

Internet Electronic Journal of Molecular Design

March 2004, Volume 3, Number 3, Pages 150–162

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday
Part 9

Guest Editor: Douglas J. Klein

A Combinatorial Protocol in Multiple Linear Regression to Model Gas Chromatographic Response Factor of Organophosphonate Esters

Yenamandra S. Prabhakar¹

¹ Medicinal Chemistry Division, Central Drug Research Institute, Lucknow 226 001, India

Received: August 25, 2003; Revised: November 28, 2003; Accepted: February 19, 2004; Published: March 31, 2004

Citation of the article:

Y. S. Prabhakar, A Combinatorial Protocol in Multiple Linear Regression to Model Gas Chromatographic Response Factor of Organophosphonate Esters, *Internet Electron. J. Mol. Des.* 2004, 3, 150–162, <http://www.biochempress.com>.

A Combinatorial Protocol in Multiple Linear Regression to Model Gas Chromatographic Response Factor of Organophosphonate Esters[#]

Yenamandra S. Prabhakar^{1,*}

¹ Medicinal Chemistry Division, Central Drug Research Institute, Lucknow 226 001, India

Received: August 25, 2003; Revised: November 28, 2003; Accepted: February 19, 2004; Published: March 31, 2004

Internet Electron. J. Mol. Des. 2004, 3 (3), 150–162

Abstract

Motivation. Organophosphorus compounds are a well known class of toxic chemicals which find their way into ecosystem due to their wide spread use. Their detection, identification and quantification are cause of concern world over. In environmental samples these compounds are detected and estimated through the gas chromatographic response factor. This prompted us to study the quantitative structure–response relationships (QSRR) of gas chromatographic response factor of organophosphonate esters.

Method. In this study attempts have been made to rationalize the gas chromatographic response factor of twenty–eight organophosphonates in terms of their physicochemical and electronic descriptors. Combinatorial Protocol in Multiple Linear Regression (CP–MLR), a ‘filter’ based variable selection procedure for model development in structure–activity or property relationship studies, has been used for the variable selection and identification of diverse QSRR models of the GC response factor of organophosphonates.

Results. The study has resulted in the identification of ten models (equations), having two or three descriptor each, to account for the response factor of organophosphonates (cross–validated R^2 or Q^2 is 0.88 to 0.95). The response factor of the compounds is strongly correlated with the total refractivity (TREF), molecular weight (MW) and thermodynamic properties, *e.g.*, enthalpy of vaporization (ENTH). In the study, alkyl groups of these compounds have shown two–fold influence (namely, steric and branching effect) on the response factor. Also, the study suggests that the polarization of (d–p) π bond of P=O_a in these compounds plays a critical role in the formation of the responding species.

Conclusions. The steric and electronic properties of organophosphonates play a determining role in the predictive aspect of their gas chromatographic response factor. Also the study suggested a mechanism for the formation of the responding species.

Keywords. Quantitative structure–response relationships (QSRR); combinatorial protocol in multiple linear regression (CP–MLR); variable selection; organophosphonates; gas chromatographic response factor; mechanism of responding species formation.

Abbreviations and notations

CP–MLR, Combinatorial Protocol in Multiple Linear Regression	RF, response factor
GC–TID, Gas Chromatograph equipped with Thermionic Nitrogen–Phosphorus Detector	RRF, relative response factor
	QSRR, quantitative structure–response relationships
	QSPR, quantitative structure–property relationships

[#] Dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday.

* Correspondence author; phone: +91–0522–2212411; fax: +91–0522–2223405; E–mail: yenpra@yahoo.com.

1 INTRODUCTION

With ever-increasing awareness of the hazards associated with the toxic chemicals discharge into the ecosystem, their detection, identification and quantification have become the prime concern of the regulatory and environmental protection agencies. Organophosphorus compounds are one such well known toxic chemicals which find their way into ecosystem due to their wide spread use, for example in pesticides, insecticides and as well as in chemical warfare agents, and are cause of concern world over. The Gas Chromatograph equipped with Thermionic Nitrogen–Phosphorus Detector (GC–TID) is one of the most sensitive and important tools for the analysis of these compounds [1]. In GC–TID, interaction of these compounds with the excited hot alkali ceramic bead (rubidium, Rb), the principal component of this device, leads to the formation of specific responding species in the form of charge carriers namely, $(\text{P}=\text{O})^-$ and $(\text{O}-\text{P}=\text{O})^-$, which is measured and quantified in the detector as response factor (RF) [1]. In the light of this, Saradhi and co-workers recently reported the GC–TID response factors of some organophosphonates (Figure 1) and suggested that in thermionic detection process, the molecule and the alkali bead/electron may undergo a sort of chemical reaction but not simple combustion [2].



