Chemometrics and Intelligent Laboratory Systems 134 (2014) 1–9

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Quantitative structure–affinity relationship study of azo dyes for cellulose fibers by multiple linear regression and artificial neural network

Xiaojun Wang [a], Yongyuan Sun [a], Lei Wu [a], Shaojin Gu [a], Ruina Liu [a], Li Liu [a], Xin Liu [a], Jie Xu [a,b,*]

[a] College of Materials Science & Engineering, Wuhan Textile University, 430200 Wuhan, China
[b] State Key Laboratory of New Textile Materials & Advanced Processing Technology, Wuhan Textile University, 430200 Wuhan, China

ARTICLE INFO

Article history:
Received 12 November 2012
Received in revised form 23 January 2014
Accepted 2 March 2014
Available online 11 March 2014

Keywords:
QSPR
Affinity
Azo dyes
Cellulose fiber
Multiple linear regression
Artificial neural network

ABSTRACT

A quantitative structure–property relationship study was performed to correlate descriptors representing the molecular structures to fiber affinities for azo dyes. The complete set of 51 compounds was divided into a training set of 41 compounds and a test set of 10 compounds by using DUPLEX algorithm. Multiple linear regression analysis was used to select the best subset of descriptors and to build linear models; nonlinear models were developed by means of artificial neural network. The robustness of the obtained models was assessed by different approaches, including leave-many-out cross-validation, Y-randomization test, and external validation through test set. The obtained models with four descriptors show good predictive power: for the test set, a squared correlation coefficient (r^2) of 0.916 was achieved by the linear model; while the nonlinear model with r^2 of 0.935 for the test set performs better than the linear model. Furthermore, the applicability domain of the models was analyzed based on the Williams plot. The donor atom number for H-bonds, group polarizability and electronegativity of the dye molecules are found to play important roles for the dye-fiber affinity.

1. Introduction

Dyes are widely used in textile dyeing, paper printing, color photography, pharmaceutical, food, cosmetics, and other industries [1–3]. The history of dyes can be traced back to 2600 BC when there was the earliest written record of the use of dyes in China [4]. Originally, the dyes were obtained naturally until William Henry Perkin prepared the first synthetic dye pigment "Mauve" from coal-tar chemicals in 1856 [5]. Since the first azo dye spliced onto the fabric by coupling in 1875, thousands of organic chromogens were produced by the dye industry [4]. For example, in 1995, 90% of the 3000 compounds registered in the Color Index were used at the level of at least 100 t per year [6]. Currently, the world market of dyes and pigments is estimated to be around 1.3 million tons, valued at US$16 billion. With respect to both number and production volume, azo dyes are the largest group of colorants constituting 70% of all organic dyes produced in the world [7].

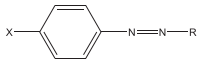
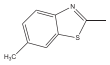
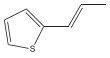
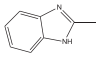
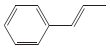
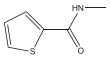
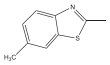
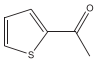
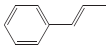
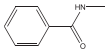
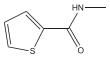
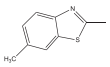
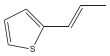
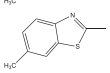
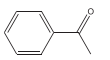
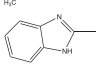
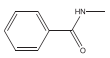
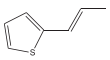
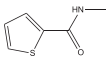
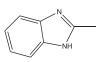
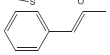
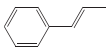
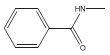
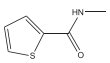
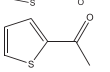
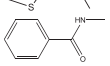
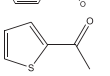
Cotton/cellulose fiber is a tremendous natural resource that has broad application in various productions including the textile industry. Cellulose is an isotactic β-1,4-polyacetal of cellobiose, in which two glucose molecules are linked together via a glycosidic bond. Hydrogen bond network between the polymer components is formed, involving hydroxyl groups and ether oxygen (glycoside oxygen). Cellulose does not contain ionic sites, and as such is not able to fixate cationic or anionic dyes through Coulomb forces [8]. Thus, direct dyeing of cotton requires specific features of the dye molecules to result in an overall affinity for the cellulose fiber. For a successful uptake from aqueous solution and stable fixation on the textile fiber, the dye molecules with hydrogen bond donor or acceptor sites are adsorbed in intermicellar cavities, where they are thought to form large aggregates and suitable dye–fiber interaction. The interaction between the dye molecule and the cellulose fiber is affected by various factors (electrostatic fields, Van der Waals forces, formation of hydrogen bonds, hydrophobicity, etc.), and the available experimental data are rather diverse. That is why theoretical methods for calculation of quantitative structure–property/activity relationships (QSPR/QSAR) are widely used to study these interactions [8–14]. The QSPR/QSAR approach is based on the assumption that the variation of the behavior of the compounds, as expressed by any measured physicochemical properties, can be correlated with numerical changes in structural features of all compounds [15–19]. The advantage of this approach lies in the fact that it requires only the knowledge of the chemical structure and is not dependent on any experimental properties. Once a correlation is established and validated, it can help to synthesize new compounds with high affinity. Classical quantitative structure–affinity relationships have been reported in the literature for the dye affinities on the cellulose fiber [9,10]. Zhokhova et al. [10] developed a multiple linear regression (MLR) model with the squared correlation coefficient R^2 of 0.971 and the cross-validated R^2 (Q^2) of 0.949 for the affinities of azo dyes based fragment descriptors. However, this model has not been evaluated with the external test set. In fact, validation is a crucial aspect

