Sci Pharm. 2008; 76: 377–394 doi:10.3797/scipharm.0803-30
© Österreichische Apotheker-Verlagsgesellschaft m. b. H., Vienna, Austria
Reproduction is permitted for non-commercial purposes.

Topological Models for Prediction of Pharmacokinetic Parameters of Cephalosporins using Random Forest, Decision Tree and Moving Average Analysis

Harish DUREJA [1], Sunil GUPTA [2], Anil Kumar MADAN * [1]

[1] Faculty of Pharmaceutical Sciences, M.D. University, Rohtak, 124 001, INDIA
[2] JCD College of Pharmacy, Sirsa, 125 055, INDIA

Abstract

The topological indices were used to encode the structureal features of cephalosporins. Both topostructural and topochemical versions of a distance based descriptor, three adjacency based descriptors and five distance-cum-adjacency based descriptors were calculated. The values of 18 indices for each cephalosporin in the dataset were computed using an in-house computer program. Multiple pharmacokinetic parameters of cephalosporins were predicted using random forest, decision tree and moving average analysis. Random forest correctly classified the pharmacokinetic parameters into low and high ranges upto 95%. A decision tree was constructed for each pharmacokinetic parameter to determine the importance of topological indices. The decision tree learned the information from the input data with an accuracy of 95% and correctly predicted the cross-validated (10 fold) data with an accuracy of upto 90%. Three independent moving average based topological models were developed using a single range for simultaneous prediction of multiple pharmacokinetic parameters. The accuracy of classification of single index based models using moving average analysis varied from 65% to 100%.

* Corresponding author: Tel.: +91-98963-46211; Fax: +91-1262-274640.
E-mail: madan_ak@yahoo.com (A. K. Madan).

Keywords

Topological indices • Random forest • Decision tree • Moving average analysis • Pharmacokinetic parameters • Cephalosporins.

Introduction

The pharmaceutical industry need to develop continuously new medicinal drugs in order to fight the development of resistance in pathogenic agents, and to cope with newly discovered types of infections [1]. Since ADME (absorption, distribution, metabolism and elimination) properties are important parameters in lead identification, the *in silico* methods to search for drug candidates with good ADME properties has attracted the pharmaceutical industry [2–4].

Various quantitative structure-activity relationship (QSAR) approaches have been applied to find relationships between ADME parameters and molecular structure and properties. The polarizability and transition state energy of a cephalosporin were used to predict permeability through the outer membrane and of the reactivity of β-lactam ring with penicillin binding proteins. The activity exhibited quadratic dependence on the variables [5]. In another QSAR study lipophilicity and electronic and hydrogen bonding parameter were used as molecular descriptors. It was found that polar-polar interactions of hydrophilic penicillins and cephalosporins could be explained on the basis of hydrogen bonding properties [6]. Turner et al. [7] predicted multiple pharmacokinetic parameters for a series of cephalosporins using artificial neural network. Further, artificial neural networks (ANNs) were used for the prediction of clearances, fraction bound to plasma proteins, and volume of distribution of a series of structurally diverse compounds. Simple methods for determining the human pharmacokinetics of known and drug-like compounds are of interest to pharmaceutical industry [8]. Genetic algorithm-combined with partial least squares were used for modeling ADME properties of structurally diverse compounds. Many ADME properties could

be well explained by simple molecular descriptors derived from 2-dimensional chemical structure [9].

Aim of the present study was to develop simple models for the prediction of multiple pharmacokinetic parameters using topological descriptors obtained from 2-dimensional chemical structure. The predictability of the proposed models using random forest, decision tree and moving average analysis has been compared in the present study. Finally, single index range models derived from moving average analysis for the simultaneous classification of multiple pharmacokinetic parameters into low and high values have also been proposed in the present study.

Computational Methods

Dataset

Turner et al [7] compiled various pharmacokinetic parameters of cephalosporins such as $t_{1/2}$, CL, CL_R, f_e, V and f_b. The half-life was reported quantitatively as $t_{1/2}$ (h). For the present study, cephalosporins were considered to exhibit low $t_{1/2}$ - labeled as “A” (N=13) if they exhibited $t_{1/2}$ value < 2.0 h and high $t_{1/2}$ - labeled as “B” (N=7) if the $t_{1/2}$ value was 2.0 or more. Similarly, the clearance was reported quantitatively as CL ($mL.min^{-1}.kg^{-1}$), the renal clearance was reported as CL_R ($mL.min^{-1}.kg^{-1}$), and the volume of distribution at steady state was reported as V (L/kg). The fraction excreted unchanged in the urine was reported quantitatively as f_e and fraction bound to plasma proteins was reported as f_b. The cephalosporins were considered to exhibit low CL – labeled as “A” (N=5) if they exhibited CL < 1.0 $mL.min^{-1}.kg^{-1}$ and high CL – labeled as “B” (N=15) if they exhibited CL ≥ 1.0 $mL.min^{-1}.kg^{-1}$. These cephalosporins were considered to exhibit low CL_R – labeled as “A” (N=7) if they exhibited CL_R < 1.0 $mL.min^{-1}.kg^{-1}$ and high CL_R – labeled as “B” (N=13) if the $CL_R \geq 1.0$ $mL.min^{-1}.kg^{-1}$. Cephalosporins were also considered to exhibit low f_e – labeled as “A” (N=8) if they exhibited f_e < 0.7 and high f_e – labeled as “B” (N=12) if they exhibited $f_e \geq 0.7$. These cephalosporins were considered to exhibit low V – labeled as “A” (N=8) if they exhibited V< 0.2 and high V – labeled as “B” (N=12) if