Figure 1. (a) General structure for the 28 compounds considered in the QSRR study. Here R_1 is H, Me, Et, and *i*-Pr and R_2 is Me, Et, *n*-Pr, *i*-Pr, *n*-Bu,*s*-Bu, and *i*-Bu. (b) Schematic representation of Figure 1a showing central phosphorus and its immediate neighboring atoms.

The derivation of an empirical relationship (model) between a chosen property of a set of compounds, a dependent variable, and a group of structural descriptors provides the means to understand the interrelationship between structural descriptors and property. Different models address different sub-structural regions/ attributes in predicting the chosen property. While the simplest among the models will be the best (principle of parsimony) to explain the chosen property, the study of a population of various models provides an opportunity to understand the diagnostic aspects of different sub-structural regions as well as in averaging and extrapolating the predictive aspect beyond the individual models. The Genetic Function Approximation (GFA) [3], Mutation and Selection Uncover Models (MUSEUM) [4] and Combinatorial Protocol in Multiple Linear Regression (CP–MLR) [5] are a few approaches for the evolution of multiple models in quantitative structure–activity relationship (QSAR) and quantitative structure–property relationship (QSPR)/ quantitative structure–response relationship (QSRR) studies. Although several QSPR/QSRR reports are available on the GC responses of different types of compounds [6–11], little or scant information exists on organophosphorus compounds in this direction. Here we present a CP–MLR based QSRR study of the GC–TID response factor of organophosphonate esters [2] reported by Saradhi and co-workers in terms of the physicochemical and electronic descriptors. CP–MLR is a

‘filter’ based variable selection procedure for multiple models evolution [5]. Here, the combination rule suggests the groups of descriptors to be evaluated, and filters set the statistical criteria to decide the significance of so developed QSAR model. Four filters (inter descriptor correlation limits, t -values of regression coefficients of descriptors, adjusted multiple correlation coefficient and cross-validated R^2 value) have been embedded in this procedure. The modulation of filters’ cutoff limits offers an opportunity to study different models.

2 MATERIALS AND METHODS

2.1 Chemical Data

The molar response factors RF, $RF = (\text{GC Peak area})(\text{molecular mass})/(\text{volume of injection})(\text{amount of sample})(\% \text{purity})$ [2], of organophosphonate esters (Figure 1) [2] have been transformed as a relative response factor (RRF), that is, RF of compound i as relative to that of dimethyl methylphosphonate (DMMP):

$$RRF_i = RF_i / RF_{\text{DMMP}} \quad (1)$$

As the physicochemical and electronic environment surrounding central phosphorus and its immediate neighbors influence the degree of formation of the GC–TID responding species of the organophosphonates in the detector, the descriptors of phosphorus, its surrounding atoms and groups, and the bonds between them have been considered as the independent parameters of the study. Accordingly, the bond lengths and electronic charges of the concerning groups and atoms along with thermodynamic descriptors were adopted from Saradhi *et al.* report [2]. These descriptors were computed with HyperChem [12]. The van der Waals volumes (Vw) of alkyl substituents (R_1 and R_2 ; Figure 1a) surrounding phosphorus have been calculated to parameterize their steric features [13,14]. The alkyl branching in the R_1 and R_2 of the compounds have been defined with indicator parameters, I_1 and I_2 respectively, which take a value of one for branched alkyl and zero for the others. The total refractivity (TREF) of the compounds has been computed using the Insight II [15–17] software. For this, using the standard fragments library, structures of the compounds have been generated in the Builder module of Insight II [15]. They have been optimized for geometry, in vacuum, by using the consistent valence force field (cvff) implemented in the software. Total refractivity (TREF) of the so generated structures has been computed by using Ghose and coworkers program [16,17] implemented in the Apex–3D module of the software. With this a total of eighteen parameters are selected as the independent descriptors of the study. Table 1 lists the names of all descriptors along with a unique number for each one of them. All these descriptors can be broadly categorized as (a) global descriptors to describe the given property or measurement made on the whole molecule, (b) local or site specific descriptors to address the property or measurement of a specific site or a portion of the molecule, and (c) indicator variables meant for quantum changes in the defined portions of the molecule.

