Carbohydrate Polymers xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Carbohydrate Polymers

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

A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks

Jamshed Akbar[a], Mohammad S. Iqbal[b,*], Muhammad T. Chaudhary[c,d], Tallat Yasin[e], Shazma Massey[b,f]

[a] Department of Chemistry, University of Sargodha, Sargodha 40100, Pakistan
[b] Department of Chemistry, Forman Christian College, Lahore 54600, Pakistan
[c] Department of Pharmacy, University of Sargodha, Sargodha 40100, Pakistan
[d] Standpharm Pakistan (Pvt.) Ltd, Lahore 54600, Pakistan
[e] Department of Chemistry, Punjab College for Women, Sargodha 40100, Pakistan
[f] Department of Chemistry, GC University, Lahore 54000, Pakistan

ARTICLE INFO

Article history:
Received 20 January 2012
Received in revised form 2 February 2012
Accepted 7 February 2012
Available online xxx

Keywords:
Arabinoxylan matrix
Controlled drug delivery
Quantitative–structure–property relationship
Density functional theory
Heuristic method
Neural networks

ABSTRACT

A QSPR study on release of pharmacologically diverse drugs from a biocompatible matrix, arabinoxylan, by use of *ab initio* structure optimization and neural networks is reported. A total of 1685 quantum mechanical, physico-chemical and structural descriptors were calculated for 16 drug molecules. A heuristic approach combined with unsupervised forward selection was used to identify descriptors mechanistically related to response variables. The release models were developed using multiple linear regression (MLR) and artificial neural networks (ANN) and were validated by leave-one-out cross validation and y-scrambling techniques. The release was found to be controlled by softness, lipophilicity, unsaturation, atomic polarization, cyclic topology and geometry of the molecules. The quantitative–structure–property relationship (QSPR) models were found to be robust and highly predictive of release profile and mechanism of a drug molecule from the arabinoxylan matrix.

1. Introduction

Arabinoxlans (AXs) are polysaccharides isolated from various sources like cereals such as wheat, rye, barley, oat, rice and corn and from some other plants like bamboo shoots and *Plantago* herbs. AXs from various sources are of similar nature having molar masses in different ranges. They are known to stabilize foams and emulsions and AXs from ispaghula (*Plantago ovata*) seeds and other sources have been reported to be the potential candidates for application in controlled delivery of drugs (Iqbal, Akbar, Hussain, Saghir, & Sher, 2011). They have been isolated and characterized from ispaghula seeds or husk by different methods (Iqbal, Akbar, Hussain, et al., 2011; Iqbal, Akbar, Saghir, et al., 2011) and are considered to be safe (Iqbal, Akbar, Hussain, et al., 2011) for use as drug carriers. The ispaghula seeds are abundantly available throughout the world and AXs can be isolated from them at a very low cost.

Several mathematical models have been developed to explain the drug release from various polymeric matrices, but none of them is universally applicable (Siepmann & Siepmann, 2008). Therefore, efforts are still going on to develop more robust models for the purpose. It is generally emphasized that release profiles are controlled by solubilities of drug molecules (Rujivipat & Bodmeier, 2010; Sumathi & Ray, 2002; Tahara, Yamamoto, & Nishihata, 1996), whereas, it has been demonstrated that drugs having similar solubilities can exhibit different release profiles from the same matrix (Baveja, Ranga Rao, Singh, & Gombar, 1988). This suggests that the release may be controlled by several factors including structural differences and physical properties of the drug and the polymer (Gafourian, Safari, Adibkia, Parviz, & Nokhodchi, 2007; Ranga-Rao, Padmalatha-Devi, & Buri, 1990; Tahara et al., 1996). Thus there is a need to develop predictive QSPR models which can facilitate the designing of drug delivery devices. The models should be robust and applied carefully as the mistakes can lead to erroneous results. We noticed that a previous study (Gafourian et al., 2007) consisted of some serious mistakes, such as: (i) the solubility and the experimental release profiles of several drugs were determined by using their hydrochlorides whereas the descriptors were calculated from the base molecules, (ii) the data have not been standardized for regression, and (iii) multicollinearity of the descriptors and validation of QSPRs were not duly taken into account.

The present work was carried out to study the release of pharmacologically diverse drug molecules from a biopolymer,

* Corresponding author. Tel.: +92 300 4262813.
E-mail address: saeediq50@hotmail.com (M.S. Iqbal).

0144-8617/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carbpol.2012.02.016

arabinoxylan (AX) isolated from ispaghula seeds, by QSPR methods, which has been demonstrated (Iqbal, Akbar, Hussain, et al., 2011; Niño-Medina et al., 2010) to be a useful alternate of hydroxypropylmethyl cellulose (HPMC).

2. Materials and methods

2.1. Materials

Arabinoxylan (prepared as reported by Iqbal, Akbar, Hussain, et al., 2011). Benzoic acid, BE (E. Merck), salicylic acid, SA (E. Merck), sodium hydroxide (Riedel-de Haën, Germany), potassium dihydrogen phosphate (Fluka, USA), caffeine, CAF (ServaFeinbiochemica GMBH, Heidelberg), atenolol (AT) and lamivudine, LAM (gifts from Novamed Pharmaceuticals, Lahore, Pakistan), flurbiprofen (FBI), mefenamic acid (MEF) and piroxicam, PIR (gifts from Quaper (Pvt.) Ltd., Sargodha, Pakistan), diclofenac (DA), famotidine (FAM), gliclazide (GLI), ibuprofen (IBU), lamotrigine (LAT), meloxicam (MEL), paracetamol (PAR), and ofloxacin, OF (gifts from Standpharm, Lahore, Pakistan) were used as received. Double-distilled water was used throughout this study.

2.2. Drug release studies

2.2.1. Preparation of tablets

Tablets were prepared by wet granulation technique using a wetting solution of AX (0.70%, w/v on dry basis in water). Powdered drug and the AX were passed separately through sieve no. 16. Drug and AX (1:1) were mixed, homogenized and the wetting solution (equivalent to 7.0%, w/w AX on dry basis) was applied uniformly. The mixture was homogenized and granulated by passing through sieve no. 16. The granules were dried at 40 °C for 6 h (to a moisture content of 8 ± 1%). The dried granules were mixed with magnesium stearate (1%, w/w) as lubricant and compressed under 10 kN force on a rotary tablet press, fitted with 8 mm flat punches, to give tablets with an average weight of 215 ± 3 mg, each containing 100 mg of the drug.

2.2.2. Dissolution study

Tablets were subjected to dissolution test in phosphate buffer (pH 7.4, 900 cm^3) using USP paddle dissolution apparatus II (Pharmatest, Germany) at 37 ± 0.1 °C and 50 rpm. Samples (2 cm^3 each) were withdrawn at predetermined intervals, appropriately treated, filtered and assayed spectrophotometrically using UV-1700 double beam spectrophotometer (Schimadzu, Japan). The spectrophotometric methods were validated by use of certified reference standards. An equal volume of the buffer was replaced immediately after the removal of the sample solution. Amount of drug released was expressed as percent of the total loaded drug.

2.2.3. Mathematical and statistical analysis of release profiles

Drug release data were analyzed using various models including zero order (Eq. (1)), first order (Eq. (2)) (Gibaldi & Feldman, 1967), Higuchi equation (Eq. (3)) (Higuchi, 1961, 1963) and power law (Eq. (4)) (Peppas, 1985; Ritger & Peppas, 1987a, 1987b).

$$M = k_0 t \tag{1}$$

where k_0 is the zero order release constant, M is the amount of drug released in time t.

$$\ln \quad M = -k_1 t + \ln \quad M_0 \tag{2}$$

where k_1 is the first order release constant, M is the remaining amount of drug in the tablet after time t and M_0 is the initial amount of drug in the tablet.

$$M = k_H\, t^{1/2} \tag{3}$$

where M is the amount of drug released in time t, k_H is the Higuchi release constant.

$$\ln \frac{M_t}{M_\infty} = \ln \quad k_p + n \quad \ln \quad t \tag{4}$$

where M_t/M_∞ is the fraction of drug released in time t, k_p is the power law constant characteristic of the drug-matrix system and n is the release exponent. The value of n identifies different mechanism for drug release. For different geometries the limits of n are different (Lopes, Sousa Lobo, Pinto, & Costa, 2007; Ritger & Peppas, 1987a, 1987b). For cylindrical systems (tablets), where $n = 0.45$, the drug release is termed as case-I transport (Fickian diffusion), for $n = 0.89$ it is case-II transport (swelling-controlled mechanism), for $0.45 < n < 0.89$, the mechanism is anomalous transport, *i.e.*, the release governed by swelling and diffusion phenomena and for $n > 0.89$, the mechanism is super case-II transport where the rate remains constant for a long time leading to an exponential drug release towards the later stages due to fast matrix erosion.

