AMECJ

Determination and prediction of peptide mobilities by micellar electro-kinetic chromatography using adaptive neuro-fuzzy inference system as a feature selection method

Mostafa Hassanisadi [a,*], Morteza G. Khaledi [b] and Mehdi Jalali-Heravi [c]

[a]*Nanotechnology Research Center, Research Institute of Petroleum Industry, Tehran, Iran.*
[b]*Department of Chemistry, Sharif University of Technology, P.O.Box9516-11365, Tehran, Iran*
[c]*Department of Chemistry, North Carolina State University, NC8204-27695, USA*

ABSTRACT

Mobility of 128 peptides composed of up to 14 amino acids is determined for sodium dodecyl sulfate (SDS) micellar systems using micellar electrokinetic chromatography (MEKC). The mobilities of these peptides are predicted using back propagation of error artificial neural networks (BP-ANNs). Adaptive neuro-fuzzy inference system (ANFIS) which can deal with linear and nonlinear phenomena is used to select the inputs of BP-ANN. A 3:4:1 BP-ANN model with four variables of Kappa substituent constant, Kappa(H), number of peptide bonds, (lnN), molar refractivity of C-terminal, MRC, and steric effects at N-terminal, ES,N, which incorporate substituent, steric and molar refractivity effects as its inputs was developed. Comparison of Multiple Linear Regression (MLR) and ANN results shows the nonlinear characteristic of the phenomena. The nonlinear model was successful in predicting the mobilities of 120 peptides except for the ones (8 peptides) with negatively charged amino acids. It is shown that that most outlier peptides contain middle glutamic acid (E) and aspartic acid (D) amino acids and their mobilities follow a similar mechanism in MEKC.

1. Introduction

Micellar electrokinetic chromatography (MEKC) is a widely used technique in capillary electrophoresis (CE) and is capable of separating neutral compounds as well as charged solutes by including a pseudo-stationary phase [1-6]. This technique has great utility in separating mixtures that contain both ionic and neutral species, and has become a valuable tool in separating very hydrophobic pharmaceuticals from their very polar metabolites. The creation of the pseudo-stationary phase is most easily achieved using micelles of surfactants and depending on the hydrophobicity, analytes partition between these micelles and the mobile phase. The significant features of MEKC are the availability of a wide range of pseudo-stationary phases that provide unique selectivities for peptides and feasibility of manipulating the comparison of the pseudo-phases since it is a completely solution-based technique. Therefore, selectivity in MEKC can be varied by altering the nature of the micelles [6-8]. This could be achieved by altering the surfactant and changing the size, charge or geometry of the micelles. It is shown that not only the predominant hydrophobic interaction but also other important solute-micelle

---

interactions such as electrostatic and hydrogen bonding could manipulate the separation. In most MEKC studies, sodium dodecyl sulphate (SDS) which is an anionic micelle has been successfully used to separate the hydrophobic and cationic analytes. An important parameter for the separation of the peptides and modeling of the electropherograms in CE and MEKC is their mobilities. This parameter can be converted to migration time and then electropherograms can be simulated using gaussian function. A long-range goal of our laboratory is developing experimental and theoretical methods for peptide separations, and mapping two-dimensional MEKC-CZE schemes. Reaching this goal requires an in-depth understanding of the effects of different factors on the CZE and MEKC peptide mobilities. Quantitative models such as quantitative structure-mobility relationships (QSMR) can help us to gain this knowledge. We have started with the prediction of electrophoretic mobilities of 125 peptides using CZE technique [9]. A QSMR model has been developed using Offord's charge-over-mass term (Q/M2/3) together with the corrected steric substituent constant (ES,C)and molar refractivity (MR) as descriptors. The latter two parameters, account for the steric effects and bulkiness of amino acid side chains, respectively [9]. The robustness of this work was shown by artificial neural network (ANN) modeling of the mobilities of 102 larger peptides – up to 42 amino acid residues – that also included highly charged and hydrophobic peptides [10]. Besides, to explore the utility of the ANN model in simulation of peptide maps, the profile for the endoproteinase digest of the melittin, glucagon and horse cytochrome C, was also studied in the latter work [10]. We intended to examine the same route for the MEKC modeling as we did for the CZE. Therefore, the main aim of the present work was the determination of the mobilities of a set of small peptides – up to 14 amino acid residues – using MEKC and then modeling the mobilities by applying different chemometric techniques. Artificial neural networks (ANNs) are among the most popular methods for modeling of the linear/nonlinear phenomena [11]. ANN-based approaches have the ability of modeling the complex data without the need for a detailed understanding of the underlying

phenomena. Back propagation (BP) learning rule is the most popular learning algorithm adopted in neural network technology. Hence in the present research, a back-propagation of error artificial neural network (BP-ANN) was used to predict the mobilities of 128 peptides obtained using MEKC with SDS micellar system. However, the main problem in developing the ANNs, is the selection of suitable descriptors for their inputs. This is especially serious when the mechanism of the phenomenon is complex or unknown. In order to overcome this problem one needs to use a powerful method for the feature selection. Therefore, in the present work we have chosen adaptive neuro-fuzzy inference system (ANFIS) for selecting the most effective parameters on MEKC mobilities. This method is capable in dealing with linear and nonlinear phenomena. Success in modeling of the electrophoretic mobilities of peptides using MEKC, together with our previous achievements in modeling of CZE mobilities might pave the way for developing and predicting the two-dimensional MEKC-CZE maps of peptides.

2. Experimental
2.1. Chemicals and Materials
Sodium dodecyl sulfate (SDS), decanophenone, sodium phosphate monobasic ($NaH_2PO_4.H_2O$), hexanol, and peptides were obtained from Sigma Chemical Co. (St. Louis, MO). Different concentrations of 40, 60 and 80 mM SDS were prepared in 20 mM phosphate buffer at pH 7 with 1.15 % (v/v) hexanol. The solutions were filtered through 0.2 μm acrodisc filter (STRL, Eatontown, NJ) before use. All experiments were carried out on a home-built CE system comprised of a 0-30 kV high voltage power supply (Series EH, Glassman High Voltage, Inc., White house Station, NJ). Fused silica capillary (Polymicro Technologies, Phoenix, AZ), with an inner diameter of 50 μm and an outer diameter of 375 μm was used. The total capillary length and the length from the inlet to the detector were 71 and 47 cm, respectively. A circulating mineral oil bath was used to maintain the temperature of the two buffer reservoirs and the capillary at a designed temperature in this experiment. A positive voltage of 25 kV was applied during the experiments. A variable-wavelength UV detector (Model 200,

Scientific System, Inc., State College, PA) was used, and a wavelength of 214 nm was set in this work. The chromatograms were collected using acquisition software written in LabView (Austin, TX). Before any injection was made, the untreated capillary was conditioned by rinsing with deionized (DI) water for 20 minutes, sodium hydroxide dissolved in methanol for 10-12 minutes, DI water for 20 minutes and finally with the buffer for 15 minutes. The capillary was vacuum rinsed with the buffer solution between each injection.

2.2. Determination of mobility by measurement of migration time of peptides

Electrophoretic mobility at a micelle concentration can be determined from the migration times using equation 10 [2,20]:

$$\mu = \frac{L_t L_d}{V}\left(\frac{1}{t_r} - \frac{1}{t_o}\right) \qquad (10)$$

where Lt is the total length of the capillary, Ld is the separation length (from the upstream end of the capillary to the detection window). V is the applied voltage. tr is the retention time of a solute at a given micelle concentration, and to is the retention time of an unretained solute. The determined mobility values of peptides in 40, 60 and 80 mM SDS solutions together with the peptides studied in this work are shown in Table 1.