Table 1. Molecular descriptors of organophosphonates considered for the study (Figure 1)

S.No.	Descriptor Name	Abbreviation	Ref.
<i>(a) Global descriptors</i>			
1	Total refractivity	TREF	
2	Molecular weight (mass)	MW	2
3	Enthalpy of vaporization	ENTH	2
4	Gibbs' free energy	GFE	2
5	Boiling point	BP	2
<i>(b) Local or site specific descriptors</i>			
6	Van der Waals volume of R ₁	VwR ₁	
7	Van der Waals volume of R ₂	VwR ₂	
8	Bond length of P–O _b	LPO	2
9	Bond length of P–X	LPX	2
10	Bond length of O _b –C _d	LOC	2
11	Bond length of P=O _a	LPDO	2
12	Charge on phosphorus	CP	2
13	Charge on O _a of P=O _a	COPDO	2
14	Charge on O _b of P–O _b	COPO	2
15	Charge on C _d of O _b –C _d	CCOC	2
16	Charge on X of P–X	CXPX	2
<i>(c) Indicator variable</i>			
17	For alkyl branching in R ₁	I ₁	
18	For alkyl branching in R ₂	I ₂	

Table 2. The relative response factor (RRF), total refractivity (TREF) and van der Waals volumes (Vw) of organophosphonates (Figure 1a)

Comp No	R ₁	R ₂	TREF	VwR ₁	VwR ₂	RRF
1	Me	Me	26.50	0.245	0.245	1.00
2	Me	Et	36.00	0.245	0.399	0.96
3	Me	Pr	45.05	0.245	0.553	0.91
4	Me	i-Pr	44.85	0.245	0.503	0.90
5	Me	Bu	54.25	0.245	0.707	0.86
6	Me	s-Bu	53.90	0.245	0.657	0.85
7	Me	i-Bu	54.00	0.245	0.657	0.84
8	Et	Me	31.25	0.399	0.245	0.98
9	Et	Et	40.75	0.399	0.399	0.93
10	Et	Pr	49.80	0.399	0.553	0.88
11	Et	i-Pr	49.60	0.399	0.503	0.86
12	Et	Bu	59.00	0.399	0.707	0.82
13	Et	s-Bu	58.65	0.399	0.657	0.80
14	Et	i-Bu	58.75	0.399	0.657	0.81
15	i-Pr	Me	35.70	0.503	0.245	0.95
16	i-Pr	Et	45.15	0.503	0.399	0.89
17	i-Pr	Pr	54.20	0.503	0.553	0.83
18	i-Pr	i-Pr	54.00	0.503	0.503	0.81
19	i-Pr	Bu	63.45	0.503	0.707	0.77
20	i-Pr	s-Bu	63.05	0.503	0.657	0.74
21	i-Pr	i-Bu	63.15	0.503	0.657	0.75
22	H	Me	16.35	0.056	0.245	1.01
23	H	Et	25.85	0.056	0.399	0.96
24	H	Pr	34.90	0.056	0.553	0.92
25	H	i-Pr	34.65	0.056	0.503	0.91
26	H	Bu	44.10	0.056	0.707	0.87
27	H	s-Bu	43.70	0.056	0.657	0.86
28	H	i-Bu	43.85	0.056	0.657	0.87

All the compounds of the study, the values of van der Waals volumes of R_1 and R_2 groups (VwR_1 and VwR_2) and the total refractivity (TREF) are enumerated along with RRF in Table 2. The electronic charges and bond lengths of the concerning groups and atoms are available in the previous report [2] and are not listed here. On this data set CP–MLR has been applied to identify diverse structure GC response factor relationships of these compounds. The details of the computational procedure are presented below.