* Corresponding author at: State Key Laboratory of New Textile Materials & Advanced Processing Technology, Wuhan Textile University, 430200 Wuhan, China.
E-mail address: xujie0@ustc.edu (J. Xu).

http://dx.doi.org/10.1016/j.chemolab.2014.03.001
0169-7439/

of any QSPR/QSAR [20]. Based on the hypothesis of specific dye–fiber interactions and as an alternative to classical QSPR studies, Comparative Molecular Field Analysis (CoMFA) [8,11] and Comparative Molecular Surface Analysis (CoMSA) [12,13] were also used to predict dye adsorption properties. Schüürmann et al. [8] established models for the azo dye-fiber affinities using CoMFA without external validation.

Table 1
Experimental and calculated dye affinity (A, kJ mol^{-1}) of azo dyes.

No.	X	R[a]	Expt. A	Cacl. A		No.	X	R[a]	Expt. A	Cacl. A	
				MLR	ANN					MLR	ANN
											
1		γ_b	22.26	21.05	22.03	27		C	9.45	9.20	9.61
2		γ_b	15.69	14.72	15.04	28[b]		B	9.2	7.92	7.74
3		γ_a	14.35	15.24	15.39	29		C	9.03	8.36	8.61
4		E	9.62	12.09	10.28	30		A	8.78	8.41	8.32
5		E	8.79	7.57	7.37	31		C	8.4	10.05	10.48
6		γ_b	13.18	14.13	12.55	32		C	8.28	8.41	8.76
7[b]		γ_b	10.92	12.33	11.30	33		B	7.15	8.15	8.32
8[b]		C	14.48	15.31	16.52	34		D	7.06	6.02	5.88
9		D	10.5	11.49	9.79	35		A	7.02	7.82	7.45
10[b]		D	7.7	7.60	7.58	36		B	6.52	7.05	6.93
11		E	5.23	6.58	6.32	37		E	6.27	5.99	5.69
12		γ_b	8.58	10.14	9.53	38		D	6.23	5.18	5.16
13		C	13.56	10.91	11.90	39[b]		D	6.02	6.87	6.72
14		E	4.48	4.39	3.87	40		E	5.81	6.19	5.81
15		E	4.6	4.39	3.87	41		D	5.18	5.23	5.22
16[b]		D	3.59	6.10	5.95	42		E	5.1	5.27	5.12
17		D	1.92	2.36	2.16	43		C	4.64	6.07	6.00
18		D	2.97	3.91	3.56	44[b]		E	4.26	5.19	5.04
19		C	9.49	9.28	9.28	45		B	4.22	4.20	4.18
20		C	7.24	7.60	7.29	46		C	4.1	6.00	5.82
21		C	6.61	6.05	5.42	47		B	4.05	4.13	4.06
22[b]		A	15.8	13.73	14.72	48		D	3.85	2.26	3.30
23		A	14.25	13.17	13.73	49		D	3.43	2.83	3.54
24[b]		A	13.08	12.36	13.21	50		E	3.22	2.34	3.28
25		A	12	10.10	10.73	51[b]		E	2.84	2.27	3.26
26		B	9.66	8.48	8.40						