Model selection criterion (MSC) was used (Eq. (5)) (Scientist-handbook & rev. 7EEF, 1995) to select the best fit kinetic model for the drug release.

$$\mathrm{MSC} = \ln \left[\frac{\sum_{i=1}^{n} w_i (Y_{\mathrm{obs}_i} - \bar{Y}_{\mathrm{obs}})^2}{\sum_{i=1}^{n} w_i (Y_{\mathrm{obs}_i} - Y_{\mathrm{cal}_i})^2} \right] - \frac{2p}{N} \tag{5}$$

where Y_{obsi} and Y_{cali} are the observed and calculated value of ith data point respectively, $\bar{Y}_{obs}$ is the mean of observed data points, w_i the optional weight factor, N the number of data points and p the number of parameters. MSC is independent of the scaling of data points and the model with largest MSC value is the most appropriate. To see the similarity level of drug release profiles, hierarchical cluster analysis (HCA) was performed and dendrograms were drawn using weighted- pair-group average and Euclidean distance.

2.3. Development of QSPRs

2.3.1. Dissolution parameters

The dissolution parameters selected were amount of drug released after 12 h (M_{12}), rate constants (k_0, k_1, k_H, $\ln k_p$) from various kinetic models, and release exponent (n).

2.3.2. Calculation of structural descriptors

Structures of drug molecules were drawn by ChemSketch 12 and pre-optimized using PM6 Hamiltonian provided in semi-empirical software MOPAC2009 (Stewart, 2009). The geometries were finally *ab initio* optimized with Density Functional Theory (DFT) at B3LYP level using 6-31G(d,p) basis set as implemented in Firefly 7.1.G (Granovsky, 2010). The calculated descriptors are; electronic and quantum mechanical descriptors by DFT calculations, physico-chemical properties by ACD/I-lab 2.0 (available at http://ilab.acdlabs.com) and twenty logical blocks of structural descriptors by E-Dragon applet 1.0 (Tetko et al., 2005). The calculated descriptors are listed in Table 1.

2.3.3. Selection of descriptors and QSPR generation

A total of 1685 descriptors were calculated. The calculated molecular descriptors were collected in an $i \times j$ data matrix, where 'i' and 'j' are the number of rows and columns, representing molecules and descriptors, respectively. A heuristic method (HM) was employed to select the best descriptors describing a response. The procedure is detailed as follows. Initially, the descriptors were rejected for which value is not available for all the molecules. Next the descriptors with a constant value for 50% molecules and with variance less than 0.0001 were rejected because of inadequate information therein. In the next step, descriptors showing

Please cite this article in press as: Akbar, J., et al. A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks. *Carbohydrate Polymers* (2012), doi:10.1016/j.carbpol.2012.02.016

Table 1
List of generated and selected descriptors.

Generated descriptors	
Method/type	Descriptors
FIRELFY/DFT	Total energy, dipole moment, logarithm of dipole moment, energies of highest occupied molecular orbital (E_{HOMO}) and lowest unoccupied molecular orbitals (E_{LUMO}), difference of HOMO and LUMO, hardness ($\eta = 0.5 \times$ (LUMO–HOMO)) and softness ($1/\eta$)
ACD/ilab/physico-chemical	Molar volume, density, polarizability, pK_a, Log P, Log $D_{7.4}$ and solubility at pH 7.4, logarithms of molar volume, density, polarizability and solubility
DRAGON	Constitutional, topological, walk and path counts, connectivity indices, information indices, 2D autocorrelations, edge adjacency indices, eigenvalue-based indices, randic molecular profiles, geometrical, radial distribution functions (RDF), three-dimensionalmolecule representation of structures based on electron diffraction (3D-MoRSE), weighted holistic invariant molecular (WHIM), geometry, topology and atomic weights assembly (GETAWAY), functional group counts, atom centered fragments, charge and molecular properties
Descriptors selected by heuristic method	
Response	Descriptors
M_{12}	Softness, PJI2, Mor08e, Mor30e, R5v, Mor26p, R1m+, ALOGPS_logP
k_0	Softness, PJI2, Mor08e, Mp, Mor28p, Mor20m, Mor18m, Mor24u, EEig01r, R1m+, ALOGPS_logP
k_1	Softness, PJI2, Mp, GATS1e, Mor18m, Mor20v, Mor08e, HATS2u, R1v, Ui, ALOGPS_logP
k_H	Softness, PJI2, Mor08e, Mp, EEig01r, Mor24u, Mor18m, Mor20m, Mor08e, Mor28p, R1m+
$\ln k_p$	Softness, RDF055m, Mor04m, Mor24m, Mor20v, Mor18e, G2u, E2s, Du, R1m+
n	BELv1, GGI4, SP18, Mor24m, Mor18v, E1e, E2s, Ku, Du

weak ($R < 0.5$) correlation with the response, were rejected. Multicollinearity and redundancy among descriptors can pose serious problems in QSPR generation (Whitley, Ford, & Livingstone, 2000). This problem was resolved by application of supervised learning concept of non-redundant descriptors (NRD) (Ghavami & Sajadi, 2010) followed by unsupervised forward selection (UFS) algorithm (Whitley et al., 2000), because in case of larger descriptors dataset, direct use of UFS usually results in descriptors which are difficult to interpret. In the use of NRD, if two descriptors are highly correlated with each other ($R > 0.8$), the one with stronger correlation with the response is retained. The remaining descriptors were subjected to UFS algorithm, repeatedly used on the reduced data set having R^2_{max} values 0.1–0.9 (increment of 0.1) and 0.95. UFS was performed with UFS-1.8 (Centre for Molecular Design, University of Portsmouth available at http://www.port.ac.uk). The descriptors thus selected are listed in Table 1.

For QSPR generation, MLR models were built with the UFS's selected descriptors. Initially, the data were standardized to zero mean and unit variance in order to avoid any bias, which may lead to serious errors in the generation and application of the models. The procedure consisted of testing, in sequence, 1–6 descriptor models in all possible combinations. The models with lower Mallows *Cp*, and best R^2, R^2_{cv} and physical interpretability were selected for prediction. Where R^2_{cv} provides for a good measure of the predictive power of a model as R^2 alone can increase due to artifacts, whereas, R^2_{cv} decreases if a model is over parameterized (Hawkins, Basak, & Mills, 2003).

In order to obtain optimized models with maximum predictive ability, ANN were used on the descriptors selected by MLR. The ANN is a powerful multivariate data analysis technique, capable of both linear and non-linear modeling and has been widely used in modeling structure–property relationships (Fatemia, Ghorbanzad'ea, & Baher, 2010; Mittermayr, Olajos, Chovan, Bonn, & Guttman, 2008). It mimics the human brain intelligence system and consists of various interconnecting neurons organized, in a sequential manner, into an input layer, one or more hidden layers and an output layer. In the present study the MLR-selected descriptors were entered as continuous input signals into ANNs and outputs were various release parameters. ANNs were trained by use of Statistica 8.0 neural network implementation subroutine consisting of: multilayer-perceptrons (MLP) type network with feed-forward topology, Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm and normal randomization. The sum-of-squares (SOS) error function was used to test their performances. *Identity*, *logistic*, *exponential* and *tanh* activation functions both for hidden and output layers and number of hidden units from 3 to 10 were used. The models with least *SOS* errors, and higher R^2_{train} and R^2_{test} values were considered as optimal. In ANNs building process, an early stopping technique was employed to avoid over-training of the ANN models. The data set was subdivided randomly into a subset of 13 molecules (approx. 80%) for training and a subset of 3 molecules (approx. 20%) as a test set to avoid over-fitting.