they exhibited V ≥ 0.2. The cephalosporins were considered to exhibit low f_b – labeled as "A" (N=14) if they exhibited $f_b < 0.8$ and high f_b – labeled as "B" (N=6) if they exhibited $f_b \geq 0.8$.

Tab. 1. Topostructural and topochemical indices

Code	Index	Reference
A1	Molecular connectivity topochemical index	10, 11
A2,	Eccentric adjacency topochemical index	12
A3	Augmented eccentric connectivity topochemical index	13
A4	Superadjacency topochemical index	14
A5	Eccentric connectivity topochemical index	15
A6	Connective eccentricity topochemical index	16
A7	Zagreb topochemical index, M_1^c	17
A8	Zagreb topochemical index, M_2^c	17
A9	Wiener's topochemical index	18
A10	Molecular connectivity index	19
A11	Eccentric adjacency index	20
A12	Augmented eccentric connectivity index	21
A13,	Superadjacency index	14
A14	Eccentric connectivity index	22
A15	Connective eccentricity index	23
A16	Zagreb group parameter, M_1	24, 25
A17	Zagreb group parameter, M_2	24, 25
A18	Wiener's index	26, 27

Topostructural and topochemical indices

The nine topostructural and nine topochemical indices used for the present study are presented in Tab. 1 [10–27]. The distance based topological descriptor (Wiener's index), adjacency based descriptors (Zagreb group parameter, M_1 and

M_2, molecular connectivity index) and distance-cum-adjacency based topological descriptors (eccentric adjacency index, augmented eccentric connectivity index, superadjacency index, eccentric connectivity index, connective eccentricity index) were calculated using in-house computer program. The topochemical descriptors of topostructural descriptors calculated above were calculated from distance and adjacency matrices weighted by molecular mass with respect to that of carbon atom.

Random Forest

Random forest (RF) was grown for each pharmacokinetic parameter separately. Random forest is an ensemble of unpruned classification trees created by using bootstrap samples of the training data and random feature selection in tree induction. Prediction was made by majority vote of the individual trees. In this study, the RFs were grown with the R program (version 2.1.0) using the randomForest library.

Decision tree

A single decision tree [28] was grown, for each property, to identify the importance of topological indices. In a decision tree, the molecules at each parent node are classified, based on the index value, into two child nodes. The prediction for a molecule reaching a given terminal node is obtained by majority vote of the molecules reaching the same terminal node in training set. The tree giving the lowest value of error in cross-validation is selected as optimal tree. In this study, R program (version 2.1.0) along with the RPART library was used to grow decision tree.

Moving average analysis

To construct single topological index based model for predicting property/activity based ranges, moving average analysis of correctly predicted compounds was used [20]. According to this method the minimum size of range is based on moving average of 65% of the correctly predicted compounds. However if

the moving average percentage of correct prediction lies between 50±15%, it is classified as transitional range. The characterisitic property assigned to each drug was compared with reported property.

Results and Discussion

The random forests were grown with 18 topologicaldescriptors. The importance of node was determined by mean decrease in accuracy and purity of the node was determined by mean decrease in Gini. The precision and sensitivity of classification was also determined. The precision is a measure of accuracy, provided that a specific class has been predicted. The sensitivity is the ability of a predicted model to select certain instances of a certain class from a dataset. The RF classified the $t_{1/2}$ of cephalosporins with an accuracy of 85% and out-of-bag (OOB) estimate of error was 15%. The precision and sensitivity of low $t_{1/2}$ was of the order of 92% and 85%, whereas the precision and sensitivity of high $t_{1/2}$ was of the order of 75% and 86% respectively. A1, molecular connectivity topochemical index and A12, augmented eccentric connectivity index were identified as the most important descriptors. The RF classified the CL of cephalosporins with an accuracy of 90% and OOB estimate of error was 10%. The precision and sensitivity of low CL was of the order of 80% and 80%, whereas the precision and sensitivity of high CL was of the order of 93% and 93% respectively. A8, Zagreb topochemical index, M_2^c and A14, eccentric connectivity index were identified as the most important descriptors. The RF classified the CL_R of cephalosporins with an accuracy of 90% and OOB estimate of error was 10%. The precision and sensitivity of low CL_R was of the order of 100% and 71%, whereas the precision and sensitivity of high CL_R was of the order of 87% and 100% respectively. A5, eccentric connectivity topochemical index and A8, Zagreb topochemical index, M_2^c were identified as the most important descriptors. The RF OOB estimate of error was 40% for f_e because only 3 out of 8 compounds were correctly classified as low f_e, although 9 out of 12 compounds were correctly classified as high f_e. The precision and sensitivity of low f_e was of the order of 50% and 38%, whereas the precision and sensitivity of high f_e