Table 1. The values of electrophoretic mobilities of peptides using MEKC with different SDS concentrations together with the calculated values of descriptors

#	Peptide	Descriptors				Mobility		
		$E_{S,C}$	MR_n	ln(N)	Kappa(H)	SDS 80 mM	S D S 60mM	S S D S 40mM
1	$A_L Y_L$	-0.70	5.65	0.69	0.67	-6.38	-6.10	-5.73
2	GY	-0.50	1.03	0.69	0.46	-6.49	-5.55	-4.31
3	AY	-0.70	5.65	0.69	0.67	-6.88	-6.23	-4.76
4	ASTTTNYT	-3.88	5.65	2.08	-0.22	-8.71	-7.36	-5.15
5	VY	-1.79	14.95	0.69	1.09	-8.74	-8.13	-6.60
6	YV	-1.79	31.83	0.69	1.09	-10.04	-9.60	-8.58
7	YA	-0.70	31.83	0.69	0.67	-10.17	-9.30	-8.12
8	GGF	-0.30	1.03	1.10	1.06	-10.92	-9.61	-7.10
9	GF	-0.50	1.03	0.69	1.06	-12.09	-10.49	-8.30
10	YY	-1.40	31.83	0.69	0.92	-12.16	-11.49	-9.88
11	YG	-0.50	31.83	0.69	0.46	-12.23	-10.71	-8.12
12	IY	-2.31	19.59	0.69	1.41	-12.86	-11.42	-8.88
13	AF	-0.70	5.65	0.69	1.27	-13.18	-11.25	-7.89
14	HY	-1.36	23.79	0.69	0.90	-13.20	-11.63	-8.66
15	YI	-2.31	31.83	0.69	1.41	-13.52	-11.70	-9.80
16	LY	-1.94	19.59	0.69	1.40	-13.60	-11.97	-8.53
17	YGG	-0.30	31.83	1.10	0.46	-16.03	-14.47	-11.45
18	FA	-0.70	30.01	0.69	1.27	-16.43	-14.56	-11.52
19	FV	-1.79	30.01	0.69	1.69	-16.43	-14.65	-11.79
20	YL	-1.94	31.83	0.69	1.40	-16.72	-15.05	-11.45
21	GW	-0.46	1.03	0.69	1.01	-16.79	-14.85	-10.49
22	AW	-0.66	5.65	0.69	1.22	-17.89	-15.65	-11.99
23	VF	-1.79	14.95	0.69	1.69	-19.52	-16.75	-12.04
24	YAG	-0.50	31.83	1.10	0.67	-19.89	-17.00	-12.17
25	YYY	-2.10	31.83	1.10	1.38	-20.00	-18.09	-13.31
26	WS	-0.94	39.81	0.69	0.90	-20.57	-18.21	-13.89
27	FG	-0.50	30.01	0.69	1.06	-21.47	-18.60	-15.43
28	PW	-0.66	13.95	0.69	1.56	-21.78	-17.65	-13.05

#	Peptide	$E_{s,c}$	MR_n	ln(N)	Kappa(H)	SDS 80 mM	SDS 60 mM	SDS 40 mM
29	VW	-1.75	14.95	0.69	1.64	-22.08	-19.31	-13.51
30	FI	-2.31	30.01	0.69	2.01	-22.10	-19.09	-15.68
31	WD	-1.44	39.81	0.69	1.45	-22.22	-22.25	-19.11
32	DF	-1.48	11.58	0.69	1.50	-22.54	-22.18	-21.82
33	WA	-0.66	39.81	0.69	1.22	-22.60	-19.65	-14.85
34	$D_L F_L$	-1.48	11.58	0.69	1.50	-23.16	-22.61	-21.64
35	FM	-1.53	30.01	0.69	1.66	-23.23	-20.79	-15.57
36	GGFM	-1.13	1.03	1.39	1.66	-23.74	-20.27	-14.12
37	WV	-1.75	39.81	0.69	1.64	-23.83	-20.62	-16.14
38	FGG	-0.30	30.01	1.10	1.06	-24.86	-22.50	-16.57
39	YGGF	-1.00	31.83	1.39	1.52	-25.14	-23.13	-15.69
40	EW	-1.28	16.23	0.69	1.43	-25.54	-22.75	-21.33
41	YW	-1.36	31.83	0.69	1.47	-25.54	-21.90	-18.32
42	YYL	-2.64	31.83	1.10	1.86	-25.68	-23.87	-16.44
43	DW	-1.44	11.58	0.69	1.45	-25.77	-22.99	-21.18
44	MW	-1.49	23.12	0.69	1.61	-25.77	-21.96	-18.07

		Descriptors				Mobility		
#	Peptide	$E_{s,c}$	MR_n	ln(N)	Kappa(H)	SDS 80 mM	SDS 60 mM	SDS 40 mM
45	WE	-1.28	39.81	0.69	1.43	-26.89	-21.82	-18.29
46	FL	-1.94	30.01	0.69	2.00	-28.13	-25.32	-18.93
47	WG	-0.46	39.81	0.69	1.01	-28.40	-25.57	-19.98
48	IW	-2.27	19.59	0.69	1.96	-28.84	-25.41	-18.06
49	IF	-2.31	19.59	0.69	2.01	-29.15	-25.04	-19.64
50	PPGFSP	-0.78	13.95	1.79	2.60	-29.19	-25.75	-17.08
51	GGFL	-1.54	1.03	1.39	2.00	-29.26	-26.34	-19.32
52	WM	-1.49	39.81	0.69	1.61	-29.56	-26.25	-20.98
53	LF	-1.94	19.59	0.69	2.00	-31.16	-28.17	-20.70
54	WP	-0.66	39.81	0.69	1.56	-31.20	-27.36	-21.35
55	KF	-1.32	25.05	0.69	2.20	-31.69	-29.11	-21.43
56	YPF	-1.40	31.83	1.10	2.07	-31.97	-28.32	-20.78
57	WY	-1.36	39.81	0.69	1.47	-32.60	-26.46	-20.86
58	$L_L W_L$	-1.90	19.59	0.69	1.95	-32.61	-28.95	-22.63
59	FF	-1.40	30.01	0.69	2.12	-32.72	-29.83	-22.99
60	$F_L F_L$	-1.40	30.01	0.69	2.12	-32.84	-30.44	-23.48
61	$K_L F_L$	-1.32	25.05	0.69	2.20	-32.87	-30.20	-21.63
62	LW	-1.90	19.59	0.69	1.95	-33.11	-29.65	-21.60
63	GLF	-1.74	1.03	1.10	2.00	-33.96	-31.49	-23.70
64	WGG	-0.26	39.81	1.10	1.01	-34.06	-30.35	-24.09
65	GFL	-1.74	1.03	1.10	2.00	-34.55	-31.48	-24.40
66	FW	-1.36	30.01	0.69	2.07	-35.09	-32.21	-26.14
67	YGGFM	-1.83	31.83	1.61	2.12	-35.13	-31.37	-22.40
68	WL	-1.90	39.81	0.69	1.95	-35.27	-31.65	-25.71
69	MLF	-2.77	23.12	1.10	2.60	-35.62	-32.99	-26.10
70	WGY	-1.16	39.81	1.10	1.47	-36.11	-32.72	-26.61
71	YAGFL	-2.44	31.83	1.61	2.67	-37.52	-34.85	-26.72
72	YGGFL	-2.24	31.83	1.61	2.46	-37.59	-36.08	-26.09
73	WGGGY	-0.76	39.81	1.61	1.47	-38.83	-35.78	-29.77
74	KW	-1.28	25.05	0.69	2.15	-38.88	-36.79	-30.76
75	WGGY	-0.96	39.81	1.39	1.47	-38.88	-35.47	-29.39
76	FGGF	-1.00	30.01	1.39	2.12	-38.95	-35.19	-29.22
77	WF	-1.36	39.81	0.69	2.07	-39.97	-37.28	-31.14
78	YSGFLT	-3.25	31.83	1.79	2.20	-39.98	-35.47	-28.51