2.3.4. Validation of QSPRs

Several techniques are in use for validation of models (Gramatica, 2007; Hawkins et al., 2003; Tropsha, Gramatica, & Gombar, 2003; Wold, 1991). In the present study, the QSPR models were validated by leave-one-out cross validation (LOOCV) technique and any chance correlation was tested by the use of y-scrambling technique, a method frequently used for the purpose. LOOCV involves leaving each molecule out of the training set in turn and train the model on the remaining molecules. The parameters of the left-out molecule are calculated from the new model. The model quality was assessed by calculating the cross-validated squared correlation coefficient (R^2_{cv}) by use of Eq. (6).

$$R^2_{cv} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_{train})^2} \quad (6)$$

where $\hat{y}_i$ and y_i are respectively the predicted and observed values for the ith left-out molecule, and $\bar{y}_{train}$ is the mean of the experimental values of the rest of the molecules. y-Scrambling is used to ascertain the robustness of the model. For each model the values of response variables were randomly assigned to the molecules and the models were rebuilt with the randomized data. Lower level correlations in the randomized data models than the original model would indicate the robustness of the original model (Gramatica, 2007; Tropsha et al., 2003).

The QSPR models were also validated by determination of coefficients of determinations, R^2_0, $R^{2'}_0$, R^2_m, and slopes, k_0, k_0', from linear regression plots between observed and predicted data where the intercept is set at zero (Goodarzi, Freitas, & Heyden, 2011). The R^2_0 and k_0 are the coefficient of determination and slope respectively

Please cite this article in press as: Akbar, J., et al. A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks. *Carbohydrate Polymers* (2012), doi:10.1016/j.carbpol.2012.02.016

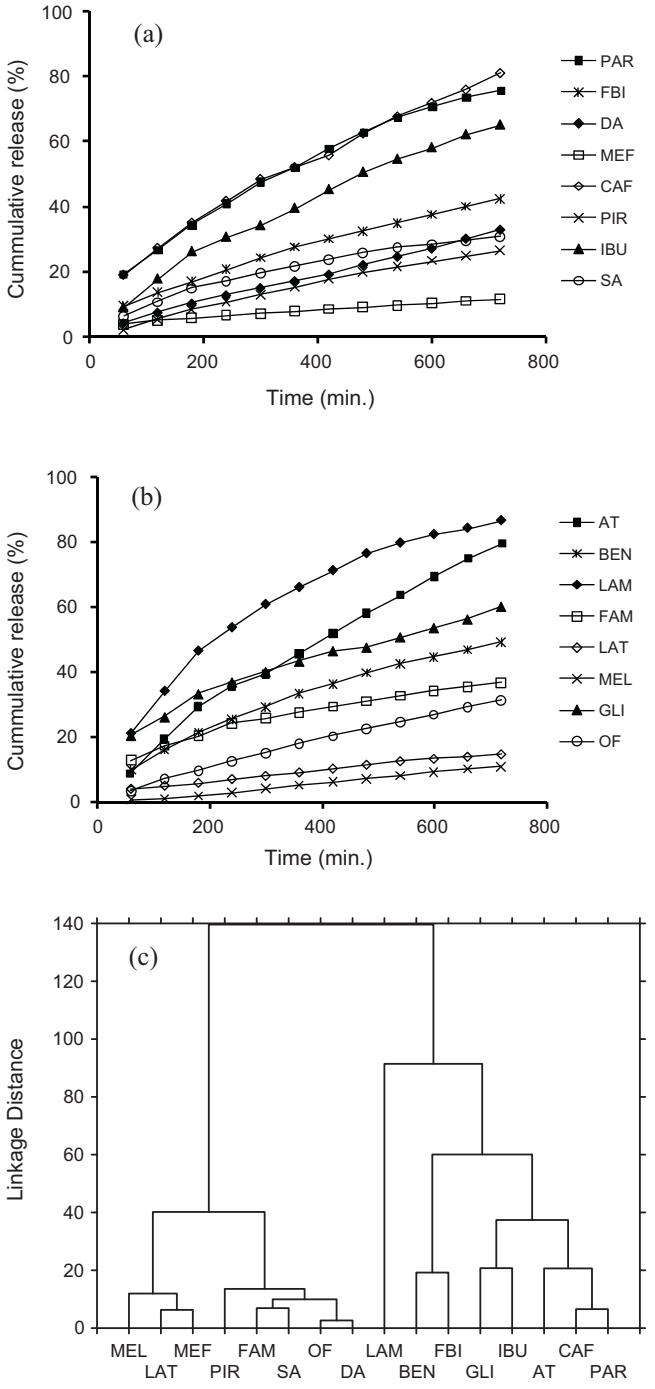

Fig. 1. (a), (b) Release profiles of drug molecules, and (c) dendrogram showing the similarity level of release profiles of drugs.

for the linear regression between "observed *versus* predicted" values and $R_0^{2\prime}$ and $k_0\prime$ respectively are for "predicted *versus* observed" values. R_m^2 is defined as $R_m^2 = R^2 \times (R^2 - R_0^2)^{1/2}$. The acceptable validation criteria are; R_0^2, $R_0^{2\prime}$, R_m^2, k_0 and $k_0\prime \approx 1$; similar R_0^2, $R_0^{2\prime}$ and R_m^2 values; $R^2 > 0.6$, $R_{cv}^2 > 0.5$, $(R^2 - R_0^2)/R^2 < 0.1$ and $0.85 \leq k_0 \leq 1.15$, $(R^2 - R_0^{2\prime})/R^2 < 0.1$ and $0.85 \leq k_0\prime \leq 1.15$ and $|R_0^2 - R_0^{2\prime}| < 0.3$.

All the calculations were performed by use of Minitab 15, Statistica 8 and MS Excel® 2007 as appropriate.

3. Results and discussion

3.1. Dissolution profiles

Drug release from the hydrophilic matrix started by penetration of water into the device forming a gel-like structure. The release data showed good linear relationships with the applied kinetic models and various parameters including R^2, rate constants, release exponents, and *MSC* values were determined (Table 2).

The drugs CAF, AT, PAR and LAM showed a release in the range 75–85%, whereas MEF, LAT and MEL were released to a smaller extent (10–15%) in 12 h time. The release of BEN, IBU and GLI was moderate (50–65%). The drugs FBI, OF, PIR, SA, DA and FAM exhibited a release less than the moderate (26–42%), (Table 2). Due to differences in mechanism (as depicted by *n*, Table 2), a single model cannot be considered as sufficient for predicting a release pattern of all the drugs; the best-fitted model was identified from the MSC values (Table 2). The release profiles of drugs are presented in Fig. 1a and b. The similarity of release profiles was determined by HCA and dendrograms showing the formation of groups were obtained by using weighted-pair-group average and Euclidean distance (Fig. 1c).

3.2. QSPRs

3.2.1. Selection of descriptors

In QSPRs, generally a large descriptor pool is generated, which is then reduced to a smaller data set by a suitable algorithm. For the reduction of descriptors, both supervised and unsupervised learning algorithms are used, which have their own advantages and disadvantages. In case of unsupervised learning, however, the major disadvantage is the lack of direction for the learning algorithm which may result in the information of no importance for a given phenomenon. An HM approach is mainly a supervised learning method, which can explore the rationale behind the phenomenon. Therefore, in this study, the HM approach was combined with UFS in order to discover rational descriptors. Initially, HM approach resulted in a small dataset of 15–25 descriptors out of 1685. The redundant and multicollinear descriptors were then removed by NRD and UFS algorithm which provided a small subset of 8–11 descriptors for different response variables. These are summarized in Table 1 and were used for MLR analysis.