Tab. 2. Confusion matrix for multiple pharmacokinetic parameters ($t_{1/2}$, CL, CL_R, f_e, V and f_b) using the models based on random forest, decision tree and moving average analysis

Property	Ranges	Number of compounds predicted using random forest		Number of compounds predicted using decision tree[a]	
		Low	High	Low	High
$t_{1/2}$	Low $t_{1/2}$	11	02	08	05
	High $t_{1/2}$	01	06	01	06
CL	Low CL	04	01	04	01
	High CL	01	14	02	13
CL_R	Low CL_R	05	02	05	02
	High CL_R	00	13	03	10
f_e	Low f_e	03	05	03	05
	High f_e	03	09	03	09
V	Low V	06	02	06	02
	High V	00	12	01	11
f_b	Low f_b	14	00	13	01
	High f_b	01	05	01	05

[a] The predictions from decision tree were obtained by tenfold cross-validation.

was of the order of 64% and 75% respectively. A11, eccentric adjacency index and A13, superadjacency index were identified as the most important descriptor. The RF classified the cephalosporins with regard to V with an accuracy of 90% and out-of-bag estimate of error was only 10%. The precision and sensitivity of low V was of the order of 100% and 75%, whereas the precision and sensitivity of high V was of the order of 86% and 100% respectively. A5, eccentric connectivity topochemical index, A9, Wiener's topochemical index and A14, eccentric connectivity index were identified as the most important descriptors The RF classified the cephalosporins with regard to f_b with an accuracy of 95% and out-of-bag estimate of error was only 5%. The precision and sensitivity of low f_b was of the order of 93% and 100%, whereas the precision and sensitivity of high f_b was of the order of 100% and 83% respectively. A5, eccentric connectivity topochemical index, A7, Zagreb topochemical index, M_1^c and A8, Zagreb topochemical index, M_2^c were identified as

the most important descriptors. The predictions for multiple pharmacokinetic parameters using RF were found to be upto 95% (Tab. 2).

The decision tree was built from a set of 18 topological indices. The index at the root node is most important and the importance of index decreases as the length of tree increases. The classification of $t_{1/2}$ using a single tree, based on A1, molecular connectivity topochemical index and A2, eccentric adjacency topochemical index is shown in Fig. 1. The decision tree identified molecular connectivity topochemical index (A1) as the most important index. The decision tree classified the cephalosporins in the training set with an accuracy of 95% and in 10 fold cross-validation, 70% cephalosporins were correctly classified with regard to $t_{1/2}$. In cross-validation, the precision and sensitivity of low $t_{1/2}$ was of the order of 89% and 62%, whereas the precision and sensitivity of high $t_{1/2}$ was of the order of 55% and 86% respectively (Tab. 2). The classification of CL using decision tree, based on A5 eccentric connectivity topochemical index is shown in Fig. 1. The tree correctly classified cephalosporins in the training set with an accuracy of 95%. In 10 fold cross-validation, 85% cephalosporins were correctly classified with regard to CL. In cross-validation, the precision and sensitivity of low CL was 67% and 80%, whereas the precision and sensitivity of high CL was 93% and 87% respectively (Tab. 2). The classification of CL_R using single tree based on A5, eccentric connectivity topochemical index is shown in Fig. 1. The tree correctly classified cephalosporins in the training set with an accuracy of 95%. In 10 fold cross-validation, 75% cephalosporins were correctly classified with regard to CL_R. In cross-validation, the precision and sensitivity of low CL_R was 63% and 71%, whereas the precision and sensitivity of high CL_R was 83% and 77% respectively (Tab. 2). The classification of f_e using A11, eccentric adjacency index and A5, eccentric connectivity topochemical index is shown in Fig. 1. According to decision tree, eccentric adjacency index (A11) was the most important index. The tree classified the cephalosporins in the training set with an accuracy of 90%. In 10 fold cross-validation, 60% cephalosporins were classified correctly with regard to f_e. In cross-validation, the precision and sensitivity of low f_e was 50% and 38%, whereas