#	Peptide	$E_{S,C}$	MR_n	ln(N)	Kappa(H)	SDS 80 mM	SDS 60 mM	SDS 40 mM
79	WW	-1.32	39.81	0.69	2.02	-41.50	-39.25	-33.67
80	TRSAW	-2.09	11.82	1.61	2.53	-41.52	-40.67	-33.07
81	YGGWL	-2.20	31.83	1.61	2.41	-42.21	-40.85	-32.63
82	FGFG	-1.00	30.01	1.39	2.12	-42.78	-39.83	-34.28
83	RW	-1.28	30.05	0.69	2.58	-43.08	-41.39	-33.55
84	FFF	-2.10	30.01	1.10	3.18	-44.33	-43.35	-38.04
85	RPPGK	-1.04	30.05	1.61	3.81	-44.43	-42.77	-35.55
86	DRVYIHP	-5.46	11.58	1.95	5.04	-44.53	-43.75	-37.60
87	KYK	-1.94	25.05	1.10	2.74	-45.47	-45.20	-41.39
88	RPPGFSP	-1.40	30.05	1.95	4.17	-45.63	-45.23	-41.06
89	WR	-1.28	39.81	0.69	2.58	-46.11	-44.77	-38.45
90	DRVYIHPF	-6.16	11.58	2.08	6.10	-46.54	-45.96	-41.39
91	ELYENKPRRPY	-6.52	16.23	2.40	7.90	-46.61	-46.27	-43.06
92	DRVYVHPFHL	-7.54	11.58	2.30	7.16	-47.11	-47.08	-45.15
93	NRVYVHPF	-5.64	14.46	2.08	5.16	-47.12	-47.29	-44.15

		Descriptors				Mobility		
#	Peptide	$E_{S,C}$	MR_n	ln(N)	Kappa(H)	SDS 80 mM	SDS 60 mM	SDS 40 mM
94	YMEHFRW	-4.79	31.83	1.95	5.56	-47.21	-46.98	-43.99
95	RYLGYL	-4.30	30.05	1.79	4.37	-47.31	-46.99	-45.06
96	ELYENKPRRPYIL	-9.37	16.23	2.56	9.79	-47.31	-46.57	-44.76
97	FFFF	-2.80	30.01	1.39	4.24	-47.35	-47.43	-44.36
98	CGYGPKKKRKVGG	-4.09	13.90	2.56	9.05	-47.48	-46.26	-41.74
99	RPKPQQFFGLM	-5.75	30.05	2.40	7.59	-47.55	-47.82	-46.31
100	YRPPGFSPFR	-3.42	31.83	2.30	7.26	-47.55	-47.82	-46.31
101	MEHFRWG	-3.89	23.12	1.95	5.10	-47.56	-47.47	-43.97
102	DRVYIHPFHL	-8.06	11.58	2.30	7.48	-47.66	-46.85	-43.59
103	ELYENKPRRPFIL	-9.37	16.23	2.56	10.39	-47.71	-46.77	-44.90
104	AGCKNFFWKTFTSC	-5.92	5.65	2.64	8.65	-47.71	-47.66	-45.63
105	RPKPQQF	-3.18	30.05	1.95	4.99	-47.83	-47.28	-44.08
106	RPPGFSPFR	-2.72	30.05	2.20	6.80	-47.85	-47.71	-46.02
107	RVYIHPI	-6.29	30.05	1.95	5.55	-47.86	-47.41	-46.15
108	SYSMEHFRWG	-5.15	7.20	2.30	5.34	-47.88	-47.54	-44.80
109	RVYVHPF	-4.86	30.05	1.95	5.34	-47.92	-47.27	-45.34
110	FFFFF	-3.50	30.01	1.61	5.30	-47.92	-48.27	-46.11
111	RPGFSPFR	-2.72	30.05	2.08	6.25	-48.07	-46.91	-44.71
112	DRVYIHPFHLVIHN	-12.20	11.58	2.64	9.32	-48.10	-47.43	-45.25
113	WQPPRARI	-4.13	39.81	2.08	6.47	-48.11	-47.40	-45.77
114	RGPFPI	-2.73	30.05	1.79	4.68	-48.12	-47.14	-42.36
115	IARRHPYFL	-6.15	19.59	2.20	7.75	-48.21	-47.30	-44.58
116	WHWLQL	-5.08	39.81	1.79	4.40	-48.28	-48.46	-45.62
117	PPGFSPFR	-2.10	13.95	2.08	5.23	-48.42	-47.61	-45.54
118	YGGFMRF	-3.15	31.83	1.95	4.75	-48.49	-47.09	-44.86
119	EGKRPWIL	-5.17	16.23	2.08	6.58	-48.57	-47.47	-45.87
120	WWW	-1.98	39.81	1.10	3.03	-48.66	-48.09	-45.97
121	HW	-1.32	23.79	0.69	1.45	-40.97	-38.71	-31.77
122	RRPYIL	-4.79	30.05	1.79	6.04	-32.39	-29.11	-22.96
123	YPFVEPI	-4.72	31.83	1.95	4.62	-32.39	-29.11	-22.96
124	YLEPGPVTA	-3.98	31.83	2.20	3.61	-18.05	-16.71	-14.09
125	RKDVY	-3.81	30.05	1.61	4.24	-42.31	-40.11	-33.63
126	RFDS	-2.38	30.05	1.39	2.96	-31.80	-28.57	-19.86
127	FLEEI	-4.79	30.01	1.61	3.79	-27.60	-27.03	-25.53
128	VEPIPY	-4.02	14.95	1.79	3.56	-14.58	-14.09	-12.70

2.3. Sequential forward search for input selection using ANFIS

In this work, ANFIS-based sequential variable selection program written in MATLAB is used as a feature selection method [21]. The algorithm is based on selecting the best descriptor which minimizes standard errors of calibration and prediction and then repeatedly adds next best descriptor to the previous one(s). In the first step after sorting the dataset based on the mobility values (Table 1), training and test sets in a ratio of 4:1 were randomly chosen such that the test set adequately represented the training set. Then based on three iterations and two Gaussian bell membership functions, 5 out of 41 descriptors were selected using ANFIS. The selected parameters were used as inputs for developing ANN models. Analysis of the results obtained by the ANN model showed some outliers. These outliers were removed from the original dataset and the sequential variable selection was repeated using the remaining peptides of the dataset. In the final stage, a total of four descriptors were selected for developing neural networks and further studies. The values of these parameters are given in Table 1 for all peptides studied in this work.

2.4. Descriptors

The following structural parameters for amino acids were considered for calculating the descriptors of peptides: The substituent constants (κ), steric effects (ES,C) and molecular refractivity (MR) [22,23]. The values of these descriptors for twenty amino acids are listed in Table 2. Molar refractivity and residues mass are scaled by a factor of 0.1, such that they will have the same scale according to Hansch and Leo [18]. Taft defined the steric constant, ES, as log (k/ko), where k and ko are the rate constants for the acidic hydrolysis of a substituted ester and of a reference ester (methyl group is usually used as the reference, but H is sometimes used), respectively [24]. Hankcock has stated that there is contribution of hyper conjugation (α-hydrogen bonding) to the Taft Es; therefore, it must be corrected as defined by [25]:

$$ES,C = ES + 0.306 (n-3) \qquad (11)$$

where ES,C is a corrected steric substituent constant and Es is the "revised" Taft steric constant[26]; n is the number of α-hydrogens. As can be seen in Table 2, a small value of steric effect was observed for "crowded" structures of α-branched side chain (V, I, L). This means that they are large resistance to the hydrolysis.

Table 2. Physicochemical substituent parameters for amino acids

AA Side Chains	K	$E_{s,c}$	MR
alanine (A)	0.21	0.00	0.57
arginine (R)	1.57	-0.62	3.01
asparagine (N)	-0.18	-0.78	1.45
aspartic acid (D)	0.44	-0.78	1.16
cysteine (C)	1.28	0.00	0.00
glutamine (Q)	0.06	-0.62	1.91
glutamic acid (E)	0.42	-0.62	1.62
glycine (G)	0.00	0.20	0.10
histidine (H)	0.44	-0.66	2.38
isoleucine (I)	0.95	-1.61	1.96
leucine (L)	0.94	-1.24	1.96
lysine (K)	1.14	-0.62	2.51
methionine (M)	0.60	-0.83	2.31
phenylalanine (F)	1.06	-0.70	3.00
proline (P)	0.55	0.00	1.40
serine (S)	-0.11	-0.28	1.18
threonine (T)	-0.15	-0.53	1.18
tryptophan (W)	1.01	-0.66	3.98
tyrosine (Y)	0.46	-0.70	3.18
valine (V)	0.63	-1.09	1.50

2.5. Artificial neural network

In the present work, the feed forward back propagation of error artificial neural network (BP-ANN) is written in C++. The input layer consisted of the four parameters selected by ANFIS. The output layer represents the electrophoretic mobilities of the peptides. In this investigation, the bipolar sigmoid function, i.e., f (x) = (1-exp (-x)) / (1+exp (-x)), is used as the transfer function. The initial weights were chosen randomly, and were optimized based

on the delta rule through back propagation of errors. The program is written in such a way that the range for initialization of the weights depends on the number of input and hidden nodes. Before training, the inputs were normalized between -2 and 2 and the output between 0 and 1. The network parameters such as the number of hidden layer nodes, learning rate and momentum were optimized. Optimizations of these parameters were based on obtaining the minimum standard error of calibration and prediction. Program automatically avoids overfitting by stopping the training when the increase in standard error of prediction commences. After the analysis of ANN results and removing the outliers, the new dataset consisted of 118 peptides were sorted based on mobility values. This set was divided into training, test and validation sets (in a ratio of 4:1:1). However, in order to test the stability of the model and making sure that the results are not due to the chance, six different sets of training, test and validation sets for each concentration of SDS were created.

