A* value (Table 2 and Figure 6). On the other hand, the results obtained from solid state show that one hydrogen atom of the amino group is intramolecularly hydrogen bonded to the adjacent carbonyl oxygen atom. However it should be noted that we did not considered the intramolecular hydrogen bonds [50]. The lower $R_M$ values in compound 20 and 22 can be attributed to the presence of methyl and methoxy groups, which probably cause steric strain in the affinity of dye molecules for cellulose [39, 43].

## 4. Conclusion

In the present work, we have constructed a quantitative structure-retention relationship model, applying the genetic algorithm based multiple linear regression strategy, for a series of 23 disperse dyes tested by Sramek. The statistical quality of obtained GA-MLR with three theoretical descriptors was validated by use of leave-one-out cross-validation, leave-many-out cross-validation, external validation, Y-randomization and applicability domain analysis. The analysis of appeared theoretical descriptors provided some useful structural information which was associated with paper chromatic retention behavior of disperse dyes. The simultaneous appearance of *T(N..Br)*, *JGT* and *R5u_A* reveals the importance of the topological and GETAWAY descriptors in the chromatographic behavior of disperse dyes, thus allowing us to use such regressed model in predicting retention behavior of other disperse dyes not included in the study. Moreover, due to the similar mode of action in both paper chromatography and fabric dyeing, the developed model has the potential to assist in better understanding the dye-fabric interaction in order to reduce the unfixed dye concentrations in the dye bath after dying to consider environmental and economical aspects.

**Acknowledgment**

The authors thank Dr P. Gramatica who provided the QSARINS software.

## 5. References

1. K. S. Bharathi, S. T. Ramesh, Removal of dyes using agricultural waste as low-cost adsorbents: a review, *Appl. Water Sci.,* 3(2013) 773-790.
2. T. Robinson, P. Chandra, P. Nigam, Removal of dyes from a synthetic textile dye effluent by biosorption on apple pomace and wheat straw, *Water Research,* 36(2002) 2824-2830.
3. T. S. Anirudhan, P. S. Suchithra, P. G. Radhakrishnan, Synthesis and characterization of humic acid immobilized-polymer/bentonite composites and their ability to adsorb basic dyes from aqueous solutions, *Appl. Clay Sci.,* 43(2009) 336-342.
4. H. Tavakkoli, D. Beiknejad, T. Tabari, Fabrication of perovskite-type oxide La0.5Ca0. $5CoO_3$- δ nanoparticles and its dye removal performance, *Desalin. water Treat.,* 52(2014), 7377-7388.
5. K. T. Chung, G. E. Fulk, A. W. Andrews, Mutagenicity testing of some commonly used dyes, *Appl. Environ. Microbiol.,* 42(1981) 641-648.
6. H. M. Pinherio, E. Touraud, O. Tomas, Aromatic amines from azo dye reduction: status review with emphasis on direct UV spectrophotometric detection in textile industry wastewaters, *Dyes Pigment.,* 61(2004) 121-139.
7. N. M. Mahmoodi, S. Soltani-Gordefaramarzi, Dye Removal from Single and Quaternary Systems Using Surface Modified Nanoparticles: Isotherm and Kinetics Studies, *Prog. Color Colorants Coat,.* 9(2016), 85-97.
8. M. Clark (ed.), Handbook of textile and industrial dyeing, Woodhead Publishing, Cambridge, 2011.
9. T. I. Oprea, L. Kurunczi, S. Timofei, QSAR studies of disperseazo dyes. Towards the negation of the pharmacophore theory of dye-fiber interaction?, *Dyes Pigment.,* 17(1997), 41-64.