3.2.2. QSPR generation

The selected descriptors were regressed upon response variables and MLR models consisting of 1–6 descriptors were built in all possible combinations. The best-fitting models were selected as described earlier. None of the one-descriptor model was found to be significant. However, two-descriptor models had somewhat improved statistical significance. These models are:

$$M_{12} = 173 - 10.3 \ \text{Softness} \ - 6.73 \ \text{ALOGPS\_log} \ P \quad (7)$$

$$(N = 16, \quad R^2 = 0.664, \quad R_{adj}^2 = 0.613, \quad R_{cv}^2 = 0.588, \quad p = 0.001)$$

$$k_0 = 0.152 - 0.00952 \ \text{Softness} + 0.0216 \ \text{Mor08e} \quad (8)$$

$$(N = 16, \quad R^2 = 0.659, \quad R_{adj}^2 = 0.607, \quad R_{cv}^2 = 0.573, \quad p = 0.001)$$

$$k_1 = -0.00411 - 0.000424 \ \text{Mor08e} + 0.000988 \ \text{Ui} \quad (9)$$

$$(N = 16, \quad R^2 = 0.753, \quad R_{adj}^2 = 0.715, \quad R_{cv}^2 = 0.699, p < 0.001)$$

$$k_H = 1.47 - 0.293 \ \text{Softness} - 14.1 \ \text{Mp} \quad (10)$$

Please cite this article in press as: Akbar, J., et al. A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks. *Carbohydrate Polymers* (2012), doi:10.1016/j.carbpol.2012.02.016

Table 2
Various mathematical models fitness data and difference (f_1) and similarity (f_2) factors for comparison between experimental and ANN predicted release profiles.

	Zero order			First order			Higuchi			Power law			n	M_{12}	f_1	f_2
	R^2	$k_0 \times 10^2$	MSC	R^2	$k_1 \times 10^2$	MSC	R^2	$k_H \times 10^2$	MSC	R^2	$k_P \times 10^2$	MSC				
PAR	0.976	8.635	3.395	0.999	−0.187	**6.175**	0.997	312.139	5.464	0.997	1.751	5.642	0.58	76	1.5	93.2
FBI	0.991	4.914	4.337	0.998	−0.068	**6.139**	0.995	176.167	5.070	0.998	0.736	5.753	0.61	43	1.6	97.0
DA	0.998	4.193	**5.877**	0.994	−0.052	4.821	0.975	148.233	3.358	0.998	0.156	5.686	0.81	33	1.3	93.1
MEF	0.998	1.035	**5.780**	0.997	−0.011	5.646	0.967	36.437	3.070	0.932	0.956	2.357	0.37	12	14.3	89.2
CAF	0.988	9.034	4.117	0.984	−0.208	3.815	0.996	324.274	5.110	0.998	1.711	**6.026**	0.58	81	1.3	93.1
PIR	0.989	3.662	4.211	0.996	−0.043	5.128	0.996	131.406	**5.214**	0.989	0.046	4.173	0.98	27	4.1	94.5
IBU	0.984	8.220	3.817	0.997	−0.143	**5.546**	0.996	295.786	5.251	0.992	0.431	4.440	0.77	65	2.5	89.7
SA	0.957	3.541	2.819	0.973	−0.044	3.266	0.996	129.198	**5.242**	0.990	0.525	4.266	0.63	31	4.4	92.3
AT	0.990	10.241	4.258	0.979	−0.214	3.536	0.993	366.919	**4.683**	0.988	0.318	4.098	0.84	79	2.6	83.9
BEN	0.976	5.807	3.381	0.994	−0.086	4.773	0.999	210.193	**7.016**	0.998	0.689	5.711	0.65	49	1.1	97.8
LAM	0.925	9.343	2.252	0.998	−0.267	**5.690**	0.984	344.694	3.800	0.982	2.333	3.707	0.56	86	1.7	89.2
FAM	0.960	3.442	2.886	0.976	−0.047	3.405	0.997	125.425	5.384	0.998	2.173	**6.001**	0.43	37	28.2	55.5
LAT	0.996	1.704	5.080	0.996	−0.019	**5.295**	0.980	60.475	3.574	0.976	0.295	3.415	0.59	15	4.7	97.1
MEL	0.997	1.662	**5.497**	0.997	−0.018	5.430	0.968	58.566	3.097	0.992	0.001	4.455	1.48	11	13.1	94.6
GLI	0.971	5.512	3.196	0.989	−0.096	4.205	0.995	199.581	4.950	0.996	3.461	**5.144**	0.43	60	1.8	91.4
OF	0.995	4.162	5.024	0.999	−0.051	**7.025**	0.993	148.627	4.561	0.997	0.093	5.454	0.89	31	6.5	88.2

Note: MSC value for the best fit model are in boldface.

$$(N = 16, \quad R^2 = 0.656, \quad R^2_{adj} = 0.603, \quad R^2_{cv} = 0.570, p = 0.001)$$

$$\ln \ k_p = -11.8 + 44.2 \ \ G2u - 9.33 \ \ E2s \qquad (11)$$

$$(N = 16, \quad R^2 = 0.659, \quad R^2_{adj} = 0.607, \quad R^2_{cv} = 0.374, \quad p = 0.001)$$

$$n = 1.28 + 1.52 \ \ Ku - 3.31 \ \ Du \qquad (12)$$

$$(N = 16, \quad R^2 = 0.65, \quad R^2_{adj} = 0.596, \quad R^2_{cv} = 0.480, \quad p = 0.001)$$

where N is the number of observations. In these models, ln k_p and n models showed very weak R^2_{cv} than others; however, all these models were improved by the addition of more descriptors. Three- and four-descriptor models showed an optimal statistical significance. Addition of more descriptors in these models resulted in lowering of R^2_{cv}. Thus three- and four-descriptor models, as appropriate, were considered for prediction purpose and are presented in Table 3. The models for final prediction were selected on the basis of higher R^2 and R^2_{cv} values, lower Mallows Cp, and better relationship of model descriptors with the drug release phenomenon. The optimal models for M_{12}, k_0, k_1, k_H, ln k_p and n are represented by Eqs. (19), (20), (21), (22), (17) and (18), respectively. In order to determine the robustness and neglect of chance correlation, y-scrambling analysis was performed, where the robust models showed a decrease in significance after y-scrambling (y-intercept < 0.2).

The predictions made by selected MLR models are given in Fig. 2. Three-descriptor model for M_{12} (Eq. (13)) incorporating Softness, ALOGPS_logP and Mor08e showed a good predictive ability (R^2 = 0.803, R^2_{cv} = 0.739). Addition of fourth descriptor, PJI2, further improved the model (Eq. (19); R^2 = 0.842, R^2_{cv} = 0.728). Therefore, Eq. (19) was used for further prediction. Softness is a quantum mechanical descriptor and its negative relation to M_{12} suggested that with increased softness of the molecule M_{12} would decrease. This means that soft molecules were unable to diffuse into the hard dissolution medium because of strong interaction with the polymer system. For example MEL being the softest (Table S1, supplementary information) of all the drugs had least M_{12} while AT, CAF, LAM and PAR being on the lower side exhibited higher values of M_{12}. The ALOGPS_logP and PJI2 were also inversely related to M_{12}. The ALOGPS_logP is a well-known lipophilicity/hydrophobicity index of a substance; the molecules with higher values will not allow them to pass into aqueous dissolution medium, thus slowing down the release. LAM, AT and CAF having very low ALOGPS_logP values (more hydrophilicity) showed an enhanced release (79–86%) than most of others (<50%). The PJI2, a two-dimensional Petitjean Shape index describing the degree of deviation of a shape from a perfect cyclic topology, varied between 0 and 1 (Borges de Melo, Ataide Martins, Marinho Jorge, Friozi, & Castro Ferreira, 2010). Therefore, a perfection towards cyclic topology will result in a decrease in M_{12} value because of its negative relationship with M_{12} in the model. This may be attributed to the tight packing of regularly shaped molecules between layers of the polymer. Thus this will result in better encapsulation of such molecules and delayed release. LAT and MEF present a good example of this phenomenon in this study. The Eq. (19) contains another descriptor Mor08e, which is a three-dimensional molecule representation of structures based on electron diffraction, MoRSE, descriptor. These descriptors are derived from infrared spectra simulations using a generalized scattering function. These descriptors acquire information from the 3D atomic coordinates, calculated by summing atom weights viewed by a different angular scattering function (Schuur, Selzer, & Gasteiger, 1996). The Mor08e is 3D-MoRSE-signal 8/weighted by atomic Sanderson electronegativities; it describes

Please cite this article in press as: Akbar, J., et al. A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks. *Carbohydrate Polymers* (2012), doi:10.1016/j.carbpol.2012.02.016

Table 3
Multiple linear regression models.

