ELSEVIER

Dyes and Pigments 46 (2000) 101–107

DYES and PIGMENTS

The prediction of the substantivity of anionic, water soluble dyes on a biomass using neural networks

A.J. Greaves *

Department of Colour Chemistry, University of Leeds, Leeds LS2 9JT, UK

Received 21 March 2000; received in revised form 26 April 2000; accepted 6 May 2000

Abstract

The substantivity of series of anionic, water soluble dyes for a biomass has been measured. Using feed-forward back propagation neural networks and relatively simple empirical and semi-empirical molecular descriptors as input variables, the correlation between the predicted substantivity and the observed substantivity was determined. Depending on the molecular descriptors used as input variables and the number of hidden layers in the neural network the correlation between the predicted substantivity and the observed substantivity was 0.80 and 0.76.

Keywords: Substantivity; Dyes; Biomass; Neural networks; Correlation

1. Introduction

Neural networks have seen an explosion of interest over the last few years, and are being successfully applied across an extraordinary range of problem domains, in areas as diverse as finance, medicine, engineering, physics and colour chemistry [1]. Indeed, anywhere that there are problems of prediction, classification or control, neural networks are being introduced. This sweeping success can be attributed to their power and ease of use.

Neural networks are sophisticated modelling techniques, capable of modelling extremely complex functions and are also intuitively appealing, based as they are, on a crude low-level model of biological neural systems. The human brain is principally composed of circa 10^9 neurons, massively interconnected (with an average of several thousand interconnects per neuron). Each neuron is a specialised cell, which can propagate an electro-chemical signal. Thus, from a very large number of extremely simple processing units (each performing a weighted sum of its inputs and then firing a binary signal if the total input exceeds a certain level) the brain manages to perform extremely complex tasks. Artificial neural networks can achieve remarkable results using a model not much more complex than this.

A simple neural network has a feed-forward structure in which signals flow from inputs, forwards through any hidden units, eventually reaching the output units. A typical feed-forward network is shown in Fig. 1.

When the network is executed (used), the input variable values are placed in the input units and

* Tel.: +44-113-233-2811; fax: +44-113-233-2947.
E-mail address: a.greaves@leeds.ac.uk .

0143-7208/00/$ - see front matter
PII: S0143-7208(00)00041-3

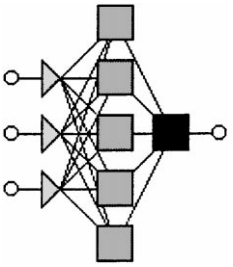

Fig. 1. A typical feed-forward network with an architecture of 3 input:5 hidden:1 output.

then the hidden and output layer units are progressively executed. Each of them calculates its activation value by taking the weighted sum of the outputs of the units in the preceding layer, and subtracting the threshold. The so-called activation value is passed through the activation function to produce the output of the neuron. When the entire network has been executed, the outputs of the output layer act as the output of the entire network.

The network thus has a simple interpretation as a form of input–output model, with the weights and thresholds (biases) the free parameters of the model. Such networks can model functions of almost arbitrary complexity, with the number of layers, and the number of units in each layer, determining the function complexity.

The error of a particular configuration of the network can be determined by running all the training cases through the network, comparing the actual output generated with the desired or target outputs. The differences are combined together by an error function to give the network error. The most common error function is the sum-squared error, where the individual errors of output units on each case are squared and summed together. The root mean squared (RMS) error is the above normalised for the number of cases and variables, then square-rooted. The RMS neatly summarises the error over the entire training set and set of output units; it is thus used herein.

Prediction of dye properties such as substantivity, exhaustion, fixation and even properties such as aggregation and solubility would be of immense benefit. The ability to predict the dye properties would have huge consequences on the synthesis and development of novel dye molecules for applications such as textile dyeing and printing and paper printing to applications as diverse as antibody–dye molecular recognition in techniques such as photodynamic therapy in biological systems. To date, several linear modelling techniques, such as multiple linear regression analysis, have been used in attempts to predict the properties of dye molecules [2–6]. Unfortunately, none of these linear techniques have been successful enough to be of significant use. Greaves et al. have already demonstrated that the prediction of the substantivity of anionic, water soluble dyes using linear modelling techniques is immensely complex [7]. The complexity of this problem was deemed a suitable challenge for a neural network approach. In this paper feed-forward back propagation neural networks are used to predict the substantivity of anionic, water soluble dyes.