10. K. L. Hatch, P. S. Magee, A discriminant model for allergic contact dermatitis in anthraquinonedisperse dyes, *Quant. Struct.-Act. Relat.*, 17 (1998) 20-26.
11. L. M. L. Nollet (ed.) Chromatographic Analysis of the Environment, Taylor& Francis Group, Boca Raton, 2006.
12. A. Sarafraz-Yazdi, D. Beiknejad, Z. Es'haghi, LC Determination of mono-substituted phenols in water using liquid–liquid–liquid phase microextraction, *Chromatographia*, 62(2005) 49-54.
13. AOAC International. Official Methods of Analysis, 16th edition; 4th revision, version CD-ROM; AOAC International: Gaithersburg, MD, 1998. Methods: 930.38, 961.20, 963.16, 964.26, 967.17, 969.30, 969.52, 970.52 and 990.25.
14. M. B. Lyne, in: Kirk-Othmer Encyclopedia of Chemical Technology, Wiley, New York, 1993.
15. U. Hachuła, S. Anikiel, M. Połowniak, Determination of gallic acid after thin-layer chromatographic and paper chromatographic separations, *J. Planar Chromatogr.,* 17(2004), 51-53.
16. R. Chakravarty, A. Dash, M. Venkatesh, A novel electrochemical technique for the production of clinical grade 99m Tc using (n, γ) 99 Mo. *Nucl. Med. Biol.*, 37(2010) 21-28.
17. U. Pandey, P.S. Dhami, P. Jagesia, M. Venkatesh, M. R. A. Pillai, Extraction paper chromatography technique for the radionuclidic purity evaluation of 90Y for clinical use, *Anal. Chem.*, 80(2008) 801-807.
18. A. Gáspár, I. Bácsi, Forced flow paper chromatography: A simple tool for separations in short time, *Microchem. J.*, 92(2009) 83-86.
19. P. Doll, F. Shi, S. Kelly, W. Wnek, The problem of catalytic fading with ink-jet inks, in Proceedings of the International Conference on Digital Printing Technologies, NIP14, Toronto, Ontario, Canada, (1998) 118-121.
20. Y. Okada, T. Hihara, Z. Morita, Analysis of the catalytic fading of pyridone-azo disperse dyes on polyester using the semi-empirical, molecular orbital PM5 method, *Dyes Pigment.*, 78(2008), 179-198.
21. H. P. Mehta, A. T. Peters, M. S. Wild, Problems arising during the chromatographic purification of some monoazo and styryl disperse dyes, *J. Chromatogr.*, 77(1973), 455-457.
22. A. J. P. Martin, Some theoretical aspects of partition chromatography, *Biochem. Soc. Symp.*, 3(1950), 4-20.
23. M. H. Fatemi, H. Malekzadeh, In-silico prediction of gas chromatographic retention indices of some terpenols, *J. Sep. Sci.*, 35(2012), 2088-2094.
24. M. H. Fatemi, H. Shamseddin, Prediction of immobilized artificial membrane-liquid chromatography retention of some drugs from their molecular structure descriptors and LFER parameters, *J. Sep. Sci.* 32(2009), 3395-3402.
25. M. H. Fatemi, H. Malekzadeh, H. Shamseddin, Prediction of supercritical fluid chromatographic retention factors at different percents of organic modifiers in mobile phase, *J. Sep. Sci.*, 32(2009), 653-659.
26. M. H. Fatemi, H. Shamseddin, H. Malekzadeh, Quantitative structure migration relationship modeling of migration factor for some benzene derivatives in micellarelectrokinetic chromatography, *J. Sep. Sci.*, 32(2009), 1934-1940.
27. F. Levy, An Approach to the Correlation of R Value with Structure in the Paper Chromatography of Carbohydrate Compounds, *Anal. Chem.*, 26(1954), 1849-1850.
28. T. Tabari, H. Tavakkoli, P. Zargaran, D. Beiknejad, Fabrication of perovskite-type oxide BaPbO3 nanoparticles and their efficiency in photodegradation of methylene blue, *S. Afr. J. Chem.*, 65(2012), 239-244.
29. D. Beiknejad, M. J. Chaichi, Estimation of photolysis half-lives of dyes in a continuous-flow system with the aid of quantitative structure-property relationship, *Front. Environ. Sci. Eng.*, 8(2014), 683-692.
30. J. Sramek, Paper chromatography of dyes. I. Paper chromatography of disperse dyes, J. Chromatogr., 9 (1962) 476-484.
31. R. Singh, A. Meena, A. S. Negi, K. Shanker, Quantitative relationships between molecular descriptors, chromatographic retention behavior, and in vitro antituberculosis activity of phytol derivatives, *J. Planar Chromatogr.,* 25(2012), 10-18.
32. Hyperchem Program Release 7 for Windows; Hypercube, Inc.: Gainesville, FL, 2002.
33. M. J. S. E. Dewar, G. Zoebisch, E. F. Healy, J. J. P. Stewart, Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, *J. Am. Chem. Soc.*, 107(1985), 3902-3909.
34. DRAGON ver. 3.0, TALETE srl, Italy; http://www.talete.mi.it.
35. QSAR research unit, Department of theoretical and applied sciences, University of Insubria, Italy; http://www.qsar.it.
36. M. Khoshneviszadeh, N. Edraki, R. Miri, A. Foroumadi, B. Hemmateenejad, QSAR study of 4-aryl-4H-chromenes as a new series apoptosis inducers using different chemometrics tools, *Chem. Biol. Drug Des.*, 79(2012), 442-458.
37. P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS: a new software for the development, analysis, and validation of QSAR