2.6. Methodology

The present work consists of three steps: (1) experimental determination of the mobilities of peptides using CE system in 40, 60 and 80 mM SDS solutions for two purposes. First, to investigate the effects of change of surfactant concentration on the mobilities of peptides and secondly, exploration of the ability and robustness of the generated theoretical models in the prediction of the MEKC mobilities at different SDS concentrations. (2) Selecting the structural parameters which play the major role in the migration behavior of peptides in MEKC experiments. This is a challenging process, since the mechanism of partitioning of the peptides into the micelles and migration of the micelles due to the electrophoretic and electroosmotic phenomena are complex. In modeling, choosing suitable features/descriptors is critical, because without success in this step the development of a robust and interpretable model is impossible. Therefore, we were very anxious to search for a powerful method as a feature selection technique. We have chosen

a neurofuzzy system for this purpose, which is a combination of the neural network and fuzzy rules. A neural network can model a process by means of a linear/nonlinear regression algorithm, for which the result is a network with adjusted weights and approximates the property of interest. However, the problem is that the knowledge is stored in an opaque fashion; the learning result is a set of parameter values, almost impossible to interpret them in words. Conversely, a fuzzy rule-base consists of readable if-then statements which are very close to natural language, but cannot learn the rules. These two are combined in neurofuzzy systems in order to achieve readability and learning ability at the same time. In this work a sequential ANFIS is used as feature selection technique. We have chosen ANFIS because of its much faster convergence, much more repeatability and much less preprocessing compared with ANN. (3) In order to develop a model for predicting the MEKC mobilities of peptides and also inspecting the linear/nonlinear characteristics of the migration behavior of peptides in MEKC, simple MLR as a linear method and BP-ANN as a nonlinear method are used. In both cases we use ANFIS for selecting the features. These methods are very common and frequently have been used in our laboratory and by several other researchers [12-16]. Therefore, for the sake of brevity their description is not given here.

2.6.1. Adaptive Neuro-Fuzzy Inference System (ANFIS)

By definition fuzzy logic is a type of logic that recognizes more than simple true and false values [17]. Fuzzy logic can represent propositions by degrees of truthfulness and falsehood and has proved to be particularly useful in expert systems and other artificial intelligence applications. One of the hallmarks of fuzzy logic is that it allows nonlinear input/output relationships to be expressed by a set of qualitative "if-then" rules. The theory of fuzzy logic provides a mathematical morphology to emulate certain perceptual and linguistic attributes associated with human cognition. Most fuzzy systems are hand-crafted by a human expert to

capture some desired input/output relationships that the expert has in mind. However, often an expert cannot express his or her knowledge explicitly; and, for many applications, an expert may not even exist. Hence, there is considerable interest in being able to automatically extract fuzzy rules from experimental input/output data. While fuzzy theory provides an inference mechanism under cognitive uncertainty, computational neural networks offer exciting advantages such as learning, adaptation, fault-tolerance, parallelism and generalization. The computational neural networks are capable of coping with computational complexity, nonlinearity and uncertainty. In fact, the neural network approach fuses well with fuzzy logic and by combining these two techniques, benefits of both would be acquired. Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic (Fig. 1). The mapping then provides a basis from which decisions can be made, or patterns discerned. One type of fuzzy inference systems is based on Takagi-Sugeno model [18]. In the Takagi-Sugeno model the idea is that each rule in a rule-base defines a region for a model, which can be linear. The left-hand side of each rule defines a fuzzy validity region for the linear model on the right-hand side. The inference mechanism interpolates smoothly between each local model to provide a global model. As an example consider a single input, single output system with the following two rules: 1) IF input is large THEN output is line 1. 2) IF input is small THEN output is line 2. Where line 1 is defined as 0.2 * input + 90 and line 2 is 0.6 * input + 20. The rules interpolate between the two lines in the region where the membership functions overlap (Fig. 2). Outside of that region the input is a linear function of the error.
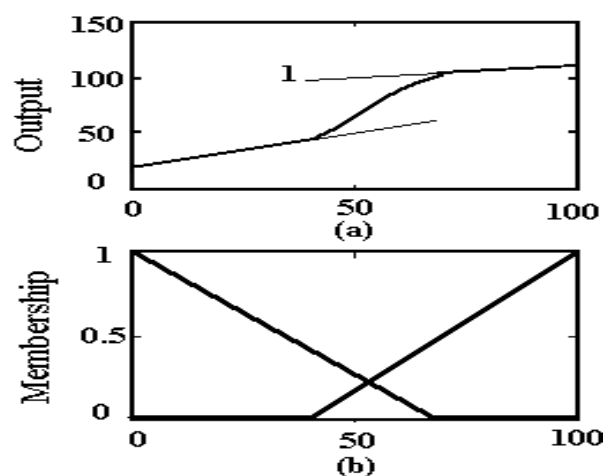

Fig. 2. Interpolation between two lines (top) in the overlap of input sets (bottom).

2.6.2.ANFIS architecture

Without loss of generality we assume two inputs, u1 and u2, and one output, y. Assume for now a first order Sugeno type rule-base composed of the following two rules [19]:

If u_1 is A_1 and u_2 is B_1 then

$$y_1 = c_1 u_1 + c_2 u_2 + c_0 \qquad (1)$$

If u_1 is A_2 and u_2 is B_2 then

$$y_2 = c_2 u_1 + c_2 u_2 + c_0 \qquad (2)$$

Incidentally, this fuzzy controller could interpolate between two linear controllers depending on the current state. If the firing strengths of the rules are α1 and α2 respectively, for two particular values of the inputs u1 and u2, then the output is computed as a weighted average

$$y = \frac{\alpha_1 y_1 + \alpha_2 y_2}{\alpha_1 + \alpha_2} = \overline{\alpha}_1 y_1 + \overline{\alpha}_2 y_2 \qquad (3)$$

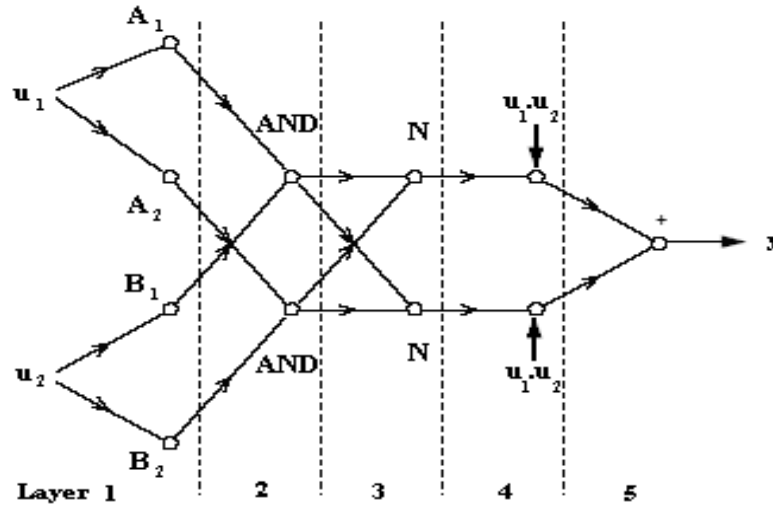**Fig. 1.** The basic structure of fuzzy inference system

Fig. 3. Structure of the ANFIS network

Figure 3 shows corresponding ANFIS network. The descriptions for the layers shown in the network are as follows:

1. Each neuron i in layer one is adaptive with a parametric activation function. Its output is the grade of membership to which the given input satisfies the membership function, i.e.,

$$\mu_{A_1}(u_1) \; \mu_{B_1}(u_2) \; \mu_{A_2}(u_1) \; \text{or} \; \mu_{B_2}(u_2).$$

An example of a membership function is the generalized bell function:

$$\mu(x) = \frac{1}{1 + \left| \dfrac{x - c}{a} \right|^{2b}} \tag{4}$$

where {a, b, c} is the parameter set. As the values of the parameters change, the shape of the bell-shaped function varies. Parameters in that layer are called premise parameters.