the precision and sensitivity of high f_e was 64% and 75% respectively (Tab. 2). The classification of V using decision tree based on A5, eccentric connectivity topochemical index is shown in Fig. 1. The tree correctly classified cephalosporins in the training set with an accuracy of 100%. In 10 fold cross-validation, 85% cephalosporins were correctly classified with regard to V. In cross-validation, the precision and sensitivity of low V was 86% and 75%, whereas the precision and sensitivity of high V was 85% and 92%, respectively (Tab. 2). The classification of f_b using a single tree based on A5, eccentric connectivity topochemical index is shown in Fig. 1. The tree classified the cephalosporins in the training set with an accuracy of 100%. In 10 fold cross-validation, 90% cephalosporins were classified correctly with regard to f_b. In cross-validation, the precision and sensitivity of low f_b was 93% and 93%, whereas the precision and sensitivity of high f_b was 83% and 83% respectively (Tab. 2). The decision tree learned the information from the input data with an accuracy of more than 95 % and predicted the cross-validated (10 fold) data with an accuracy of up to 90%.

The result obtained using single tree agree in principle with those obtained using random forest. The strength of random forest lies in out-of-bag error of estimate. Since decision tree is easy to interpret and can be visualized, the importance of descriptors was taken from decision trees. The variables selected by the tree can be different from random forest because decision tree results are based on single tree while random forest results are average of many trees. The single decision tree sometime assigns the same importance to more than one descriptor and selects one descriptor at random whereas random forest assign importance based on the average of all the individual trees.