2.2 Computational Procedure

CP–MLR is a ‘filter’ based variable selection procedure for model development in QSAR and QSPR/ QSRR studies [5]. To extract the optimum and diverse structure–property relationship information from the parameter set considered in MLR, a combinatorial strategy with appropriately placed ‘filters’ is developed to recurrently select the non–repetitive k independent variables, at a time, from a total of p variables for the regression model development. Here, different statistical measures are formed as filters to evaluate the significance of the models. If we call a group of variables as bundle, then according to the combination rule, a total of ${}_pC_k$ bundles emerge from p variables with k variables in each bundle (original variable bundle, OVB). Here ‘bundle’ represents a collection of items, in the present case group of descriptors. If the original descriptors are the items of a bundle then it is called as original variable bundle (OVB). Any transformation operation, other than scaling, on an original descriptor leads to a transformed descriptor. If the items of a bundle are from transformed descriptors then it is called as transformed variable bundle (TVB). Accordingly, a variable may contribute to a model in two different ways: (i) by itself alone and / or (ii) by itself and its functionally transformed term together. To find the influence of a selected function of any variable along with its original form in the model development, the k variables of OVB along with their meaningfully transformed functional variables will be adopted for the formation of new bundles. We may mention it here that the functionally transformed variable enters the bundle only when its original variable is part of that bundle. The OVB provides k variables to create the functionally transformed variable bundles (TVBs). In this process, the contents and number of variables in the TVBs (s) are varied from one to k' to explore the role of functionally transformed variable combinations (${}_kC_s$; $s = 1$ to k' where $k' \leq k$) along with respective OVB. Furthermore, the size of the OVB (k) is also varied from a minimum (begin, b) to maximum (end, e) value with an increment of one. The e value will be governed by the number of observations (n) with n/e ratio as large as possible ($1 \leq b \leq e \leq n/e$). This process generates different sizes of OVBs of all variables within the limits of k and joins TVBs of different sizes, to the respective OVBs to form OVB–TVBs. If no functional transformation is considered in the analysis, all the terms and corresponding to TVBs will become null and void. If the size of OVB (k) is restricted to a single value, the resultant models contain uniformly k variables in each one of them. Four layers of filters embedded in the procedure make this process efficient and offer unique solutions. They are set in

terms of inter-parameter correlation cutoff criteria for variables to stay as a bundle (filter-1), t -values of regression coefficients of variables associated with a bundle (filter-2), square-root of adjusted multiple correlation coefficient of regression equation (r -bar) [18] (filter-3), and cross-validated R^2 (Q^2) criteria (filter-4). The following equations are used for computing r -bar and Q^2 .

$$(r\text{-bar})^2 = (1.0 - (1.0 - r^2)(n-1)/(n-k-1)) \quad (2)$$

$$r^2 = 1.0 - (\sum(\mathbf{Yc} - \mathbf{Yo})^2 / \sum(\mathbf{Yo} - \mathbf{Ym})^2) \quad (3)$$

$$Q^2 = 1.0 - (\sum(\mathbf{Yp} - \mathbf{Yo})^2 / \sum(\mathbf{Yo} - \mathbf{Ym})^2) \quad (4)$$

$$SPRESS = \sqrt{(\sum(\mathbf{Yp} - \mathbf{Yo})^2 / (n - k - 1))} \quad (5)$$

$$SDEP = \sqrt{(\sum(\mathbf{Yp} - \mathbf{Yo})^2 / n)} \quad (6)$$

In the above expressions \mathbf{Yo} , \mathbf{Ym} , \mathbf{Yc} and \mathbf{Yp} are observed, mean, calculated and predicted values, respectively, of dependent variable, n is number of observations and k is number of independent variables in regression equation. $SPRESS$ and $SDEP$ are standard deviation of predictors and standard error of predictors respectively [4].

The filter-1 controls the entry of OVBS with inter-correlated variables in the model development. The default cutoff value for the tolerance of inter-parameter correlation coefficient between pairs of independent variables is set as less than or equal to 0.3 to maintain reasonably good independence among the variables of a bundle. The efficiency of CP-MLR is primarily based on this filter. The second filter (filter-2) evaluates the significance of variables in a bundle in terms of the t -values of regression coefficients. A default value of 2.0 is set for this filter; a bundle will pass this filter if the t -values of its regression coefficients are more than or equal to the set threshold value. Normally, successive additions of variables to multiple regression equation will increase successive multiple correlation coefficient (r) values. In light of this, to compare the internal explanatory power of bundles with different number of variables (variable bundles of different sizes), r -bar is adopted in this procedure [18]. Accordingly, filter-3 sets predefined threshold level for r -bar. Only those variable bundles whose r -bar with the dependent variable is more than or equal to the set threshold level will pass this filter. Also, incremental raise in r^2 (or r) is associated with the best subset regressions, especially when the total number of predictor variables (p) (corresponding to all bundles) is more than the number of observations (data points) [19]. The number of predictor variables could be very large in many a situations. The higher r^2 may be meaningful only when there is increase in the number of relevant descriptors corresponding to the phenomenon under investigation. However, to exclude false or artificial correlations arising out of these situations, it is important to validate the relevance of selected bundles in the model generation. While the first three filters are designed to check the internal consistency of the data, the fourth filter (filter-4) addresses the external consistency in the form of cross-validation of the model with leave-one-out procedure as the default option. Finally, 'goodness of fit' of the model is measured in terms of Q^2 statistics. A Q^2 is considered to be acceptable only when it has a value between zero