2. Experimental design

The software packages Statistica [8] and Statistica Neural Networks [9] were used for all analyses in this section.

3. Determination of % substantivity

The % substantivity was determined as previously described [10]. Simplistically, a defined concentration of the purified test dyestuff was stirred for 3 h in a vessel containing a given concentration of standardised biomass and specific water hardeners and pH to simulate conditions in a typical waste water treatment works (WWTW). After 3 h, the test medium was centrifuged and the optical density of the aqueous phase measured on a visible spectrophotometer to determine the amount of dye removed by adsorption onto the biomass. Selected dyes were measured five times to determine consistency and the values were found to be within $\pm$ 1% of the measured value (e.g. 15 $\pm$ 1%). At the dilutions used, absorption spectra of the dyes in solution obeyed the Beer–Lambert law so that aggregation effects could be ignored.

The test conditions reflected medium water hardness and 2.7 g/l suspended solids of an unacclimatised biomass.

4. Choice of type of neural network and learning algorithm

Since the output data (the substantivity results) was known, a supervised learning network (multilayer perceptron) was used involving a back propagation algorithm. Multilayer perceptrons are the most popular network architecture in use today. They account for 80% of all practical applications of neural networks.

5. Dyes

The dyes were either commercially available or specially synthesised and reflected either commercially important dyes or dyes specifically chosen to enable quantitative relationships to be developed. Reactive dyes were hydrolysed prior to testing to reflect the species responsible for the coloured effluent received at WWTW after a dyeing process. It has been shown that the parent dye and the hydrolysed dye can have vastly different substantivity characteristics [11]. The reactive dyes were heated in aqueous solution at 80°C and pH 10 (technical grade Na_2CO_3) until complete hydrolysis had occurred [as assessed by thin layer chromatography (TLC) and capillary zone electrophoresis (CZE)]. The TLC procedure was silica plate ascending and the eluent was *n*-butanol: acetone:conc. ammonia: water (5:5:2:1 v/v). The CZE procedure used a Dionex CE instrument fitted with UV–VIS detector, a 50 cm × 50 μm silica capillary, 50 kV voltage, 10 mM boric acid and 10 mM sodium dodecyl sulphate at pH 9.4. All dyes were purified by desalination in distilled water using visking tubing (Fisher catalogue no. TWT-400-070M), re-precipitated with 74OP methylated spirits, triturated with methanol and oven dried at 105°C. The purities of all the dyes were assessed by TLC and CZE and the colour strengths determined by elemental analysis (%C, %N and %S) and were found to be greater than 95% pure.

6. Choice of molecular descriptors

Multiple linear regression analysis had previously demonstrated that the size/charge of the dyes (number of carbon atoms divided by the number of sulphonate groups; C/S), length of the dye (Z-3D), number of aromatic primary amino groups (AR-NH2), number of unsulphonated naphthalene nuclei (NAP) and the number of aliphatic alcohol groups (ALK-OH) were important [12]. Consequently, these descriptors were used as inputs into neural network 1.

The intention of network 2 was to determine whether it was possible to predict the substantivity using simple empirical data (molecular composition). Several of the empirical descriptors showed significance in the previous MLR studies [7].

Network 3 was conducted using a set of input variables (molecular descriptors) to try and reflect the size (MW), shape (Z-3D), bonding abilities (HBD, HBA), hydrophobicity (LOGP) and empirical make-up (SULPHONE and AR-NH2) of the dyes.

A full description of the molecular descriptors is given below.