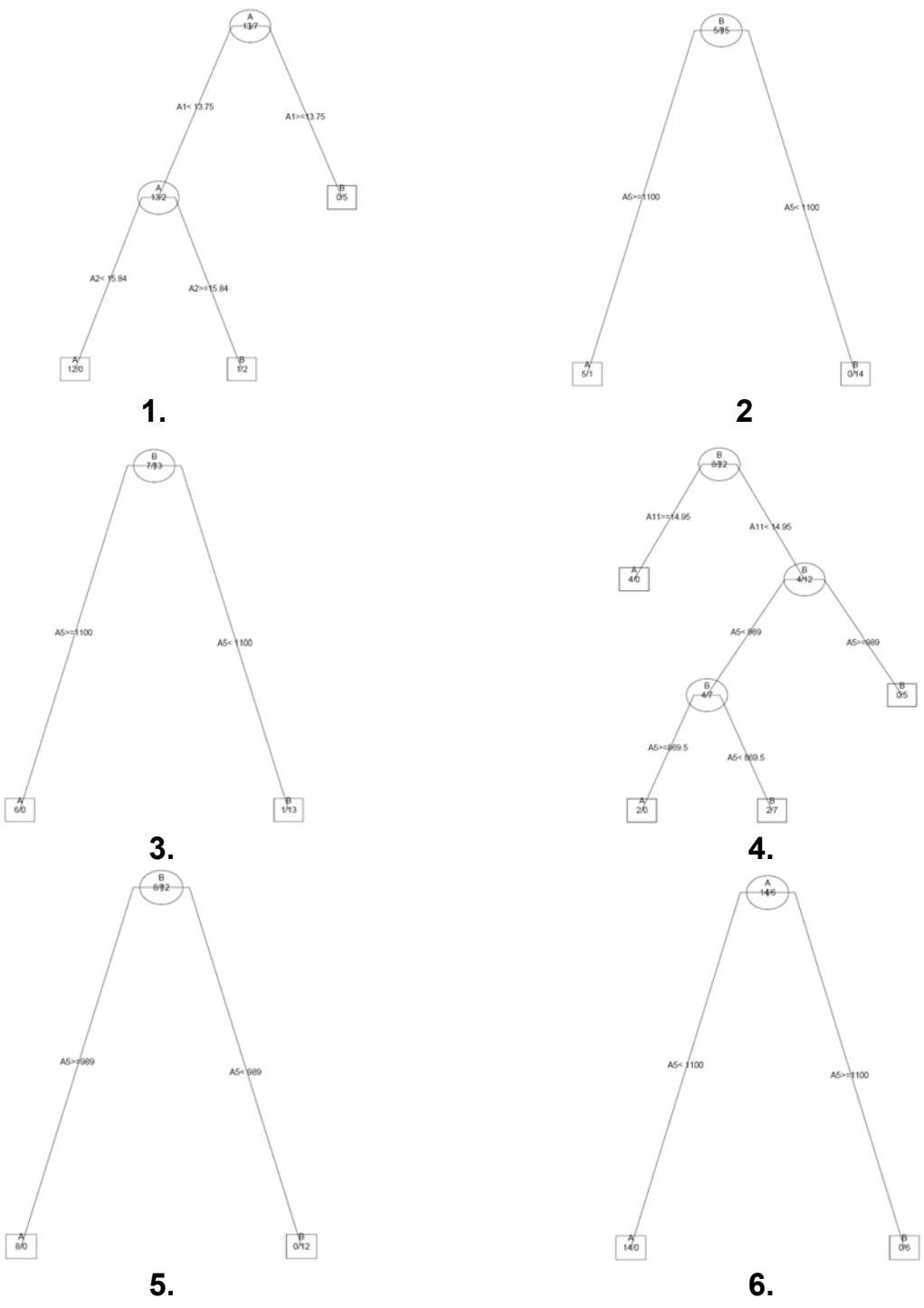

Fig. 1. The decision tree for distinguishing low value – A from high value –B; 1. t1/2 (A1, molecular connectivity topochemical index; A2, eccentric adjacency topochemical index); 2. CL (A5, Eccentric connectivity topochemical index); 3. CL_R (A5, Eccentric connectivity topochemical index); 4. f_e (A11, eccentric adjacency index, A5, Eccentric connectivity topochemical index); 5. V (A5, Eccentric connectivity topochemical index); 6. f_b (A5, Eccentric connectivity topochemical index)

Tab. 3. Accuracy of classification for multiple pharmacokinetic parameters ($t_{1/2}$, CL, CL_R, f_e, V and f_b) using the models based on moving average analysis.

Index	Property	Nature of range	Index value	Total drugs in the range	Number of drugs predicted correctly	Precision (%)	Sensitivity (%)
A5	$t_{1/2}$	Low $t_{1/2}$	< 1132.326	14	12	92	86
		High $t_{1/2}$	≥ 1132.326	06	05	71	83
	CL	High CL		14	14	93	100
		Low CL		06	05	100	83
	CL_R	High CL_R		14	13	100	93
		Low CL_R		06	06	86	100
	f_e	High f_e		14	09	75	64
		Low f_e		06	03	38	50
	V	High V		14	12	100	86
		Low V		06	06	75	100
	f_b	Low f_b		14	14	100	100
		High f_b		06	06	100	100
A1	$t_{1/2}$	Low $t_{1/2}$	< 14.266	15	13	100	87
		High $t_{1/2}$	≥ 14.266	05	05	71	100
	CL	High CL		15	14	100	93
		Low CL		05	05	83	100
	CL_R	High CL_R		15	13	100	87
		Low CL_R		05	05	71	100
	f_e	High f_e		15	10	83	67
		Low f_e		05	03	38	60
	V	High V		15	12	100	80
		Low V		05	05	63	100
	f_b	Low f_b		15	14	100	93
		High f_b		05	05	83	100
A11	$t_{1/2}$	Low $t_{1/2}$	< 15.076	16	13	100	81
		High $t_{1/2}$	≥ 15.076	04	04	57	100
	CL	High CL		16	13	87	81
		Low CL		04	02	40	50
	CL_R	High CL_R		16	13	100	81
		Low CL_R		04	04	57	100
	f_e	High f_e		16	12	100	75
		Low f_e		04	04	50	100
	V	High V		16	11	92	69
		Low V		04	03	38	75
	f_b	Low f_b		16	13	93	81
		High f_b		04	03	50	75

A5, eccentric connectivity topochemical index; A1, molecular connectivity topochemical index; A11, eccentric adjacency index.

The property based ranges were identified using moving average analysis [18]. Three independent moving average analysis based models were developed using a single index at a time. The three topological indices identified as most important indices by decision trees were used to construct single index based model for simultaneous prediction of multiple pharmacokinetic parameters. The

precision and sensitivity of classification for multiple pharmacokinetic parameters ($t_{1/2}$, CL, CL_R, f_e, V and f_b) using moving average analysis is summarized in Tab. 3.

Tab. 4. Prediction of multiple pharmacokinetic parameters ($t_{1/2}$, CL, CL_R, f_e, V and f_b) by moving average analysis using eccentric connectivity topochemical index (A5).

S. N.	Drug	Index Value A5	Multiple Pharmacokinetic parameters											
			Reported						Predicted using A5					
			$t_{1/2}$	CL	CL_R	f_e	V	f_b	$t_{1/2}$	CL	CL_R	f_e	V	f_b
1	Cefaclor	624.38	−	+	+	+	+	−	−	+	+	−	+	−
2	Cefadroxil	641.27	−	+	+	+	+	−	−	+	+	+	+	−
3	Cefamandole	1067.879	−	+	+	+	+	−	−	+	+	+	−	−
4	Cefazolin	1132.326	+	−	−	−	−	+	−	−	−	+	−	+
5	Cefixime	866.83	−	+	+	+	+	−	+	+	−	−	+	−
6	Cefmetazole	999.611	−	+	+	+	+	−	−	+	+	+	−	−
7	Cefonicid	1613.06	+	−	−	−	−	+	+	−	−	+	−	+
8	Cefoperazone	1860.369	+	−	−	−	−	+	+	+	−	−	−	+
9	Ceforanide	1334.465	+	−	−	−	−	+	+	−	−	+	−	+
10	Cefotaxime	926.885	−	+	+	+	+	−	−	+	+	−	+	−
11	Cefotetan	1603.015	+	−	−	−	−	+	+	−	−	−	−	+
12	Cefpodoxime	812.176	−	+	+	+	+	−	+	+	+	+	+	−
13	Cefprozil	741.614	−	+	+	+	+	−	−	+	+	+	+	−
14	Ceftizoxime	657.066	−	+	+	+	+	−	−	+	+	+	+	−
15	Ceftriaxone	1467.932	+	−	−	−	−	+	+	−	−	−	−	+
16	Cephalexin	563.729	−	+	+	+	+	−	−	+	+	+	+	−
17	Cephalothin	767.439	−	+	+	+	+	−	−	+	+	−	+	−
18	Cephapirin	978.374	−	+	+	+	+	−	−	+	+	−	+	−
19	Cephradine	563.729	−	+	+	+	+	−	−	+	+	+	+	−
20	Loracarbef	582.757	−	+	+	+	+	−	−	+	+	+	+	−