(no predictive power) and one (perfect predictive power). A value between zero and one for Q^2 will result in the collection of a good number of models with different degree of predictive power. Only those models whose Q^2 value is in the predefined limits are retained for the further study.

3 RESULTS AND DISCUSSION

Among the global descriptors of these compounds, the molecular weight (MW) and total refractivity (TREF) of the analogues have shown excellent correlation with RRF (r is -0.979 and -0.955 respectively). The remaining three global parameters, namely enthalpy of vaporization (ENTH), Gibbs' free energy (GFE) and boiling point (BP) of the compounds have shown almost the same degree of correlation with the RRF ($r = -0.877$). Also, these three parameters are almost perfectly inter-correlated with one another ($r = 0.998$ to 1.0). This indicates that for all practical purposes the information content of these three parameters is almost the same. Hence, only one descriptor that is ENTH, among these three is considered in the parameter set for the purpose of analysis. Apart from global parameters, local descriptors, VwR_2 and COPO *i.e.*, charge on oxygen (O_b) of P- O_b (Figure 1b), have also shown good correlation with RRF (r is -0.817 and 0.752 , respectively). The other parameters which displayed directly noticeable correlation with RRF ($r \sim 0.5$) are VwR_1 , CCOC *i.e.*, charge on the carbon (C_d) of O_b - C_d , LOC *i.e.*, bond length of O_b - C_d (Figure 1b) and the indicator parameter I_2 . In this background, the CP-MLR protocol has been applied to the sixteen molecular descriptors ($p = 16$) to identify the best possible regression model(s) with complete or near complete explanation to the variation in the RRF of the compounds under study. For this in CP-MLR, filter-1 was assigned with a value of 0.3, and filter-2 with 2.0. As MW and TREF have shown excellent correlation with RRF, for filter-3 a value of 0.95 has been assigned to collect all possible equations with r -bar value of 0.95 and above. The filter-4 is set to collect all models with Q^2 value between 0.3 and 1.0 ($0.3 \leq Q^2 \leq 1.0$). With these filter thresholds a search has been carried out for models having up to four original variables (OVBs) and up to two transformed variables (TVBs) ($k = 1$ to 4 ; $s = 1$ to 2). As squared term of a descriptor is the most often and widely used functional transformation of variables in QSAR and QSPR studies, the same is opted for TVBs of this data set also. This resulted in the identification sixteen structure-response relationship equations (models) for the RRF of organophosphonates. In the discussion a statistically significant regression equation explaining the structure-response relationship is called as a model. A model and the corresponding equation are identified with the same number. A model contains one or more explanatory descriptors. The descriptors involved in each one of these models along with selected statistical indices are listed in Table 3. Furthermore to maintain brevity and facilitate easy comparison, in Table 3, the explanatory descriptors of each model are identified with the serial numbers allotted to them in Table 1; an 's' character after descriptor number indicates that both the descriptor and its squared term are involved in the equation; an asterisk sign before the descriptor

number indicates that the regression coefficients of the descriptor and its squared term (transformed descriptor) have the same mathematical sign (both are positive or both are negative).

Table 3. CP-MLR models for the estimation of RRF of organophosphonates and the associated statistical parameters.[#]