- ALK-OH: number of aliphatic alcohol groups.
- AR-NH2: number of aromatic primary amines.
- C/S: number of carbon atoms divided by number of sulphonate groups.
- HBD: hydrogen bond donating ability. Sum of the atomic charges on all the hydrogen atoms excluding those involved in conjugation and intra-molecular hydrogen bonds divided by the number of atoms. The atomic charges were obtained from Hyperchem [13] after molecular mechanics (MM+) minimisation (using the Fletcher–Reeves algorithm) of the ionised form (SO_3^-) of the dye.
- HBA: hydrogen bond accepting ability. Sum of the atomic charges on all the oxygen and nitrogen atoms (excluding those involved in conjugation and intramolecular hydrogen bonds) and dividing by the total number of atoms in the molecule. The atomic charges were obtained from Hyperchem after molecular

Table 1
The values of the molecular descriptors

Dye	% SUB	HBD	HBA	Z-3D	LOGP	ALK-OH	AR-NH2	NAP	%C	%N	%S	%O	%H	C/S	MW	SULPHONE
1	42.0	0.040	0.000	10.70	−5.20	0	1	0	60.91	7.10	8.13	20.28	3.58	20	393	0
2	46.0	0.043	−0.013	12.93	−6.47	0	1	0	59.35	9.03	6.89	20.62	4.11	23	464	0
3	51.0	0.045	−0.009	10.80	−6.77	0	2	0	59.57	9.92	7.57	18.89	4.05	21	422	0
4	33.0	0.039	0.000	11.31	−4.25	0	1	0	59.99	7.00	8.01	19.98	5.03	20	398	0
5	45.0	0.040	0.000	11.89	−5.05	0	1	0	61.75	6.86	7.85	19.59	3.95	21	407	0
6	44.0	0.043	−0.007	12.39	−6.62	0	1	0	58.53	9.31	7.10	21.26	3.80	22	450	0
7	50.0	0.046	0.000	11.59	−6.92	0	2	0	58.67	10.26	7.83	19.54	3.69	20	408	0
8	48.0	0.044	−0.007	13.34	−6.00	0	1	0	59.35	9.03	6.89	20.62	4.11	23	464	0
9	50.0	0.044	−0.006	14.45	−5.60	0	1	0	60.12	8.76	6.69	20.02	4.41	24	478	0
10	64.0	0.043	−0.005	18.11	−4.41	0	1	0	62.17	8.06	6.15	18.40	5.22	27	521	0
11	59.0	0.043	−0.006	15.37	−6.92	0	1	0	63.15	8.18	6.24	18.69	3.73	27	512	0
12	11.0	0.032	0.000	18.40	−8.83	0	0	0	58.12	4.84	11.08	22.12	3.83	14	576	0
13	0.0	0.031	0.000	14.31	−8.83	0	0	0	58.12	4.84	11.08	22.12	3.83	14	576	0
14	84.0	0.034	0.000	24.41	−6.45	0	0	0	61.61	4.23	9.68	19.31	5.17	17	659	0
15	0.0	0.024	0.000	12.94	−13.26	0	2	0	39.07	6.51	14.90	37.18	2.34	7	428	0
16	22.0	0.040	0.000	13.47	−2.14	0	0	1	58.53	8.53	9.77	19.49	3.68	16	327	0
17	29.0	0.040	0.000	12.91	−2.14	0	0	1	58.53	8.53	9.77	19.49	3.68	16	327	0
18	11.0	0.047	0.000	13.12	−2.27	0	0	1	58.53	8.53	9.77	19.49	3.68	16	327	0
19	79.0	0.040	0.000	13.27	−2.07	0	0	2	63.48	7.40	8.47	16.91	3.73	20	378	0