−, low $t_{1/2}$, CL, CL_R, f_e, V and f_b drug; +, high $t_{1/2}$, CL, CL_R, f_e, V and f_b drug.

Though three independent models were developed, the classification of multiple pharmacokinetic parameters ($t_{1/2}$, CL, CL_R, f_e, V and f_b) was based on single range of the topological indices A5, eccentric connectivity topochemical index, A1 molecular connectivity topochemical index and eccentric adjacency index, A11 (Tab. 4–5).

Tab. 5. Prediction of multiple pharmacokinetic parameters ($t_{1/2}$, CL, CL_R, f_e, V and f_b) by moving average analysis using molecular connectivity topochemical index (A1) and eccentric adjacency index (A11).

S. N.	Drug	Index Value		Multiple pharmacokinetic parameters predicted using A1						Multiple pharmacokinetic parameters predicted using A11					
		A1	A11	$t_{1/2}$	CL	CL_R	f_e	V	f_b	$t_{1/2}$	CL	CL_R	f_e	V	f_b
1	Cefaclor	10.264	14.336	−	+	+	+	+	−	−	+	+	+	+	−
2	Cefadroxil	10.854	13.88	−	+	+	+	+	−	−	+	+	+	+	−
3	Cefamandole	13.227	14.674	−	+	+	+	+	−	−	+	+	+	+	−
4	Cefazolin	11.788	14.04	−	+	+	+	+	−	−	+	+	+	+	−
5	Cefixime	12.466	15.162	−	+	+	+	+	−	+	−	−	−	−	+
6	Cefmetazole	12.584	14.322	−	+	+	+	+	−	−	+	+	+	+	−
7	Cefonicid	14.266	13.905	+	−	−	−	−	+	−	+	+	+	+	−
8	Cefoperazone	18.897	16.227	+	−	−	−	−	+	+	−	−	−	−	+
9	Ceforanide	15.022	14.496	+	−	−	−	−	+	−	+	+	+	+	−
10	Cefotaxime	12.362	14.643	−	+	+	+	+	−	−	+	+	+	+	−
11	Cefotetan	14.369	15.299	+	−	−	−	−	+	+	−	−	−	−	+
12	Cefpodoxime	11.561	14.600	−	+	+	+	+	−	−	+	+	+	+	−
13	Cefprozil	11.892	13.319	−	+	+	+	+	−	−	+	+	+	+	−
14	Ceftizoxime	10.248	14.822	−	+	+	+	+	−	−	+	+	+	+	−
15	Ceftriaxone	14.660	15.076	+	−	−	−	−	+	+	−	−	−	−	+
16	Cephalexin	10.531	14.336	−	+	+	+	+	−	−	+	+	+	+	−
17	Cephalothin	10.804	14.002	−	+	+	+	+	−	−	+	+	+	+	−
18	Cephapirin	11.757	12.754	−	+	+	+	+	−	−	+	+	+	+	−
19	Cephradine	10.531	14.336	−	+	+	+	+	−	−	+	+	+	+	−
20	Loracarbef	10.676	14.336	−	+	+	+	+	−	−	+	+	+	+	−

−, low $t_{1/2}$, CL, CL_R, f_e, V and f_b drug; +, high $t_{1/2}$, CL, CL_R, f_e, V and $f_{b\,0}$ drug.

It is surprising that topostructural eccentric adjacency index was identified as one of the important index along with topochemical indices and also a single range could be identified. One would expect this to happen because topostructural indices are insensitive to topochemical isomers. Therefore, we evaluated the intercorrelation of eccentric adjacency index values with that of A5 and A1 using all possible structures upto 5 vertices (all 29 structures varying only with respect to connectivity and not topochemical nature). A11 indeed exhibited poor correlation with A5 and A1.