Model No.	Model [¶]	Normal stat. [§]				Cross-validation stat.			
		var	<i>r</i>	<i>r</i> -bar	SE	F	Q ²	SPRESS	SDEP
7	2	1	0.979	0.978	0.015	602.00	0.950	0.016	0.016
8	1	1	0.955	0.953	0.022	270.40	0.895	0.024	0.023
9	1,16	2	0.983	0.982	0.014	364.94	0.958	0.015	0.015
10	2,16	2	0.983	0.982	0.014	362.35	0.957	0.016	0.015
11	1,13	2	0.978	0.976	0.016	275.82	0.946	0.017	0.016
12	6,7	2	0.954	0.951	0.022	127.86	0.890	0.025	0.024
13	7,9s	3	0.981	0.978	0.016	199.84	0.941	0.019	0.017
14	3,17,18	3	0.979	0.977	0.015	188.51	0.942	0.018	0.017
15	7,8,17	3	0.960	0.955	0.022	93.24	0.896	0.025	0.023
16	3,15,17	3	0.957	0.951	0.022	87.08	0.883	0.026	0.024
17	3,9s, 18	4	0.981	0.978	0.016	147.65	0.936	0.020	0.018
18	3,9s, 15	4	0.960	0.953	0.022	67.65	0.881	0.027	0.024
19	*1s, 16	3	0.988	0.986	0.012	318.46	0.966	0.014	0.013
20	*1s, *16s	4	0.990	0.989	0.011	290.90	0.969	0.014	0.012
21	7, *16s, 17	4	0.971	0.966	0.019	96.13	0.918	0.022	0.020
22	3, *16s, 18	4	0.961	0.954	0.022	68.90	0.888	0.026	0.024

[#], Filter-1 as 0.3; filter-2 as 2.0; filter-3 as 0.95; filter-4 as $0.3 \leq Q^2 \leq 1.0$; [¶], the number corresponds to the descriptor serial number given in Table 1. A character 's' after the descriptor number indicates that both the normal and squared terms are involved in the model. An '*' before the descriptor number indicates that the regression coefficients of the descriptor and its squared term have the same mathematical sign; [§], in all the models the number of compounds are 28; *var* is number of descriptor variables in the model; *SE* is standard error of the estimate; *F* is *F*-ratio between the variance of calculated and observed activities.

In all the models the *t*-values of the regression coefficients of the descriptors are significant at more than 95% level. Even though the search is carried out up to four original variables and up to two of their squared terms, we could not find any statistically significant equation beyond four variables (including the squared terms). The identified models indicate that *k* = 1 to 3 and *s* = 1 to 2 as the optimum perimeter for model extraction for the filter thresholds adopted. Of the sixteen descriptors considered in the parameter set, twelve have found relevance in one or more multi-parametric equations in correlating the RRF. Among these twelve descriptors, seven are site-specific descriptors (VwR₁, VwR₂, CXPX, CCOC, COPDO, LPX, LPO) and two are indicator variables (I₁, I₂). The indicator parameters I₁ and I₂ account for the branching of alkyl groups R₁ and R₂ respectively. They take a value of one if the respective alkyl group is branched and zero otherwise. All the descriptors, excepting one, (LPX), have shown their presence in the first order terms in the equations. In Table 3, for models 19 to 22 the regression coefficients of some of the descriptors (identified with an asterisk sign before the descriptor number) and their squared terms have the same mathematical sign (both are positive or both are negative). Since this is not normally practiced and not expected in QSAR/QSPR studies, they are discounted from considering as good models. Table 4 presents the regression equations of the multi-parametric models listed in Table 3.

Table 4. Multi-parametric regression equations (models 9–18) showing correlation of RRF of organophosphonates with various molecular descriptors.

No*	Const	TREF/ MW/ ENTH [#]	Site Specific Descriptors								Indicator Variable	
			Steric		Bond Length			Atomic Charge			I ₁	I ₂
			VwR ₁	VwR ₂	LPO	LPX	LPX ²	COPDO	CCOC	CXPX		
9	1.139	0.006 0.0004 ^s								0.255 0.082		
10	1.262	0.002 0.0002								0.095 0.080		
11	-0.395	0.006 0.0004						1.074 0.437				
12	1.142		0.208 0.051	0.379 0.058								
13	-46.74					61.396 12.588	19.361 3.961					
14	1.752	0.019 0.002									0.051 0.014	0.049 0.012
15	-0.454			0.377 0.056	0.912 0.570						0.067 0.020	
16	2.009	0.019 0.003							0.555 0.266		0.049 0.021	
17	-30.40	0.019 0.004				41.209 13.138	12.986 4.134					0.044 0.012
18	-29.67	0.019 0.003				40.555 18.362	12.780 5.667		0.483 0.271			

*, Except for LPO and LPX, the sign of the regression coefficients of all variables (including LPX²) are negative; [#], in models 9 & 11 the regression coefficients are of TREF, in model 10 the regression coefficient is of MW and in models 14, 16–18 the regression coefficient are of ENTH; ^s, 95% confidence intervals of the regression coefficients.