20	90.0	0.041	0.000	14.60	−2.07	0	0	2	63.48	7.40	8.47	16.91	3.73	20	378	0
21	96.0	0.043	0.000	19.38	−1.77	0	0	1	61.10	12.96	7.41	14.80	3.73	22	428	0
22	0.0	0.041	0.000	14.12	−7.04	0	0	2	52.40	6.11	13.99	24.43	3.08	10	432	0
23	11.0	0.034	0.000	14.56	−7.04	0	0	2	52.40	6.11	13.99	24.43	3.08	10	432	0
24	4.0	0.038	0.000	14.71	−7.04	0	0	2	52.40	6.11	13.99	24.43	3.08	10	454	0
25	0.0	0.029	0.000	14.34	−12.01	0	0	2	44.61	5.20	17.86	29.71	2.62	6.7	538	0
26	0.0	0.029	0.000	14.40	−12.01	0	0	2	44.61	5.20	17.86	29.71	2.62	6.7	538	0
27	0.0	0.047	0.000	12.77	−3.95	0	0	1	58.53	8.53	9.77	19.49	3.68	16	327	0
28	8.0	0.035	0.000	19.21	−6.76	0	0	1	51.56	10.93	12.51	21.85	3.15	11	510	0
29	22.0	0.042	−0.034	16.50	−4.00	0	0	0	45.37	12.02	11.01	16.48	2.94	22	523	1
30	35.0	0.048	0.000	13.64	−2.85	0	0	0	48.92	14.26	8.16	16.29	3.34	16	441	0
31	40.0	0.036	0.000	25.61	−6.25	0	0	0	51.75	18.86	8.63	17.23	3.53	16	710	0
32	0.0	0.025	0.000	14.56	−8.04	0	0	0	37.88	11.04	12.64	22.08	2.38	8	505	0
33	6.0	0.027	0.000	16.19	−7.97	0	0	1	43.10	10.05	11.51	20.09	2.53	10	555	0
34	33.0	0.043	0.000	15.18	−2.63	0	0	0	53.62	15.63	8.95	17.86	3.94	16	357	0
35	0.0	0.036	0.000	17.34	−7.67	0	0	0	41.03	11.96	13.69	30.74	2.58	8	466	0
36	24.0	0.047	−0.014	19.81	−3.62	0	0	0	52.36	13.27	12.16	18.20	4.01	23	527	1
37	75.0	0.043	−0.042	16.14	−3.68	0	0	0	49.07	9.95	11.39	19.89	3.40	23	562	1
38	78.0	0.041	−0.042	17.61	−3.46	0	0	0	52.26	10.60	12.13	21.19	3.81	23	528	1
39	8.0	0.028	−0.014	18.46	−13.14	0	0	1	46.90	13.67	10.44	26.03	2.95	11	570.5	0
40	8.0	0.037	0.000	17.92	−10.18	0	2	1	48.70	15.49	11.82	20.64	3.34	11	540.5	0
41	9.0	0.041	0.000	16.54	−14.50	0	0	2	49.60	6.67	15.28	25.41	3.04	8.7	627	0
42	22.0	0.034	−0.043	15.25	−8.26	0	0	1	46.88	7.46	17.07	25.55	3.04	11	574	1
43	96.0	0.039	−0.004	29.07	−4.02	0	0	1	56.85	6.63	10.12	20.20	6.20	15	619	0
44	26.0	0.035	−0.029	14.81	−8.99	0	0	1	48.73	6.09	13.94	23.18	2.92	14.5	701.5	1
45	7.0	0.035	−0.012	18.10	−9.95	0	0	1	45.97	10.72	12.27	27.56	3.47	10	522	0
46	11.0	0.039	0.000	15.75	−10.55	0	2	1	43.83	12.78	14.63	25.54	3.22	8	436	0
47	39.0	0.039	−0.009	21.04	−7.65	0	0	1	55.72	8.66	9.92	22.27	3.43	15	646	0
48	17.0	0.037	−0.050	18.30	−9.10	0	0	1	48.25	7.50	17.17	23.57	3.51	15	746	2
49	96.0	0.045	0.000	20.63	−8.13	0	0	2	60.27	10.98	10.06	15.05	3.64	16	635	0
50	90.0	0.056	−0.044	20.32	−9.04	0	0	0	52.81	9.85	11.28	22.51	3.55	25	567	1