2. Every node in layer two is a fixed node, whose output is the product of all incoming signals. In general, any other fuzzy AND operation can be used. Each node output represents the firing strength αi of the ith rule.

3. Every node in layer three is a fixed node which calculates the ratio of the ith rule's firing strength relative to the sum of all rule's firing strengths,

$$\overline{\alpha}_i = \frac{\alpha_i}{\sum\limits_{i=1,2} \alpha_i} \tag{5}$$

The result is a normalized firing strength.

4. Every node in layer four is an adaptive node with a node output

$$\overline{\alpha}_i y_i = \overline{\alpha}_i (c_{i1} u_1 + c_{i2} u_2 + c_{i0}) \quad i = 1,2 \tag{6}$$

where i is the normalized firing strength from layer three and {ci1, ci2, ci0} is the parameter set of this node. Parameters in this layer are called consequent parameters.

5. Every node in layer five is a fixed node which sums all incoming signals. It is straightforward to generalize the ANFIS architecture in Figure 3 to a rule-base with more than two rules.

2.6.3. The ANFIS learning algorithm

When the premise parameters are fixed, the overall output is a linear combination of the consequent parameters. In symbols, the output y can be written as:

$$y = \frac{\alpha_1}{\alpha_{1+}\alpha_2} y_1 + \frac{\alpha_2}{\alpha_{1+}\alpha_2} y_2 \tag{7}$$

$$y = \overline{\alpha}_1(c_1 u_1 + c_2 u_2 + c_0) + \overline{\alpha}_2(c_1 u_1 + c_2 u_2 + c_0) \tag{8}$$

$$y = (\overline{\alpha}_1 u_1)c_1 + (\overline{\alpha}_1 u_2)c_2 + \overline{\alpha}_1 c_0 + (\overline{\alpha}_2 u_1)c_1 + (\overline{\alpha}_2 u_2)c_2 + \overline{\alpha}_2 c_0 \tag{9}$$

Which is linear in the consequent parameters cij (i=1, 2; j=0, 1, 2). A hybrid algorithm adjusts the consequent parameters cij in a forward pass and the premise parameters {ai, bi, ci} in a backward pass. In the forward pass the network inputs propagate forward up to layer 4, where the consequent parameters are identified by the least-squares method. In the backward pass, the error signals propagate backwards and the premise parameters are updated by gradient descent. Because the update rules for the premise and consequent parameters are decoupled in the hybrid learning rule, a computational speedup may be possible by using variants of the gradient method or other optimization techniques on the premise parameters.

3. Results and discussion

The main goal of this work was to study the mechanism of the migration of peptides in MEKC. We hope that the results of this work together with our previous works on CZE could pave the way for further studies on the 2D MEKC-CZE simulations. However, the best way of studying the mechanism is gathering a set of the general parameters which are responsible for the migration of the peptides in MEKC. To achieve this, the mobility of a set of different classes of peptides has to be modeled. Mobilities of a set of 128 peptides composed of up to 14 amino acids in 40, 60 and 80 mM solutions of SDS was measured. The general strategy of modeling the mobilities was as follows: Feature selection using sequential ANFIS algorithm; developing MLR and BP-ANN models using the selected descriptors as the inputs; and analysis and evaluation of the best model.

3.1. *Application of ANFIS as feature selection method*

After obtaining the values of the mobilities of 128 peptides using MEKC method and calculating 41 descriptors for each one, the dataset was divided into training (102 peptides) and test sets (26 peptides). Then by applying ANFIS, five descriptors of Kappa(H), ln(N), ARM, MRC and ES,N which minimized the standard errors of

calibration and prediction were selected. These descriptors were used as the inputs to develop a network for modeling the mobility of peptides. The results of the ANN showed that the network is not able to predict accurately the mobility of ten peptides and therefore they can be considered as statistical outliers. These peptides were DF, DLFL, DW, WD, EW, WE, FLEEI, VEPIPY, YPFVEPI and YLEPGPVTA. This means that these peptides have different characteristics compared with the rest of the dataset, and the ANN model is unable to learn and predict their behavior. Consequently, due to the special characteristics of the above-mentioned peptides there is the possibility of misleading the ANFIS and this method may model the noise. Hence, these peptides were removed from the dataset and selection of the features was repeated. Finally four descriptors (Kappa (H), ln (N), MRn and ES,C) which could model the mobility were chosen. The behavior of the outliers will be discussed later in this section.

3.2. *Modeling and prediction by Artificial Neural Network*

The investigations were started using SDS 80 mM dataset. Even though we had removed the ten outliers in feature selection step, in order to study the results of ANN calculations, we added them to the dataset again. Then the dataset was divided into training, test and validation sets. The network was trained and the results were studied. Results showed that some peptides cause instability and premature training of the network. Therefore, the outliers were removed one at a time and entered into the validation set in order to study their behavior when the network is completely trained with the remaining peptides. It is obvious that after removing the outliers the remaining peptides were divided into new training, test and validation sets. These sets are composed of random and representative samples, but the validation set encompasses the entered samples as well. In the final stage eight statistical outliers were obtained which were HW, RRPYIL, RFDS, RKDVY FLEEI, VEPIPY, YPFVEPI and YLEPGPVTA.
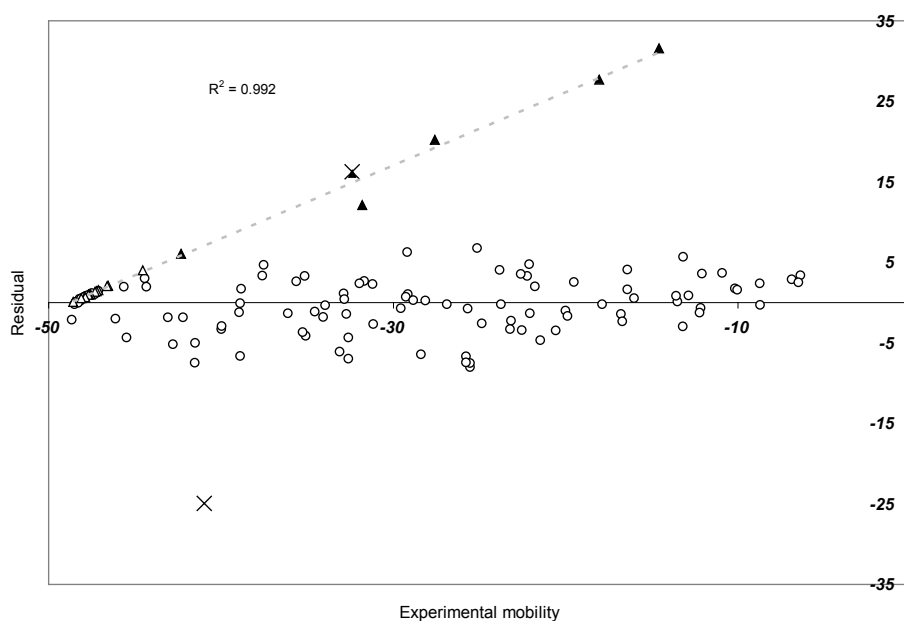
Fig. 4. Residuals of calculated ANN values of mobilities versus the experimental values. Δ, peptides containing negatively charged middle amino acids (E and D); $\times$, HW and RRPYIL; $\blacktriangle$, outliers