The cephalosporins were correctly classified as exhibiting low $t_{1/2}$ or exhibiting high $t_{1/2}$ using eccentric connectivity index (A5) with an accuracy of 85%. Eccentric

connectivity index (A5), classified the cephalosporins as exhibiting high CL, high CL_R, high f_e, high V or exhibiting low CL, low CL_R, low f_e, low V with an accuracy of 95%, 95%, 60% and 90%, respectively. All the cephalosporins were correctly classified as exhibiting low f_b or exhibiting high f_b.

The single index range model based on eccentric connetivity index can simultaneously predict the multiple pharmacokinetic parameters. Similarly, the single range model based on molecular connectivity topochemical index (A1) correctly classified the cephalosporins as exhibiting low $t_{1/2}$ or exhibiting high $t_{1/2}$ with an accuracy of 90%. The cephalosporins were also correctly classified as exhibiting high CL, high CL_R, high f_e, high V or exhibiting low CL, low CL_R, low f_e, low V with an accuracy of 90%, 90%, 65% and 85%, respectively. The cephalosporins were correctly classified as exhibiting low f_b or exhibiting high f_b with an accuracy of 95%. The cephalosporins were also correctly classified as exhibiting low $t_{1/2}$ or exhibiting high $t_{1/2}$ using eccentric adjacency index (A11) with an accuracy of 85%. Eccentric adjacency index (A11), classified the cephalosporins as exhibiting high CL, high CL_R, high f_e, high V or exhibiting low CL, low CL_R, low f_e, low V with an accuracy of 75%, 85%, 80% and 70%, respectively. The cephalosporins were correctly classified as exhibiting low f_b or exhibiting high f_b with an accuracy of 80%.

It is noteworthy that the threshold index values for classification of compounds into high or low pharmacokinetic properties using moving average analysis may appear different from those obtained using decision tree. The apparent differences can be attributed to the fact that topological index values identified using moving average analysis were strictly based on the index value of drugs in the dataset, whereas the ranges of index values obtained from decision tree may refer to drug that is not present in the dataset used to obtain decision tree.

Conclusion

To identify important descriptors and to predict the multiple pharmacokinetic parameters of cephalosporins RF and decision tree were constructed. Single index

range models derived from moving average analysis were proposed for for simultaneous classification of all pharmacokinetic parameters. Authors have comapred the calssification ability of of RF, decision tree and moving average analysis in predicting multiple pharmacokinetic parameters. The topostructural and topochemical indices utilized to classify the multiple pharmacokinetic parameters indicate that they are capable of encoding latent features of cephalosporins that are not visible in terms of structural similarity.

References

[1] Devillers J, Balaban AT, editors.
Historical Developments of Topological Indices.
In: Topological Indices and Related Descriptors in QSAR and QSPR.
The Netherlands: Gordon and Breach Science Publishers, 1999: 21–57.

[2] Bai JPF, Utis A, Crippen G, He H-D, Fischer V, Tullman R, Yin H-Q, Hsu C-P, Jiang L, Hwang K-K.
Use of Classification Regression Tree in Prediction of Oral Absorption in Humans.
J Chem Inf Comput Sci. 2004; 44: 2061–2069.
[doi:10.1021/ci040023n]

[3] van de Waterbeemd H, Gifford E.
ADMET *in silico* Modeling: Towards Prediction or Paradise?
Nat Rev Drug Discovery. 2003; 2: 192–204.
[doi:10.1038/nrd1032]

[4] Perez MAC, Sanz MB, Torres LR, Avalos RG, Gonzalez MP, Diaz HG.
A Topological Sub-Structural Approach for Predicting Human Intestinal Absorption of Drugs.
Eur J Med Chem. 2004; 39: 905–916.
[doi:10.1016/j.ejmech.2004.06.012]

[5] Choi J-H, Kim H.
Structure-Activity Relationship Study on Cephalosporins with Mechanism-Based-Descriptors.
Bull Korean Chem Soc. 1993; 14: 631–635.

[6] Ferreira MMC, Kiralj R.
QSAR Study of β-lactam Antibiotic Efflux by the Bacterial Multidrug Resistance Pump AcrB.
J Chemomet. 2004; 18: 242–252.
[doi:10.1002/cem.867]

[7] Turner JV, Maddalena DJ, Cutler DJ, Agatonovic-Kustrin S.
Multilple Pharmacokinetic Parameter Prediction for a Series of Cephalosporins.
J Pharm Sci. 2003; 92: 552–559.
[doi:10.1002/jps.10314]

[8] Turner JV, Maddalena DJ, Cutler DJ.
Pharmacokinetic Parameter Prediction from Drug Structure using Artificial Neural Networks.
Int J Pharm. 2004; 270: 209–219.
[doi:10.1016/j.ijpharm.2003.10.011]

[9] Yamashita F, Fujiwara S-I, Wanchana S, Hashida M.
Quantitative Structure/Activity Relationship Modelling of Pharmacokinetic Properties using Genetic Algorithm-Combined Partial Least Squares Method.
J Drug Target. 2006; 14: 496–504.
[doi:10.1080/10611860600844895]