(continued on next page)

Table 1 (*continued*)

Dye	% SUB	HBD	HBA	Z-3D	LOGP	ALK-OH	AR-NH2	NAP	%C	%N	%S	%O	%H	C/S	MW	SULPHONE
51	5.0	0.057	0.000	15.08	−5.82	0	0	1	60.09	5.19	11.88	17.79	5.04	15	577	0
52	0.0	0.034	0.000	13.16	−7.99	0	0	0	67.61	5.84	6.69	13.34	6.51	13.5	558	0
53	22.0	0.036	−0.014	33.20	−16.20	0	0	2	49.10	13.74	10.49	23.54	3.13	12.5	1206	0
54	12.0	0.033	−0.003	42.29	−25.50	0	0	2	43.55	11.28	15.50	27.07	2.60	7.5	1226	0
55	53.0	0.038	−0.004	30.21	−16.76	0	3	1	48.83	16.27	11.17	20.44	3.28	11.7	827	0
56	8.0	0.042	−0.019	22.01	−16.43	1	0	2	47.85	14.31	13.10	21.25	3.50	9.8	973	0
57	89.0	0.049	0.000	29.21	−17.66	0	4	2	51.91	17.89	9.45	17.29	3.47	14.7	1015	0
58	25.0	0.030	0.000	23.25	−17.88	0	0	0	49.22	6.75	11.59	21.21	2.67	10	826	0
59	7.0	0.040	0.000	38.75	−15.66	0	0	0	49.77	12.22	13.99	20.94	3.08	9	912	0
60	9.0	0.045	−0.023	28.45	−22.39	0	0	2	46.69	14.52	12.46	22.80	3.53	10	1537	0
61	50.0	0.042	−0.024	28.34	−17.15	1	0	2	53.06	15.86	7.26	19.93	3.88	19.5	881	0
62	4.0	0.035	−0.004	28.12	−17.56	0	0	2	48.16	9.63	14.69	23.83	3.69	8.8	869	0
63	28.0	0.041	−0.005	27.93	−7.01	0	0	1	55.14	11.09	10.15	20.26	3.35	14.5	629.6	0
64	5.5	0.025	0.000	27.80	−20.64	0	2	2	45.13	9.29	14.17	28.29	3.12	8.5	900.8	0
65	36.0	0.040	0.008	30.76	−7.57	0	0	2	54.61	12.74	8.33	20.78	3.54	17.5	767.7	0
66	41.0	0.039	0.000	27.06	−17.80	0	1	4	55.35	11.30	11.08	18.43	3.83	13.3	864.9	0
67	15.8	0.030	0.000	29.14	−18.34	0	2	2	46.78	9.63	14.69	25.66	3.23	8.5	868.8	0
68	16.1	0.032	0.000	27.33	−18.34	0	2	2	46.78	9.63	14.69	25.66	3.23	8.5	868.8	0
69	26.0	0.036	0.000	22.99	−7.95	0	0	1	56.71	10.33	9.46	18.89	4.61	16	675.7	0
70	8.5	0.044	0.000	21.98	−9.04	0	0	0	51.98	10.10	11.56	23.08	3.27	12	552.5	0
71	0.0	0.038	−0.009	20.30	−10.68	1	0	0	48.54	9.06	15.55	23.28	3.58	12.5	616.6	0
72	94.0	0.040	0.000	25.39	−9.97	0	2	2	59.99	12.35	9.42	14.10	4.15	17	678.7	0
73	95.0	0.044	0.000	25.14	−10.28	0	2	2	58.88	12.88	9.83	14.71	3.71	16	650.7	0
74	41.0	0.044	−0.004	25.25	−9.68	0	0	2	58.69	11.41	8.70	17.37	3.83	18	734.8	0
75	68.7	0.044	−0.003	25.32	−7.63	0	0	2	60.59	10.60	8.09	16.14	4.58	20	790.9	0
76	86.0	0.056	−0.033	31.71	−11.45	0	2	0	57.76	9.62	8.81	19.79	4.02	17.5	725.8	0
77	16.0	0.030	−0.005	20.93	−23.36	0	2	2	48.17	13.62	11.69	23.33	3.19	8.3	820.8	0
78	10.5	0.036	−0.030	27.38	−28.99	0	1	1	38.43	16.81	15.39	26.88	2.50	6.7	1243	0
79	12.9	0.034	−0.028	32.38	−16.78	0	1	1	39.47	16.44	15.05	26.29	2.76	7	1271	0
80	0.4	0.040	−0.059	23.94	−12.54	2	1	1	41.99	9.42	17.25	27.96	3.39	13	741	2
81	6.2	0.036	−0.044	24.45	−10.41	1	1	1	38.42	12.80	14.65	27.42	2.65	14	893.5	1
82	0.0	0.042	−0.052	28.09	−17.48	2	1	1	42.75	10.31	15.74	27.49	3.71	9.7	789	2
83	4.8	0.035	−0.035	24.76	−12.06	2	1	1	41.27	8.91	16.32	30.54	2.95	13.5	838	1
84	10.0	0.033	−0.021	26.05	−17.24	0	1	1	40.33	15.17	13.89	24.26	2.51	7.8	918.5	0
85	7.1	0.030	−0.019	27.19	−22.13	0	1	2	39.91	13.30	15.22	25.82	2.39	7.4	1106	0
86	3.2	0.028	−0.034	14.94	−11.04	0	0	1	37.13	13.68	15.65	31.24	2.30	6.3	610	0
87	5.0	0.034	−0.025	15.49	−12.27	0	0	1	42.27	10.96	12.54	27.11	2.50	9	762.5	0
88	2.1	0.038	−0.052	18.76	−12.74	1	0	1	40.65	12.29	16.08	28.08	2.91	9	793	1
89	7.3	0.035	−0.026	13.86	−10.54	0	0	1	44.38	13.93	13.67	25.01	3.01	8.7	700	0
90	16.4	0.030	−0.025	25.82	−22.99	0	0	2	40.61	15.07	14.79	27.05	2.48	7.3	1294	0
91	14.5	0.029	−0.029	25.44	−22.75	0	0	2	41.09	14.91	14.63	26.76	2.61	7.5	1306	0
92	0.0	0.047	−0.077	27.45	−9.67	2	0	1	39.46	11.93	15.61	29.20	3.80	13.5	818	2
93	0.0	0.051	−0.049	17.30	−12.74	1	0	1	40.65	12.29	16.08	28.08	2.91	9	793	1
94	5.7	0.044	−0.052	24.39	−18.51	2	0	2	40.52	10.90	16.64	29.06	2.88	9.8	1180	2
95	11.8	0.025	−0.022	30.83	−32.40	0	0	4	40.40	12.45	16.28	28.43	2.43	6.6	1552	0
96	0.0	0.029	−0.016	18.70	−17.16	0	0	2	41.50	9.37	14.30	28.53	2.36	7.8	892.5	0
97	1.4	0.034	−0.041	24.40	−17.63	1	0	2	40.13	10.57	17.28	29.31	2.72	7.8	923	1
98	1.3	0.030	−0.022	25.25	−17.63	1	0	2	40.13	10.57	17.28	29.31	2.72	7.8	923	1
99	12.5	0.028	−0.021	33.14	−32.86	0	0	4	40.00	12.56	16.43	28.69	2.32	6.5	1552	0
100	19.6	0.040	−0.034	11.35	−9.14	0	1	0	46.00	13.99	10.68	26.64	2.69	11.5	598	0
101	0.0	0.034	−0.029	20.31	−10.72	0	0	2	42.48	12.38	14.18	28.29	2.67	8	712	0
102	3.0	0.039	−0.024	21.83	−11.15	0	0	1	44.26	15.29	13.13	24.02	3.30	9	685	0
103	25.3	0.045	−0.041	12.80	−7.24	1	1	0	52.58	5.57	12.76	25.47	3.61	22	501	1