Model	Model parameters				
	N	R^2	R^2_{adj}	R^2_{cv}	p
Three descriptors models					
$M_{12} = 143 - 8.25$ Softness $-$ 5.97 ALOGPS_log $P + 15.2$ Mor08e (13)	16	0.803	0.753	0.739	0.001
$k_0 = 0.192 - 0.00837$ Softness $+$ 0.0201 Mor08e $- 0.0626$ PJI2 (14)	16	0.739	0.674	0.643	0.001
$k_1 = -0.00296 - 0.000464$ Mor08e $+ 0.000817$ Ui $- 0.00147$ HATS2u (15)	16	0.792	0.741	0.700	<0.001
$k_H = 6.95 - 0.303$ Softness $+ 0.725$ Mor08e $- 2.28$ PJI2 (16)	16	0.735	0.669	0.637	0.001
ln $k_p = -11.6 + 39.8$ G2u $- 8.13$ E2s $- 0.831$ Mor04m (17) [a]	16	0.769	0.712	0.540	<0.001
$n = 0.932 + 1.27$ Ku $- 2.70$ Du $+ 0.723$ E2s (18) [a]	16	0.759	0.698	0.502	0.001
Four descriptors models					
$M_{12} = 167 - 7.54$ Softness $- 5.19$ ALOGPS_log$P + 14.6$ Mor08e $- 39.6$ PJI2 (19) [a]	16	0.842	0.785	0.728	<0.001
$k_0 = 0.322 - 0.00681$ Softness $+ 0.0140$ Mor08e $- 0.0558$ PJI2 $- 0.225$ Mp (20) [a]	16	0.793	0.717	0.686	0.001
$k_1 = -0.00205 - 0.000464$ Mor08e $+ 0.000696$ Ui $- 0.00137$ HATS2u $-$ 0.000764 GATS1e (21) [a]	16	0.813	0.745	0.696	0.001
$k_H = 11.7 + 0.501$ Mor08e $- 2.03$ PJI2 $- 0.246$ Softness $- 8.20$ Mp (22) [a]	16	0.789	0.712	0.681	0.001
ln $k_p = -11.3 + 37.5$ G2u $- 5.94$ E2s $- 0.943$ Mor04m $- 0.116$ RDF055m (23)	16	0.788	0.710	0.499	0.001
$n = 0.901 + 1.20$ Ku $- 2.68$ Du $+ 0.574$ E2s $- 0.185$ Mor18v (24)	16	0.776	0.694	0.495	0.001
Five descriptors models					
$M_{12} = 166 - 6.55$ Softness $- 4.99$ ALOGPS_log$P + 12.7$ Mor08e $- 42.9$ PJI2 $-$ 37.8 R1m+ (25)	16	0.851	0.776	0.684	0.001
$k_0 = 0.391 - 0.0384$ EEig01r $+ 0.0140$ Mor08e $- 0.0503$ PJI2 $- 0.219$ Mp $-$ 0.0772 Mor28p (26)	16	0.835	0.752	0.631	0.001
$k_1 = -0.00217 - 0.000442$ Mor08e $- 0.00146$ HATS2u $- 0.00114$ GATS1e $+$ 0.000108 Softness $+ 0.000458$ Ui (27)	16	0.838	0.756	0.651	0.001
$k_H =$ 13.4 $+ 0.605$ Mor08e $- 2.02$ PJI2 $- 0.129$ Softness $- 7.65$ Mp $- 0.829$ EEig01r (28)	16	0.810	0.714	0.616	0.002
ln $k_p =$ $-9.78 + 30.8$ G2u $- 4.25$ E2s $- 1.06$ Mor04m $- 0.131$ RDF055m $- 3.86$ R1m+ (29)	16	0.801	0.702	0.506	0.003
$n = 2.52 + 1.45$ Ku $- 2.91$ Du $+ 0.651$ E2s $- 0.296$ Mor18v $- 0.890$ BELv1 (30)	16	0.790	0.686	0.471	0.004

N, number of observations, R^2_{adj},: adjusted coefficient of determination.
[a] Models carried forward to Table 4 for ANN analysis.

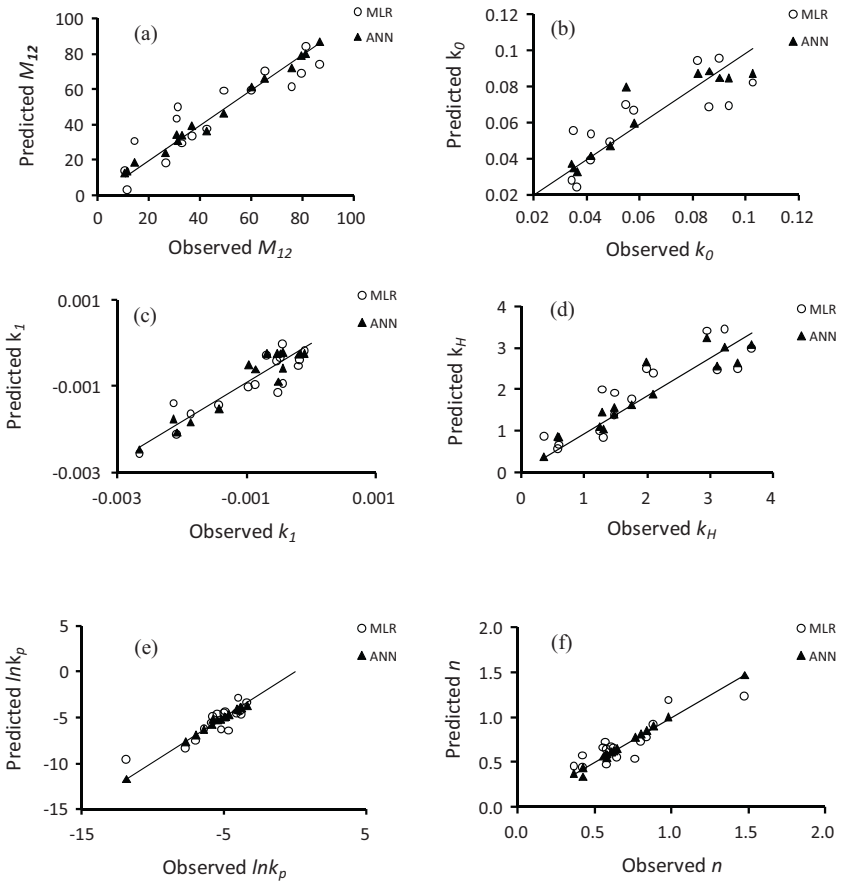

Fig. 2. Plots for observed and predicted data based on MLR and ANN models.