3.3. Analysis of residuals

Figure 4 depicts the residuals of ANN calculated values versus the experimental mobilities. It can be seen from this figure that the developed ANN is not able to predict accurately the mobilities of HW and RRPYIL. Also it can be seen that the residuals of all of the outliers are located above the zero axis. Statistically this means that there should be a systematic error in the calculated results of the ANN model. However, inspection of the residuals reveals that the peptides containing middle E and D amino acids together with the six outliers lie on a line with a correlation of R2= 0.992. This implies that the mobility of these peptides follows a similar mechanism which is different from the mechanism for the remaining peptides. Presumably a parameter appropriate for accounting the influence of charge is missing. Fundamentally a charge descriptor should be able to introduce this characteristic to the model. However, due to the small number of these type of peptides in the dataset (18 peptides), the network was not being able to receive enough information to learn their behavior. On the other hand, positive values of the residuals show that the ANN model

overestimates the mobilities of these peptides. The repulsion between the negatively charged D and E amino acids and the anionic SDS surfactant could be responsible for this overestimation. Because of this repulsion, these peptides spend a shorter time in the micellar phase and move slower. Consequently, one expects that in more dilute solutions of SDS, this effect be more pronounced.

3.4. Effect of SDS concentration on peptide mobilities

In order to investigate the effect of concentration on the migration of peptides, in addition to the original model which was developed based on the 80 mM SDS solution, two other models were developed using the same descriptors and same settings of the network for 40 and 60 mM SDS solutions. Figure 5 presents the predicted results (validation set) versus the experimental values for 80, 60 and 40 mM SDS solutions. Inspection of the figures reveals that by decreasing the SDS concentration the spread of points around the correlation line has increased. Therefore, a question arises regarding the effect of the SDS

concentration on the mobility behavior of the peptides. To explain the grounds, we refer to Figures 6 which are obtained using purely experimental values. These figures show the trend of mobility changes due to changes in SDS concentration. The vertical axis shows the change in mobility for two solutions of SDS and the horizontal axis is the mobility of the more concentrated one. By inspection of these figures one may conclude that: (1) Concentration increments have a profound effect on the curvatures of the mobility trends. Figure 6C with the highest concentration gradient shows a high curvature. On the other hand, although the concentration gradients are equal in figures 6A and 6B (20mM), Figure 6A shows a very small curvature compared with Figure 6B. This shows that 40 and 60 mM solutions are below the optimum level of SDS concentration. This may be due to the fact that when the solution is more dilute the effect of micelles is less profound, and CZE mechanism prevails over MEKC. Since the ANN model is trained based on the MEKC data, the calculated values for the more dilute solutions show a broader spread. (2) The compounds which include E and D amino acids behave different in comparison with the other peptides. This is more pronounced for the dipeptides. Points marked with hollow squares in Figures 6 belong to FLEEI, DW, WD, WE, EW, DF and DLFL. The figures show that these peptides do not obey the general
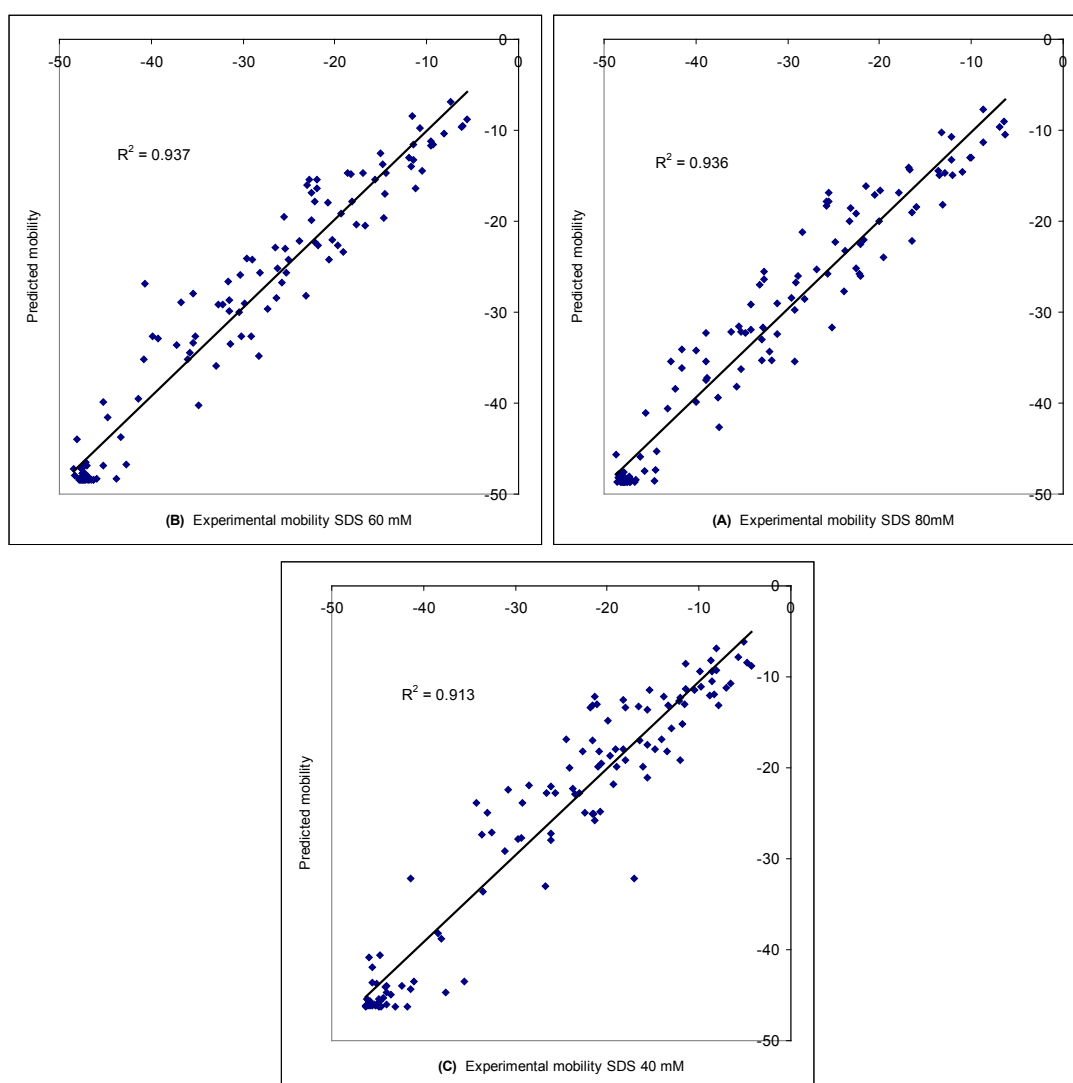

Fig. 5. Predicted values obtained by ANN model for the all 6 batches. 80 mM, (B) 60 mM and (C) 40 mM SDS solutions.

trend of mobility changes which exists for the other peptides. FLEEI is the only peptide in the dataset which contains two E amino acids, so it would suffer more of the repulsion forces which exists between E amino acids and the anionic surfactant. The footprint of this interaction exits in the outliers of the ANN (RFDS, RKDVY, FLEEI, VEPIPY, YPFVEPI and YLEPGPVTA), and ANFIS (DF, DLFL, DW, WD, EW, WE, FLEEI, VEPIPY, YPFVEPI and YLEPGPVTA) as well.

3.5. Comparison of MLR and ANN results

To investigate the linear/nonlinear characteristics of the relation between mobility and the descriptors, a similar MLR model was developed. For a meaningful comparison, both the ANN and MLR methods has to be trained using the same training set and verified by the same validation set. Despite the fact that we have used all of the peptides (outliers of the ANN model were excluded) in the regression step, the MLR calculated results for the training set show a poor correlation of 0.71. This demonstrates the inadequacy of MLR method for the modeling of the peptide mobilities, and irrational trend of residuals (Fig. 7). This could be due to nonlinear characteristics of the mobilities. Such a trend is absent in ANN residuals (Fig. 4). In order to assess the role of each variable in nonlinear characteristics of the peptide mobilities, the MLR residuals for the variables are depicted in Figure 8. These figures suggest that, kappa could be the parameter responsible for the nonlinearity, because the trend in its residuals is very similar to the trend of the MLR residual plot.