mechanics (MM+) minimisation (using the Fletcher–Reeves algorithm) of the ionised form (SO_3^-) of the dye.
- LOGP: calculated partition co-efficient in octanol–water (obtained from the ChemPlus package within Hyperchem).
- MW: molecular weight of the dye anion, excluding any counter-ion.
- SULPHONE: number of sulphone (SO_2) groups, including sulphonamide groups and sulphonate esters but excluding sulphonate groups.
- NAP: number of unsulphonated naphthalene nuclei.
- Z-3D: length of the z axis when the 3D conformation of the dye is encapsulated in a periodic box.

%C, %N, %S, %O and %H of the sulphonic acid form of the dye.

7. Experimental procedure

The values of the molecular descriptors were used as inputs in the neural network. The values are given in Table 1. The input cases were randomly divided into training cases (T) and verification cases (V) by the computer software. Approximately 10% of the cases were used as verification cases. One hidden layer was used. The Automatic Network Designer in Statistica Neural Networks [9] conducted a vast number of experiments to determine the best number of units in the hidden layer. The best network (in terms of verification error) was retained and the experimental result compared against the predicted result using standard statistical tools such as the regression coefficient (r) and the statistical significance level (P). The statistical results were considered significant at $P < 0.01$. Further details are given in Table 1.