Please cite this article in press as: Akbar, J., et al. A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks. *Carbohydrate Polymers* (2012), doi:10.1016/j.carbpol.2012.02.016

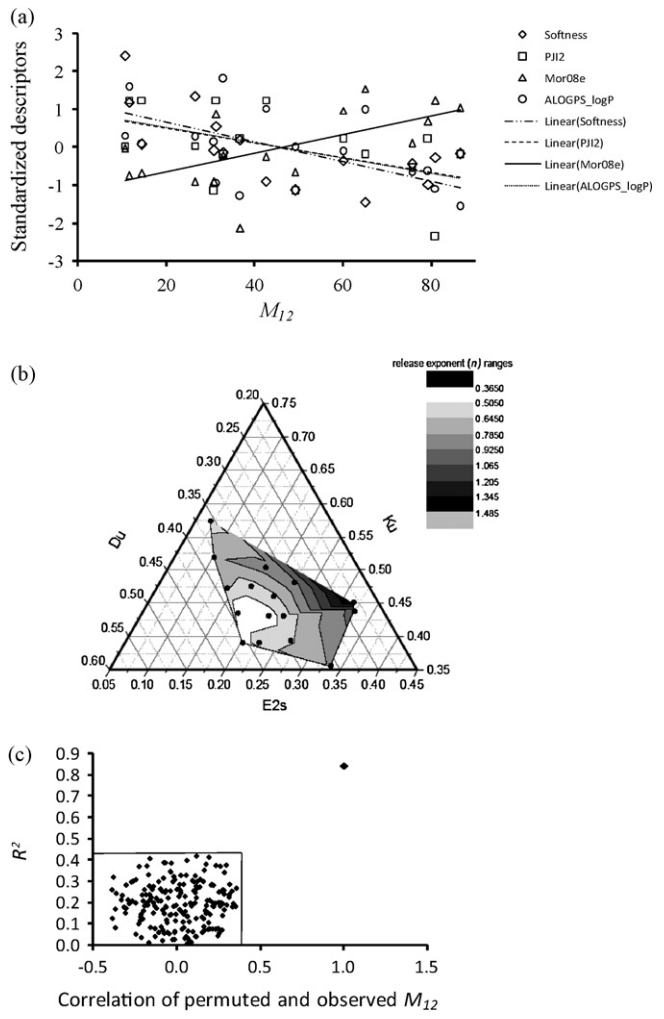

Fig. 3. (i) Representative plots showing simultaneous effects of descriptors: (a) scatter plot for *M12* and (b) ternary contour plot for *n* (ii) y-scrambling plot for *M12* MLR model.

the effect of electronegativity on the drug release. Its positive relationship with M_{12} suggests that an increase in Mor08e value will result in increase in the amount of drug released. Thus it can be concluded that softer molecules having lower lipophilicity, higher Mor08e value and irregular shape would result in an enhanced release of the drug from AX. The combined effect of all four descriptors on M_{12} is shown in the form of scatter plot (Fig. 3a) and a typical y-scrambling plot for M_{12} is shown in Fig. 3c.

In order to have insight into the drug release phenomenon form the AX matrix, other response parameters were also investigated to find other factors which may influence the release. These parameters were included in models represented by Eqs. (20)–(22), (17) and (18). The Eqs. (20) and (22) represent k_0 and k_H as optimum models respectively. In these models the descriptor ALOGPS_logP in Eq. (19) has been replaced with Mp. The descriptors including softness and PJI2 are related negatively to k_0 and k_H, therefore, an increase in their values will slow down the rate of drug release from AX. The descriptor, Mor08e, has positive relationship and therefore, an increase in its value will also increase the drug release rate. The other descriptor Mp, which describes the mean atomic polarizability (scaled on carbon atom), has negative relationship with k_0 and k_H. This suggests that molecules with higher values of Mp would have slower release rate due to local atomic charge dispersions. It may be noted that AX, being a carbohydrate polymer, has polar OH groups as a part of its structure. Therefore, drug molecules having greater atomic charge dispersion will have stronger drug-polymer interaction resulting in a lower release rate. This was witnessed in case of drugs MEL, LAT, PIR, MEF having greater π-character and greater number and/or larger size of atoms which result in greater charge dispersions.

In case of k_1 (Eq. (21)), three descriptors Ui, HATS2u and GATS1e were identified in addition to Mor08e which is positively related to k_1. Since, slope of a first order plot is always negative, therefore, an increase in Mor08e value would decrease k_1 resulting in an increase in release rate. The Ui an empirical descriptor, known as unsaturation index, is also positively related to k_1. Therefore, an increase in Ui value will increase the k_1, *i.e.*, a slower release. Highly unsaturated drug molecules like MEL, LAT, MEF and PIR exhibited a slower release. Whereas LAM, AT, PAR and CAF having lower Ui values (Table S1, supplementary information) exhibited a higher rate of release. This is because of greater π-charge dispersion and, as in case of Mp, it is expected to have stronger drug–polymer interactions resulting in a decreased release rate. Most of the other drug molecules under investigation had Mp and Ui values intermediate between those of highly and less unsaturated molecules thus showing intermediate release rates. HATS2u pertains to geometry, topology and atomic weights assembly (GETAWAY) family of descriptors (Consonni, Todeschini, & Pavan, 2002; Consonni, Todeschini, Pavan, & Gramatica, 2002) derived from leverage matrix called molecular influence matrix (MIM). These descriptors encode information about 3D geometry and atom-related molecular topology. HATS2u is the leverage weighted autocorrelation-lag2/unweighted. GATS1e, Geary autocorrelation-lag1/weighted by atomic Sanderson electronegativities (Geary, 1954), is a 2D topological descriptor additionally containing information about atomic electronegativities. Both of these descriptors have negative coefficient suggesting an inverse relationship with k_1, meaning by that with increase in 2D lag1 and 3D lag2 autocorrelations within a molecule, the release rate will increase. It may also be noted that electronegativity information contained in both Mor08e and GATS1e have similar effect, as an increase in both these descriptors increases the release rate.

The $\ln k_p$ and n response parameters were derived from power law; the optimal models representing these parameters are Eqs. (17) and (18) respectively. These models, containing mostly same class of descriptors, *i.e.*, weighted holistic invariant molecular (WHIM) type (Todeschini & Gramatica, 1997; Todeschini, Lasagni, & Marengo, 1994), provide complementary information regarding drug release phenomenon. As regards $\ln k_p$, its optimal model includes three descriptors G2u, E2s and Mor04m. The G2u belongs to the WHIM class. The WHIM descriptors are statistical indices calculated on the projections of the atoms along principal axes. Thus these descriptors are built in such a way that they acquire relevant 3D information regarding molecular size, shape, symmetry, and atom distribution with respect to invariant reference frames. One of the WHIM descriptors describes global directional symmetry index that contains mean information content of the symmetry indices along each principle component. The G2u is a second component symmetry directional WHIM index/unweighted. This index encodes symmetry information about the molecules and increases with an increase in molecular symmetry. Its high positive relationship with $\ln k_p$ suggests that an increase in molecular symmetry will lead to an increase in $\ln k_p$. The E2s is a second component accessibility directional WHIM index/weighted by atomic electrotopological states and the Mor04m is a 3D-MoRSE—signal 4/weighted by atomic masses. Both of them showed negative relationship with $\ln k_p$ describing release pattern as discussed above. The release exponent n is an important parameter as it determines the mechanism for the release of a drug. Its model solely consists of WHIM descriptors Ku, Du and E2s. The E2s is the same descriptor as described in Eq. (17). The Ku is the global shape index/unweighted

Please cite this article in press as: Akbar, J., et al. A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks. *Carbohydrate Polymers* (2012), doi:10.1016/j.carbpol.2012.02.016

and the Du is the total accessibility index/unweighted, where the Ku increases with increase in linearity of the molecule. It is positively correlated with n, showing that the n will increase with the linearity of the molecule; thus shifting the mechanism from fickian diffusion ($n = 0.45$) through anomalous transport ($0.45 < n < 0.89$) and swelling-controlled mechanism ($n = 0.89$) to super case-II transport ($n > 0.89$) domain. For example, IBU and AT possess sufficient linearity with high Ku index so their release mechanisms were found to be anomalous transport with n values 0.76 and 0.84 for IBU and AT respectively. Both MEL and PIR also possess a high Ku index thus showing a super case-II transport. A similar impact of E2s was observed on n in this model. It has direct relationship with the n, meaning by that more the accessibility of the atoms of a molecule, higher the n, therefore, favors non-Fickian type mechanisms. The Du describes the total density of atoms within a molecule; that means the greater its value, the greater will be the projected unfilled space (Saíz-Urra, Cabrera Pérez, Helguera, & Froeyen, 2011). This descriptor has negative impact on n which means that for atom dense molecules, n will be lower resulting in Fickian diffusion. For example, GLI having high Du index was found to be released through nearly ($n = 0.43$) Fickian diffusion. This suggests that more linear and less dense molecules with high atomic accessibility tend to follow non-Fickian type mechanisms. The combined effect of all three descriptors on n has been summarized in a triangular contour plot (Fig. 3b). The plot was created with Origin Pro version 8.5.