[10] Goel A, Madan AK.
Structure-Activity Study on Anti-Inflammatory Pyrazole Carboxylic Acid Hydrazide Analogs using Molecular Connectivity Indices.
J Chem Inf Comput Sci. 1995; 35: 510–514.
[doi:10.1021/ci00025a019]

[11] Dureja H, Madan, AK.
Topochemical Models for Prediction of Cyclin-Dependent Kinase 2 Inhibitory Activity of Indole-2-Ones.
J Mol Mod. 2005; 11: 525–531.
[doi:10.1007/s00894-005-0276-3]

[12] Gupta S, Singh M, Madan AK.
Novel Topochemical Descriptors for Predicting Anti-Hiv Activity.
Indian J Chem. 2003; 42A: 1414–1425.

[13] Bajaj S.
Study on Topochemical Descriptors for the Prediction of Physicochemical and Biological Properties of Molecules.
Ph.D. Thesis, Guru Gobind Singh Indraprastha University, India, 2005.

[14] Bajaj S, Sambi SS, Madan AK.
Prediction of Carbonic Anhydrase Activation by Tri-/Tetrasubstituted-Pyridinium-Azole Drugs: A Computational Approach using Novel Topochemical Descriptor.
QSAR Comb Sci. 2004; 23: 506–514.
[doi:10.1002/qsar.200439999]

[15] Kumar V, Sardana S, Madan AK.
Predicting Anti-Hiv Activity of 2,3-Diaryl-1,3-Thiazolidin-4-Ones: Computational Approach using Reformed Eccentric Connectivity Index.
J Mol Mod. 2004; 10: 399–407.
[doi:10.1007/s00894-004-0215-8]

[16] Gupta S.
Application and Development of Graph Invariants for Drug Design.
Ph.D. Thesis, Punjabi University, Patiala, India, 2002.

[17] Bajaj S, Sambi SS, Madan AK.
Prediction of Anti-Inflammatory Activity of N-Arylanthranilic Acids: Computational Approach using Refined Zagreb Indices.
Croat Chem Acta. 2005; 78: 165–174.

[18] Bajaj S, Sambi SS, Madan AK.
Predicting Anti-HIV Activity of Phenethylthiazolethiourea (PETT) Analogs: Computational Approach using Wiener's Topochemical Index.
J Mol Str. (THEOCHEM) 2004; 684: 197–203.
[doi:10.1016/j.theochem.2004.01.052]

[19] Randic M.
On Characterization of Molecular Branching.
J Am Chem Soc. 1975; 97: 6609–6615.
[doi:10.1021/ja00856a001]

[20] Gupta S, Singh M, Madan AK.
Predicting Anti-HIV Activity: Computational Approach using Novel Topological Indices.
J Comput Aided Mol Des. 2001; 15: 671–678.
[doi:10.1023/A:1011964003474]

[21] Bajaj S, Sambi SS, Madan, AK.
Model for Prediction of Anti-Hiv Activity of 2-Pyridinone Derivatives using Novel Topological Descriptor.
QSAR Comb Sci. 2006; 25: 813–823.
[doi:10.1002/qsar.200430918]

[22] Sharma V, Goswami R, Madan, AK.
Eccentric Connectivity Index; A Novel Highly Discriminating Topological Descriptor for Structure Property and Structure Activity Studies.
J Chem Inf Comput Sci. 1997; 37: 273–282.
[doi:10.1021/ci960049h]

[23] Gupta S, Singh M, Madan, AK.
Connective Eccentricity Index: A Novel Topological Descriptor for Predicting Biological Activity.
J Mol Graph Mod. 2000; 18: 18–25.
[doi:10.1016/S1093-3263(00)00027-9]

[24] Gutman I, Ruscic B, Trinajstic N, Wicox, CF.
Graph Theory and Molecular Orbitals XII Acyclic Polyenes.
J Chem Phys. 1975; 62: 3399–3405.
[doi:10.1063/1.430994]

[25] Gutman I, Randic M.
Algebric Characterization of Skeletal Branching.
Chem Phys Lett. 1977; 47:15–19.
[doi:10.1016/0009-2614(77)85296-2]

[26] Wiener H.
Correlation of Heat of Isomerization and Difference in Heat of Vaporization of Isomers among Paraffin Hydrocarbons.
J Am Chem Soc. 1947; 69: 2636–2638.
[doi:10.1021/ja01203a022]

[27] Wiener H.
Influence of Interatomic Forces on Paraffin Properties.
J Chem Phys. 1947; 15: 766.
[doi:10.1063/1.1746328]

[28] Breiman L, Friedman JH, Olshen RA, Stone CJ.
In: Classification and Regression Trees.
Chapman & Hall/CRC, Boca Raton, 2000.

Received March 27th, 2008
Accepted (after revision) June 24th, 2008
Available online at www.scipharm.at June 30th, 2008