8. Results and discussion

The values of the molecular descriptors are shown in Table 1. The results of each back propagation neural network are shown in Table 2. All the results of the networks were highly significant because the significance levels (P values) were less than 0.01 (1%).

In network 1, the result demonstrated that the size/charge and length of the dye, number of aromatic primary amino groups, number of unsulphonated naphthalene nuclei and the number of aliphatic alcohol groups were important in predicting the substantivity.

The intention of network 2 was to determine whether it was possible to predict the substantivity using simple empirical (molecular composition) data. The results show that a moderate prediction of substantivity is possible if only the molecular composition values are used as input variables in the network.

Network 3 was conducted using a set of input variables (molecular descriptors) to try and reflect the size (MW), shape (Z-3D), bonding abilities (HBD, HBA), hydrophobicity (LOGP) and empirical make-up (SULPHONE and AR-NH2) of the dyes. The magnitude of the correlation coefficient implied that a moderate relationship existed between the molecular descriptors and the substantivity.

All the correlation coefficients (r values) were of a magnitude that implied there was a moderate relationship between the predicted substantivity

Table 2
Results of the prediction of the substantivity using back propagation neural networks

Network	Optimum architecture (input×hidden×output)	No. of dyes		RMS error				
		T	V	T	V	r	r^2	P
1	5×7×1	93	10	0.19	0.08	0.80	0.65	0.00
2	5×10×1	93	10	0.20	0.16	0.77	0.60	0.00
3	7×7×1	93	10	0.20	0.15	0.76	0.58	0.00

and the observed (experimentally determined) substantivity. In other words, the molecular descriptors in each of the 3 networks were capable of describing/predicting the substantivity. Even so, the networks implied that the size, bonding ability, shape, hydrophobicity and empirical make-up of the dyes are not the only molecular features responsible for the substantivity. Further work, in particular, the input of a greater number of dyes (input cases) should help to improve the understanding of the molecular features that are responsible for the substantivity of anionic, water soluble dyes on a biomass.

9. Conclusion

It is possible to predict the substantivity of anionic, water soluble dyes on a biomass using relatively simple empirical and semi-empirical molecular descriptors and a feed-forward back propagation neural network. Depending on the input molecular descriptors and the number of hidden layers in the neural network the correlation between the predicted substantivity and the observed (experimentally determined) substantivity was 0.80 and 0.76. The magnitude of the correlation coefficients implied that the size, bonding ability, shape, hydrophobicity and empirical make-up of the dyes are not the only molecular features responsible for the substantivity.

Acknowledgements

BASF and Severn Trent Water plc for their financial support and Professor D.A.S. Phillips, Dr. M.G. Hutchings and Mr. J.H. Churchley for their valuable comments.

References

[1] Westland S. J Soc Dyers and Colourists 1998;114:309–13.
[2] Timofei S, Kurunczi L, Schmidt W, Simon Z. Dyes and Pigments 1996;32:25–42.
[3] Timofei S, Fabian WMF. Journal of Molecular Structure (Theochem) 1996;362:155–62.
[4] Timofei S, Kurunczi L, Schmidt W, Simon Z. Dyes and Pigments 1995;29:251–8.
[5] Timofei S, Kurunczi L, Fabian WMF, Suzuki T, Muresan S. Dyes and Pigments 1997;34:181–93.
[6] Timofei S, Kurunczi L, Schmidt W. Quant Struct-Act Relat 1995;14:444–9.
[7] Greaves AJ, Churchley JH, Hutchings MG, Phillips DAS, Taylor JJ. Soc. Dyers and Colourists 2000;116:25–31.
[8] Anon. Statistica, version 5.1. USA: Statsoft Inc. (www.Statsoft.com).
[9] Anon. Statistica neural networks, version 1. USA: Statsoft Inc. (www.Statsoft.com).
[10] Greaves AJ, Churchley JH, James AE, Hutchings MG, Phillips DAS. Water Research 2000;34(5):1673–9.
[11] Greaves AJ, Phillips DAS. J Soc Dyers and Colourists 1999;115:363–5.
[12] Greaves AJ, Churchley JH, James AE, Hutchings MG, Phillips DAS. Water Research, in print.
[13] Anon. Hyperchem, version 4.0. USA: Hypercube Inc. (www.hyper.com).