An ideal study should be able to predict the complete release profile of dug molecules under specified conditions. In order to achieve this, the % release (M) of drug molecules at all time points was regressed against time (t) and descriptors contained in Eq. (19). This incorporated all the release data into a model with reasonably good statistical significance (Eq. (31)).

$$M = 89.8 + 0.0531\ \ t - 4.92\ \ \text{Softness} - 27.9\ \ \text{PJI2} + 10.8\ \ \text{Mor08e} - 3.64\ \ \text{ALOGPS\_log}\ \ P \quad (31)$$

$$(N = 192,\quad R^2 = 0.798,\quad R^2_{\text{adj}} = 0.792,\quad R^2_{\text{cv}} = 0.691,\quad p < 0.001)$$

For cross validation of this model, all the data points relating to release of the one-by-one selected drug were left out and prediction was obtained by the model trained on the remaining molecules.

The behavior of descriptors; Softness, PJI2, ALOGPS_logP and Mor08e remained same in this model as previously discussed. The predicted release profile for all drug molecules is presented in Fig. 4a.

3.2.3. Artificial neural networks

The network architecture and validation statistics for ANNs is given in Table 4. The descriptors in the selected models (Table 3) were used to build ANN models. The real strength of ANN mapping technique was observed for the $\ln k_p$ and n response parameters, which showed considerably better prediction ability than MLR models (Fig. 2). The MLR models of $\ln k_p$ and n showed a low predictive ability ($R^2 < 0.75$, $R^2_{\text{cv}} < 0.55$), whereas, the ANNs, being inherently capable of both linear and non-linear modeling, afforded the models having greatly improved statistical qualities (Table 4). For each response many ANNs were built and the optimal model was selected on the basis of lower SOS error function value and higher R^2_{train} and R^2_{test} values. The ANN models were considered to be superior over the MLR models. In the ANN models all the response parameters have different types of architecture. In these models the so called 'global sensitivity analysis' ranked the descriptors in ANNs as shown in Table 4. Keeping in view the suitability of ANN models (better validation statistics: Table S2, supplementary information) and better predictive ability (Fig. 2 and 4) these models were considered to be superior to the MLR models. The ANN was also designed to simultaneously predict two important parameters for a drug molecule, M_{12} and n. The descriptors in Eqs. (18) and (19) were used as continuous input and the output was M_{12} and n (Eq. (39)).

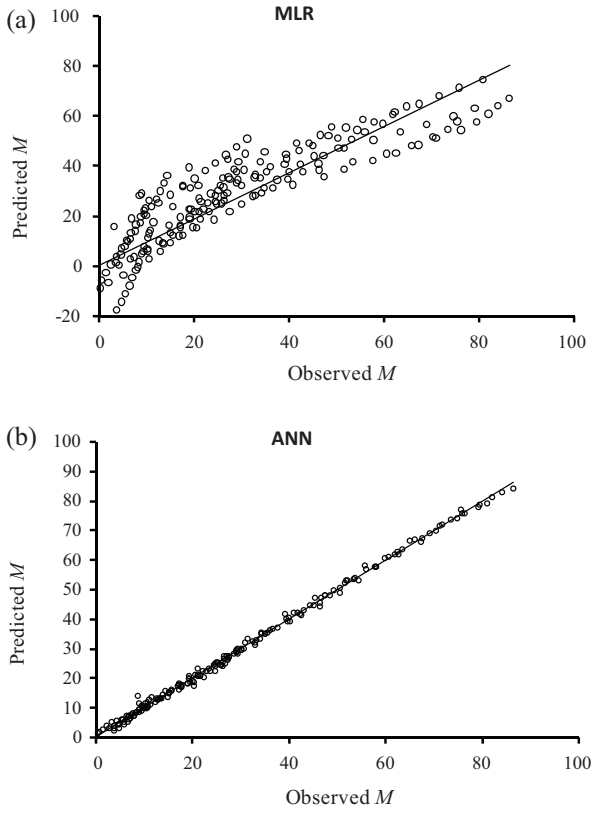

Fig. 4. Plots for observed and predicted release profiles of drug molecules (a) MLR and (b) ANN model.

Table 4
Architecture of artificial neural network models.

Response	Number of neurons			Activation function		R^2_{train}	R^2_{test}	Global sensitivity analysis of descriptors	
	Input layer	Hidden layer	Output layer	Hidden	Output				
M_{12}	4	8	1	Logistic	Identity	0.985	0.999	ALOGPS_log$P$ > Softness > PJI2 > Mor08e	(32)
k_0	4	4	1	Logistic	Logistic	0.979	0.997	PJI2 > Mor08e > Softness > Mp	(33)
k_1	4	3	1	tanh	tanh	0.910	0.947	Ui > Mor08e > HATS2u > GATS1e	(34)
k_H	4	5	1	tanh	tanh	0.905	0.988	Softness > Mor08e > PJI2 > Mp	(35)
$\ln k_p$	3	8	1	Exponential	Logistic	0.998	0.999	G2u > E2s > Mor04m	(36)
n	3	10	1	tanh	Identity	0.999	0.997	Ku > Du > E2s	(37)
M	5	6	1	tanh	Logistic	0.998	0.996	t > Softness > Mor08e > ALOGPS_log$P$ > PJI2	(38)
M_{12}, n	7	9	2	Exponential	Identity	0.984	0.950	Softness > ALOGPS_log$P$ > Mor08e > Ku > PJI2 > E2s > Du	(39)

Please cite this article in press as: Akbar, J., et al. A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks. *Carbohydrate Polymers* (2012), doi:10.1016/j.carbpol.2012.02.016

The complete release profile of the drugs was also predicted by ANN analysis. The descriptors in Eq. (31) were used to develop an MLP 5-6-1 ANN model (Table 4, Eq. (38)) that can predict the release profiles of the drugs under investigation with considerable better predictive performance over MLR model (Fig. 4). The predicted release profiles of the drugs by ANN were compared with the experimental profiles using difference (f_1) and similarity (f_2) factors (Moore & Flanner, 1996) as given in Eqs. (40) and (41).

$$f_1 = \left\{ \frac{\left[\sum_{t=1}^{n} \left|E_t - P_t\right|\right]}{\left[\sum_{t=1}^{n} P_t\right]} \right\} \times 100 \tag{40}$$

$$f_2 = 50 \times \log \left\{ \left[1 + \left(\frac{1}{n}\right) \sum_{t=1}^{n} \left|E_t - P_t\right|^2 \right]^{-0.5} \times 100 \right\} \tag{41}$$

where E_t and P_t are the experimental and predicted dissolution values respectively at time t. The summation is done over all time points n. For two curves to be similar f_1 should be close to zero and f_2 close to 100. Generally, f_1 value in the range 0–15 and f_2 greater than 50 indicate the equivalence of two dissolution profiles. The f_1 and f_2 values for the drugs under investigation are given in Table 2 and predictive release profile data is shown in Fig. 4b. The MLR and ANN models conformed to the validation standards (Goodarzi et al., 2011) for QSAR/QSPR (Table S2, supplementary information).

4. Conclusions

New QSPR models were developed by use of MLR and ANN through *ab initio* structure optimization. The models are capable of predicting release profiles of drug molecules differing in structure and pharmacological activity. Softness, lipophilicity (ALOGPS_logP), geometry (Mor08e, GATS1e, HATS2u), cyclic topology (PJI2), polarizability (Mp) and π-character (Ui) in the molecules were identified as the most significant descriptors governing the release of a drug from the arabinoxylan. The models meet the criteria for stability, robustness, and predictive power and as such can be used with confidence as effective tools for predicting, analyzing the behavior of a variety of drugs and designing controlled release tablets of different drugs.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.carbpol.2012.02.016.

References

Baveja, S. K., Ranga Rao, K. V., Singh, A., & Gombar, V. K. (1988). Release characteristics of some bronchodilators from compressed hydrophilic polymeric matrices and their correlation with molecular geometry. *International Journal of Pharmaceutics*, *41*(1–2), 55–62.