3.6. Robustness of the ANN models

After the training of the ANN for the prediction of peptide mobilities in different concentrations of SDS, the outliers were expelled and the remaining peptides were divided into six different batches of training, test and validation sets. These batches were chosen in a way that in each one every peptide appeared in the test and validation sets once. Then all six batches of training, test and validation sets
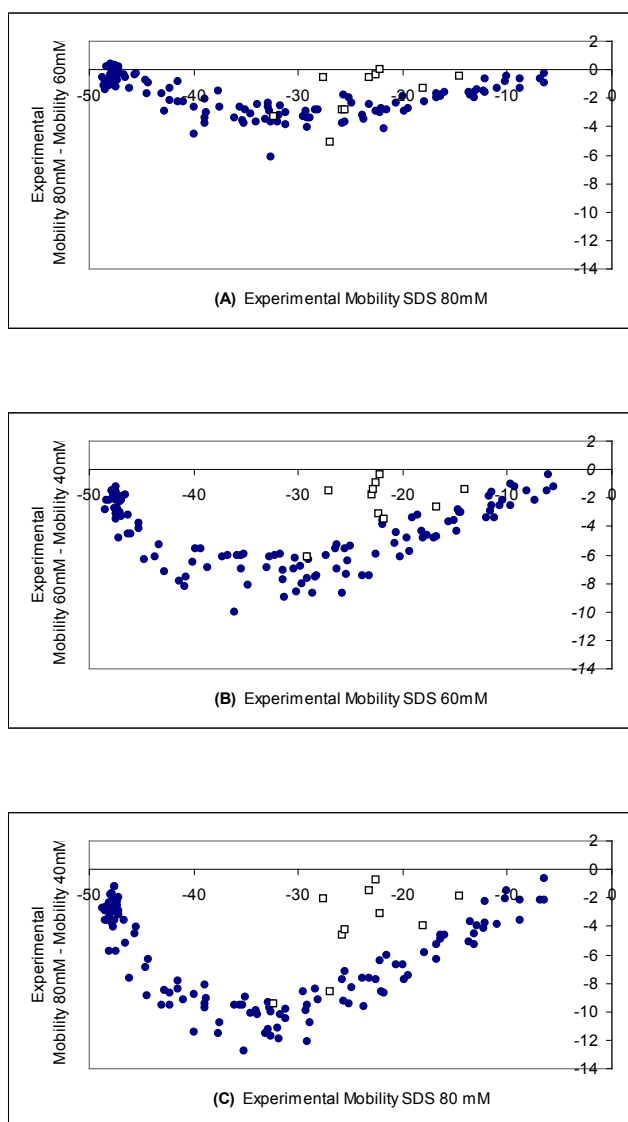

Fig. 6. Change in mobility due to SDS concentrations changes.
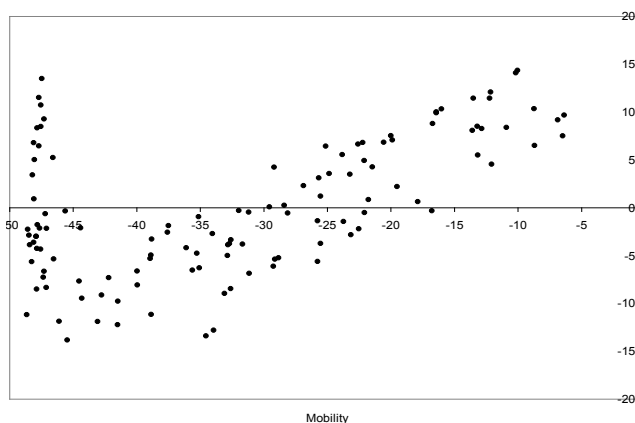the peptides with E and D amino acids

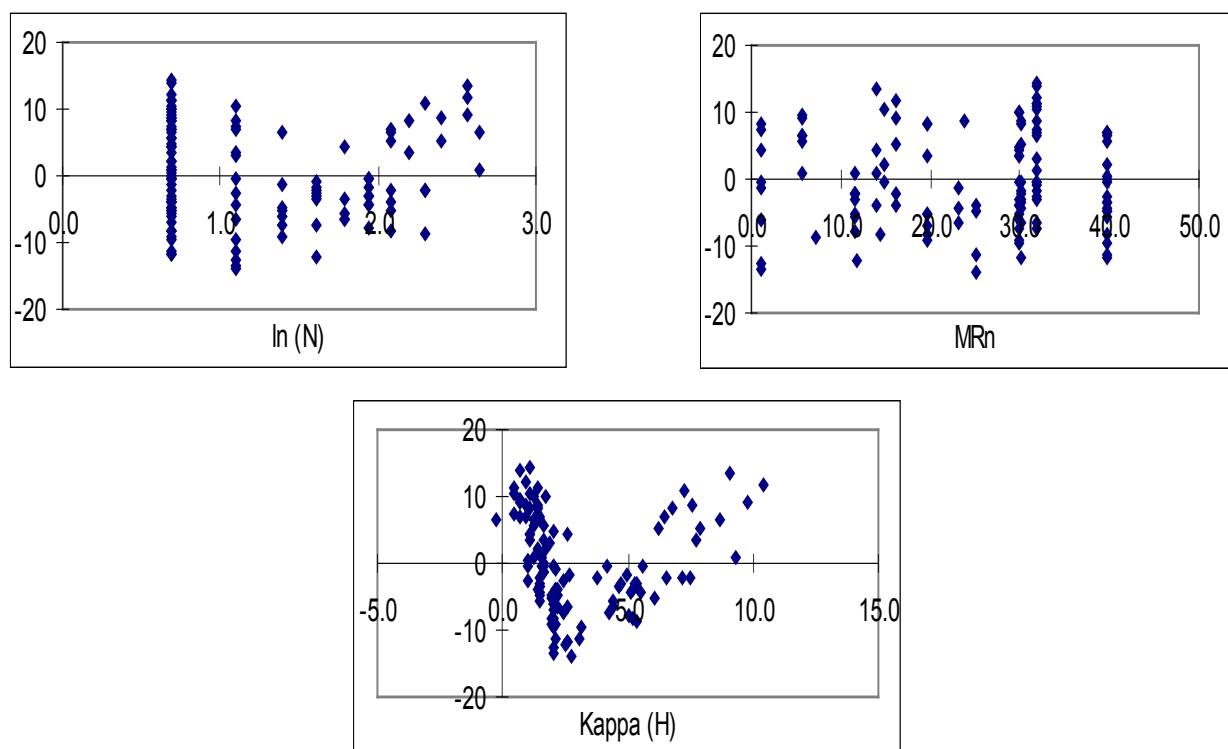**Fig. 7.** MLR prediction residual plot.

Fig. 8. MLR variables residual plots

were trained and applied for predicting the peptide mobilities in solutions with different concentrations of SDS. It can be seen from Figures 5 and Tables 3 and 4 that the residuals are promising for all of the six batches. Table 3 is devoted to correlation values for training, test and validation sets of each batch. The values for the correlations were in the range of 0.849 to 0.969 for all batches with different concentrations of SDS, which show the robustness of the model. Table 4 shows the average deviation (AD), average absolute deviation (AAD) and standard deviation (SD) of the ANN predicted values (validation sets) which have been calculated through equations 12-14.

$$AD = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right) \quad (12)$$

$$AAD = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \quad (13)$$

$$SD = \sqrt{\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2\right)} \quad (14)$$

in these equations y_i are calculated mobilities, $\hat{y}_i$ represents experimental mobilities and n is the number of samples of the set. In these calculations we have not considered the outlier peptides of the validation sets which were outliers. The small values of the deviations reveal the lack of systematic errors in the model. It is noteworthy that the SD shows a range of 3.410 to 4.040 which is close to the experimental errors. These deviations also confirm the predictive ability and robustness of the model.