MLR models, *J. Comput. Chem.*, 34(2013), 2121-2132.
38. Cazes (ed.), Encyclopedia of Chromatography. Taylor and Francis Group, LLC , 2010, p. 2199.
39. V. G. Agnihotri, C. H. Giles, The cellulose–dye adsorption process. A study by the monolayer method, *J. Chem. Soc., Perkin Trans.*, 2(1972), 2241-2246.
40. N. I. Zhokhova, I. I. Baskin, V. A. Palyulin, A. N. Zefirov, N. S. Zefirov, A Study of the Affinity of Dyes for Cellulose Fiber within the Framework of a Fragment Approach in QSPR, *Russ. J. Appl. Chem.*, 78(2005), 1013-1017.
41. S. Timofei, L. Kurunczi, T. Suzuki, W.M.F. Fabian, S. Muresan, linear regression (MLR) and neural network (NN) calculations of some disazo dye adsorption on cellulose, *Dyes Pigment.*, 34(1997), 181-193.
42. A. D. French, O. A. Battisla, J. A. Cuculo, D. G. Gray, Kirk-Othmer Encyclopedia of Chemical Technology, Wiley, New York, 1993.
43. J. Polanski, R. Gieleciak, M. Wyszomirski, Comparative molecular surface analysis (CoMSA) for modeling dye-fiber affinities of the azo and anthraquinone dyes, *J. Chem. Inf. Comp. Sci.*, 43(2003), 1754-1762.
44. R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley–VCH, Weinheim, 2000.
45. R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics, Wiley-VCH, New York, USA, 2009.
46. A. Bajpai, N. Agarwal, S. P. Gupta, A comparative 2D QSAR study on a series of hydroxamic acid-based histone deacetylase inhibitors vis-à-vis comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), *Indian J. Biochem. Biophys.*, 51(2014), 244-252.
47. S. Timofei, W. Schmidt, L. Kurunczi, Z. Simon, A review of QSAR for dye affinity for cellulose fibres, *Dyes Pigment.*, 47(2000), 5-16.
48. V. Consonni, R. Todeschini, M. Pavan, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors, *J. Chem. Inf. Comput. Sci.*, 42(2002), 682-692.
49. S. Funar-Timofei, W. M. F. Fabian, L. Kurunczi, M. Goodarzi, S.T. Ali, Y.V. Heyden, Modelling heterocyclic azo dye affinities for cellulose fibres by computational approaches, *Dyes Pigment.*, 94(2012), 278-289.
50. J. O. Morley, Molecular orbital studies of aminoanthraquinones, *J. Chem. Soc., Perkin Trans.*, 2(1972), 1223-1229.

How to cite this article:
D. Beiknejad, M. J. Chaichi and M. H. Fatemi, QSRR Study of Organic Dyes by Multiple Linear Regression Method Based on Genetic Algorithm (GA–MLR), Prog. Color Colorants Coat., 9 (2016) 195-206.

PROGRESS IN COLOR, COLORANTS AND COATINGS

available online @ www.pccc.icrc.ac.ir

*Prog. Color Colorants Coat. 9 (2016), 217-222*

Ministry of Science, Research and Technology
Institute for Color Science and Technology

# توصیف ساختار مولکولی مواد رنگزای آلی به منظور پیش‌بینی رفتار بازداری آنها در کروماتوگرافی کاغذی به روش روابط کمی ساختار- بازداری

داود بیک نژاد[۱*]، محمد جواد چایچی[۲]، محمد حسین فاطمی[۳]

[۱] دانش آموخته‌ی مقطع دکترای شیمی تجزیه، دانشکده شیمی، دانشگاه مازندران، بابلسر، ایران، صندوق پستی ۹۵۴۴۷-۴۷۴۱۶

[۲] دانشیارگروه شیمی تجزیه، دانشکده شیمی، دانشگاه مازندران، بابلسر، ایران، صندوق پستی ۹۵۴۴۷-۴۷۴۱۶

[۳] استاد گروه شیمی تجزیه، دانشکده شیمی، دانشگاه مازندران، بابلسر، ایران، صندوق پستی ۹۵۴۴۷-۴۷۴۱۶

## چکیده

از روابط کمی ساختار- بازداری (QSRRs) برای ارتباط رفتار بازداری مواد رنگزای دیسپرس در کروماتوگرافی کاغذی با توصیف‌کننده‌های مولکولی استفاده شد. سری داده‌ها شامل ۲۳ ترکیب با مقدار RF معلوم بودند. با استفاده از روش الگوریتم ژنتیک و رگرسیون خطی چندگانه (GA-MLR) مدلی حاوی سه توصیف‌کننده بدست آمد. کیفیت آماری مدل با استفاده از روش‌های ارزیابی تقاطعی، ارزیابی خارجی، درهم ریختگیY و بررسی دامنه‌ی قابل کاربرد، مورد ارزیابی قرار گرفت. تحلیل توصیف‌کننده‌های ظاهرشده در مدل حاکی از وابستگی مقادیر RF به مجموع فواصل توپولوژیک بین اتم‌های نیتروژن و برم ((T(N..Br))، شاخص بار توپولوژیک کلی (JGT) و ضریب خودهمبستگی گروه R با فاصله‌ی ۵ / وزن‌دار نشده (R5u_A) در مواد رنگزای مطالعه شده می‌باشد.

اطلاعات مقاله

**تاریخچه مقاله:**

تاریخ دریافت : ۱۴ بهمن ۱۳۹۴

تاریخ دریافت آخرین اصلاحات: ۱۲ تیر ۱۳۹۵

تاریخ پذیرش: ۱۵ تیر ۱۳۹۵

در دسترس به صورت الکترونیکی از: ۱۵ تیر ۱۳۹۵

**واژه های کلیدی:**

رنگ‌های دیسپرس

الگوریتم ژنتیک

رگرسیون خطی چندگانه (GA-MLR)

کروماتوگرافی کاغذی

روابط کمی ساختار- بازداری (QSRRs)

*Corresponding author: meshahidizandi@gmail.com.