Borges de Melo, E., Ataide Martins, J. P., Marinho Jorge, T. C., Friozi, M. C., & Castro Ferreira, M. M. (2010). Multivariate QSAR study on the antimutagenic activity of flavonoids against 3-NFA on *Salmonella typhimurium* TA98. *European Journal of Medicinal Chemistry*, *45*(10), 4562–4569.

Consonni, V., Todeschini, R., & Pavan, M. (2002). Structure/response correlation and similarity/diversity analysis by GETAWAY descriptors 1. Theory of the novel 3D molecular descriptors. *Journal of Chemical Information and Computer Sciences*, *42*(3), 682–692.

Consonni, V., Todeschini, R., Pavan, M., & Gramatica, P. (2002). Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *Journal of Chemical Information and Computer Sciences*, *42*(3), 693–705.

Fatemia, M. H., Ghorbanzad'ea, M., & Baher, E. (2010). Quantitative structure retention relationship modeling of retention time for some organic pollutants. *Analytical Letters*, *43*(5), 823–835.

Gafourian, T., Safari, A., Adibkia, K., Parviz, F., & Nokhodchi, A. (2007). A drug release study from hydroxypropylmethylcellulose (HPMC) matrices using QSPR modeling. *Journal of Pharmaceutical Sciences*, *96*(12), 3334–3351.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, *5*(3), 115–145.

Ghavami, R., & Sajadi, S. M. (2010). Semi-empirical topological method for prediction of the relative retention time of polychlorinated biphenyl congeners on 18 different HR GC columns. *Chromatographia*, *72*(5), 523–533.

Gibaldi, M., & Feldman, S. (1967). Establishment of sink conditions in dissolution rate determinations-theoretical considerations and application to nondisintegrating dosage forms. *Journal of Pharmaceutical Sciences*, *56*(10), 1238–1242.

Goodarzi, M., Freitas, M. P., & Heyden, Y. V. (2011). Linear and nonlinear quantitative structure–activity relationship modeling of the HIV-1 reverse transcriptase inhibiting activities of thiocarbamates. *Analytica Chimica Acta*, *705*(1–2), 166–173.

Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, *26*(5), 694–701.

Granovsky, A. A., Firefly, version 7.1.G, 2010. www.http://classic.chem.msu.su/gran/firefly/index.html.

Hawkins, D. M., Basak, S. C., & Mills, D. (2003). Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences*, *43*(2), 579–586.

Higuchi, T. (1961). Rate of release of medicaments from ointment bases containing drugs in suspension. *Journal of Pharmaceutical Sciences*, *50*(10), 874–875.

Higuchi, T. (1963). Mechanism of sustained-action medication. Theoretical analysis of rate of release of solid drugs dispersed in solid matrices. *Journal of Pharmaceutical Sciences*, *52*(12), 1145–1149.

Iqbal, M. S., Akbar, J., Hussain, M. A., Saghir, S., & Sher, M. (2011). Evaluation of hot-water extracted arabinoxylans from ispaghula seeds as drug carriers. *Carbohydrate Polymers*, *83*(3), 1218–1225.

Iqbal, M. S., Akbar, J., Saghir, S., Karim, A., Koschella, A., Heinze, T., et al. (2011). Thermal studies of plant carbohydrate polymer hydrogels. *Carbohydrate Polymers*, *86*(4), 1775–1783, and references therein

Lopes, C., Sousa Lobo, J., Pinto, J., & Costa, P. (2007). Compressed matrix core tablet as a quick/slow dual-component delivery system containing ibuprofen. *AAPS PharmSciTech*, *8*(3), E195–E202.

Mittermayr, S., Olajos, M., Chovan, T., Bonn, G. K., & Guttman, A. (2008). Mobility modeling of peptides in capillary electrophoresis. *TrAC Trends in Analytical Chemistry*, *27*(5), 407–417.

Moore, J. W., & Flanner, H. H. (1996). Mathematical comparison of dissolution profiles. *Pharmaceutical Technology*, *20*(6), 64–74.

Niño-Medina, G., Carvajal-Millán, E., Rascón-Chu, A., Márquez-Escalante, J. A., Guerrero, V., & Salas-Muñoz, E. (2010). Feruloylated arabinoxylans and arabinoxylan gels: structure, sources and applications. *Phytochemistry Reviews*, *9*(1), 111–120.

Peppas, N. A. (1985). Analysis of Fickian and non-Fickian drug release from polymers. *Pharmaceutica Acta Helvetiae*, *60*(4), 110–111.

Ranga-Rao, K. V., Padmalatha-Devi, K., & Buri, P. (1990). Influence of molecular size and water solubility of the solute on Its release from swelling and erosion controlled polymeric Matrices. *Journal of Controlled Release*, *12*(2), 133–141.

Ritger, P. L., & Peppas, N. A. (1987a). A simple equation for description of solute release I. Fickian and non-Fickian release from non-swellable devices in the form of slabs, spheres, cylinders or discs. *Journal of Controlled Release*, *5*(1), 23–36.

Ritger, P. L., & Peppas, N. A. (1987b). A simple equation for description of solute release II. Fickian and anomalous release from swellable devices. *Journal of Controlled Release*, *5*(1), 37–42.

Rujivipat, S., & Bodmeier, R. (2010). Modified release from hydroxypropyl methylcellulose compression-coated tablets. *International Journal of Pharmaceutics*, *402*(1–2), 72–77.

Saíz-Urra, L., Cabrera Pérez, M. Á., Helguera, A. M., & Froeyen, M. (2011). Combining molecular docking and QSAR studies for modelling the antigyrase activity of cyclothialidine derivatives. *European Journal of Medicinal Chemistry*, *46*(7), 2736–2747.

Schuur, J. H., Selzer, P., & Gasteiger, J. (1996). The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. *Journal of Chemical Information and Computer Sciences*, *36*(2), 334–344.

Scientist-handbook, & rev. 7EEF. (1995). MicroMath, Inc., Salt Lake City, p. 467.

Siepmann, J., & Siepmann, F. (2008). Mathematical modeling of drug delivery. *International Journal of Pharmaceutics*, *364*(2), 328–343.

Stewart, J. J. P., MOPAC2009, 2009. Colorado Springs, CO, USA, http://OpenMOPACnet.

Sumathi, S., & Ray, A. R. (2002). Release behaviour of drugs from tamarind seed polysaccharide tablets. *Journal of Pharmacy and Pharmaceutical Sciences*, *5*(1), 12–18.

Tahara, K., Yamamoto, K., & Nishihata, T. (1996). Application of model-independent and model analysis for the investigation of effect of drug solubility on its release rate from hydroxypropyl methylcellulose sustained-release tablets. *International Journal of Pharmaceutics*, *133*(1–2), 17–27.

Tetko, I. V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., et al. (2005). Virtual computational chemistry laboratory—Design and description. *Journal of Computer-Aided Molecular Design*, *19*(6), 453–463.

Todeschini, R., & Gramatica, P. (1997). 3D-modelling and prediction by WHIM descriptors Part 5. Theory development and chemical meaning of WHIM descriptors. *Quantitative Structure–Activity Relationships*, *16*(2), 113–119.

Please cite this article in press as: Akbar, J., et al. A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks. *Carbohydrate Polymers* (2012), doi:10.1016/j.carbpol.2012.02.016

Todeschini, R., Lasagni, M., & Marengo, E. (1994). New molecular descriptors for 2D and 3D structures. *Theory. Journal of Chemometrics*, *8*(4), 263–272.
Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, *22*(1), 69–77.
Whitley, D. C., Ford, M. G., & Livingstone, D. J. (2000). Unsupervised forward selection: A method for eliminating redundant variables. *Journal of Chemical Information and Computer Sciences*, *40*(5), 1160–1168.
Wold, S. (1991). Validation of QSAR's. *Quantitative Structure–Activity Relationships*, *10*(3), 191–193.

Please cite this article in press as: Akbar, J., et al. A QSPR study of drug release from an arabinoxylan using *ab initio* optimization and neural networks. *Carbohydrate Polymers* (2012), doi:10.1016/j.carbpol.2012.02.016