4. Conclusion

A long-range goal of our laboratory is the development of experimental and theoretical methods for peptide separations and mapping in two-dimensional MEKC-CZE scheme. We have considered the specifications of simplicity, accuracy and robustness of the models in predicting the CZE and MEKC mobilities of the peptides. In our previous works [9-10], we showed the ability of the artificial neural networks in modeling of the CZE mobilities. This paper focuses on MEKC with more complicated mechanism compared to the CZE. Adaptive neuro-fuzzy inference system (ANFIS) was successfully used to select the most

Table 3. Statistical correlations using the ANN model for six different batches

	batch I	batch II	batch III	batch IV	batch V	batch VI	average
80 mM SDS							
Training	0.956	0.952	0.951	0.964	0.951	0.949	0.954
test	0.955	0.969	0.941	0.924	0.947	0.947	0.947
prediction	0.921	0.943	0.947	0.943	0.963	0.939	0.943
60 mM SDS							
Training	0.964	0.953	0.956	0.962	0.951	0.952	0.956
test	0.956	0.969	0.951	0.913	0.953	0.953	0.949
prediction	0.900	0.953	0.954	0.958	0.956	0.947	0.945
40 mM SDS							
Training	0.937	0.941	0.948	0.953	0.948	0.945	0.945
test	0.953	0.961	0.932	0.900	0.942	0.924	0.935
prediction	0.849	0.944	0.924	0.928	0.954	0.931	0.922

Table 4. Statistical deviations for the ANN predicted values of the peptide nobilities

Prediction[a]	batch I	batch II	batch III	batch IV	batch V	batch VI	average
80 mM SDS							
AD	-1.143	-0.361	-1.712	-0.489	0.057	1.144	-0.417
AAD	3.467	2.527	2.536	2.283	1.935	2.879	2.605
SD	4.208	3.262	3.652	3.257	2.477	3.605	3.410
60 mM SDS							
AD	-1.754	0.029	-1.353	-0.769	-0.093	1.126	-0.469
AAD	3.871	2.542	2.383	2.174	2.196	2.693	2.643
SD	5.076	3.114	3.377	2.937	2.795	3.413	3.452
40 mM SDS							
AD	-0.222	-0.266	-1.724	-0.351	0.255	1.105	-0.201
AAD	4.299	2.610	3.016	2.696	2.312	2.989	2.987
SD	5.745	3.521	4.402	3.771	2.949	3.853	4.040

[a] The notations AD, AAD and SD stand for average deviation, average absolute deviation and standard deviation, respectively.

appropriate variables as ANN inputs. It is shown that except for the peptides including negatively charged amino acids the model holds promise for application in predicting the peptide mobilities in MEKC systems. However, researches are underway in our laboratory to combine the CZE and MEKC models to map the peptides in 2D CZE/MEKC scheme.

5. Acknowledgement

6. References

[1] A.H. Rageh, U. Pyell, Pseudostationary ion-exchanger" sweeping as an online enrichment technique in teh determination of nucleosides in urine via micellar electrokinetic chromatography, Chromatogra., 82 (2019) 325–345.

[2] A.H. Rageh, U. Pyell, Imidazolium-based ionic liquid-type surfactant as pseudostationary phase in micellar electrokinetic chromatography of highly hydrophilic urinary nucleosides, J Chromatogr., A 1316 (2013)135–146.

[3] R.B. Yu, J. P. Quirino, Chiral Selectors in capillary electrophoresis: trends during 2017–2018, Molecules, 24 (2019) 1135.

[4] J. Fiori, B. Pasquini, C. Caprini, S. Orlandini, S. Furlanetto, R. Gotti, Chiral analysis of theanine and catechin in characterization of green tea by cyclodextrin-modified micellar electrokinetic chromatography and high performance liquid chromatography, J. Chromatogr. A, 1562 (2018) 115–122.

[5] Y. Liu, X. Wang, Enantioseparation of ofloxacin and

its four related substances with ligand exchange–micellar electrokinetic chromatography using copper (II)-L-isoleucine complex as chiral selector, Chirality., 29 (2017) 422–429.

[6] I.J. Stavrou, E.A. Agathokleous, C.P. Kapnissi-Christodoulou, Chiral selectors in CE: Recent development and applications, Electrophoresis, 38 (2017) 786–819.

[7] V. Patel , S. A. Shamsi, Carbohydrate based polymeric surfactants for chiral micellar electrokinetic chromatography (CMEKC) coupled to mass spectrometry, Methods Mol. Biol., 1985 (2019) 417–444.

[8] Y. Liu, S.A. Shamsi, Development of novel micellar electrokinetic chromatography mass spectrometry for simultaneous enantioseparation of venlafaxine and dimethyl-venlafaxine: Application to analysis of drug-drug interactions, J. Chromatogr. A, 1420 (2015) 119–128.

[9] R. L. C. Voeten, I. K. Ventouri, R. Haselberg, G. W. Somsen, Capillary electrophoresis: trends and recent Advances, Anal Chem., 90 (2018) 1464–1481.

[10] C.-R. Chung, J.-H. Jhong, Z. Wang, S. Chen, Y. Wan, Characterization and identification of natural antimicrobial peptides on different organisms, Int. J. Mol. Sci., 21(2020) 986. doi:10.3390/ijms21030986

[11] Fausett, L, Fundamentals of neural networks, architectures, algorithms, and applications, Prentice-Hall, inc., New Jersey, 1994.

[12] M. Sugimoto, S. Kikuchi, M. Arita, T. Soga, T. Nishioka, M. Tomita, Large-scale prediction of cationic metabolite identity and migration time in capillary electrophoresis mass spectrometry using artificial neural networks, Anal. Chem., 77 (2005) 78-84

[13] K. Shinoda, M. Sugimoto, N. Yachie, N. Sugiyama, T. Masuda, M. Robert, T Soga, M. Tomita, Prediction of liquid chromatographic retention times of peptides generated by protease digestion of the Escherichia coli proteome using artificial neural networks, J. Proteome Res., 5 (2006) 3312-3317.

[14] K. Khan, K. Roy, Ecotoxicological modelling of cosmetics for aquatic organisms: A QSTR approach, SAR QSAR Environ. Res., 28 (2017) 567-594.

[15] F. Yang, J. Tian, Y. Xiang, Z. Zhang, P. de B. Harrington, Near infrared spectroscopy combined with least squares support vector machines and fuzzy rule-building expert system applied to diagnosis of endometrial carcinoma, Cancer Epidemiol., 36 (2012) 317-323.

[16] R. Darnag, B.Minaoui, M. Fakir, QSAR models for prediction study of HIV protease inhibitors using support vector machines, neural networks and multiple linear regression, Arab. J. Chem., 10 (2017) S600-S608.

[17] G.J. Klir, B.Yuan, Fuzzy sets and fuzzy logic, theory and applications. Prentice-Hall, inc., New Jersey, 1995.

[18] H. Zarei, A. Khastan, A. Zafar, Optimal control of linear fuzzy time-variant controlled systems, Iran. J. Fuzzy Sys., 17 (2020) 1-12.

[19] J. Jantzen, Neurofuzzy modeling technical report no 98-H-874, 1998.

[20] M. G. Khaledi, Micellar electrokinetic chromatography in high performance capillary electrophoresis: theory, technique and applications, Wiley, New York, 1998.

[21] Fuzzy toolbox; Copyright 1994-2002, the math works, Inc. revision: 1.8; Roger Jang, Aug. 1997.

[22] J. Liu, H. Wang, LingZhi oligopeptides amino acid sequence analysis and anticancer potency evaluation, RSC Adv., 10 (2020) 8377-8384.

[23] C. Hansch, A. Leo, D. Hoekman, Exploring QSAR, hydrophobic, electronic, and steric constants, ACS, Washington DC., 1995.

[24] R. W. Taft, Steric effects in organic chemistry, John Wiley and Sons Inc.; New York, 1956.

[25] S. S. Samanta, S. P. Roche, Synthesis and reactivity of α-haloglycine esters: hyperconjugation in action, Eur. J. Inorg. Chem., 2019 (2019) 6597-6605.

[26] J.A. Macphee, A. Panaye, J.E. Dubois, Steric effects-I: A critical examination of the taft steric parameter-Es, definition of a revised, broader and homogeneous scale, extension to highly congested alkyl groups, Tetrahedron., 34 (1978) 3553-3562