The ease of interaction of compound–alkali metal (Rb) or compound–electron emitted by alkali metal is the most crucial and determining step in the formation responding species and, in turn, in RRF. The results broadly suggest that the variations in the RRF of the compounds are primarily correlated to each of their global properties, namely Eqs. (7)–(11), (14), (16)–(18), and to the local descriptor VwR₂, namely Eqs. (12), (13), (15), Tables 3 and 4. The multi parametric models formed here by the combination of global, site-specific and/ or indicator parameters suggest the role of various structural components of organophosphonates in the formation of responding species. The discussion of models is organized to address each structural component of organophosphonates namely, P=O_a, P–O_b, R₁ (X), R₂ and O–R₂ (C_d–O_b), (Figure 1) in a stepwise manner and models (equations) are identified with their numbers mentioned in the Tables 3 and 4.

The negative regression coefficient of COPDO in Eq. 11 suggests that having electron rich O_a of P=O_a is favorable for the formation of desired responding species. In other words, a weak or polarized (d–p)π bond between phosphorus and oxygen (O_a) may act like driving force for the generation of responding species in the TID. The polarization of (d–p)π electrons towards oxygen (O_a) makes the situation conducive for the approach of alkali metal towards the central phosphorus, which, in turn, may further facilitate the shift of π electrons to O_a and subsequent transfer of an electron from the alkali metal to phosphorus. Even though the initiation of the reaction may be due to the polar nature of P=O_a, the whole event may take place like a concerted process. The

coefficient of LPDO in the following regression equation also favors polarization of P=O_a bond for the ease of formation of the desired responding species.

$$\text{RRF} = -1.028 - 0.020(\pm 0.004)\text{ENTH} + 1.742(\pm 1.560)\text{LPDO} - 0.043(\pm 0.021)\text{I}_2 \quad (23)$$

$n = 28 \quad r = 0.944 \quad SE = 0.025 \quad F = 65.67$

In addition, the positive regression coefficient of LPO in Eq. (15) suggests that a longer or less conjugated P–O_b bond (sigma bond) is favorable for the formation of responding species. This situation may point towards the more localized (d–p) π bond (high–energy state) between phosphorus and O_a to initiate the process of formation of desired species. Moreover, the charge on oxygen, O_b, of P–O_b (COPO) is positively correlated with RRF ($r = 0.752$) indicating the unfavorable conditions associated with large negative charge on oxygen O_b to initiate the process.

The site–specific descriptors, VwR₁ and VwR₂, are coupled with the negative regression coefficients in Eqs. (12), (13), (15). This indicates that the increased substitution around central phosphorus leads to higher restrictions in the formation of desired species. The Eqs. (9), (10), (13), (17) and (18) address the R₁ group but in terms of characteristics of atom (X; Figure 1b) immediately attached to the phosphorus. The negative regression coefficient of CXPX in Eqs. (9) and (10) suggests that the ease of formation of desired responding species is associated with the electron rich center (X; Figure 1b) in the neighborhood of phosphorus. Eqs. (13), (17) and (18) are quadratic with respect to LPX and suggest 1.586 Å as the optimum separation for P–X bond. While the volume addresses spatial requirements, the bond distances can be viewed as its one–dimensional operator. Here, optimally separated phosphorus and its immediate neighbor (X; Figure 1b) may facilitate the approach of alkali metal or capture of electron and thereby the formation of desired species. From these equations it is evident that both electronic and steric effects operate in the formation of desired species for the identification in TID.

The influence of alkyl branching on RRF may be seen through the Eqs. (14)–(17) where the indicator parameters, I₁ and I₂, have uniformly negative regression coefficients. This suggests that the branched alkyls in R₁ and R₂ may lead to decreased formation of the desired responding species for detection. In addition to imposing pre–discussed steric restrictions, the branched alkyls may function like a ‘curtain’ between the central phosphorus and the approaching alkali metal. It is also important to note that the regression coefficients of I₁ and I₂ are of the same order in the above equations. This suggests that restrictions imposed by them will be approximately in the same order. The following equations further confirm this observation:

$$\text{RRF} = 0.930 - 0.076(\pm 0.047)\text{I}_1 - 0.075(\pm 0.041)\text{I}_2 \quad (24)$$

$n = 28 \quad r = 0.703 \quad SE = 0.053 \quad F = 12.20$

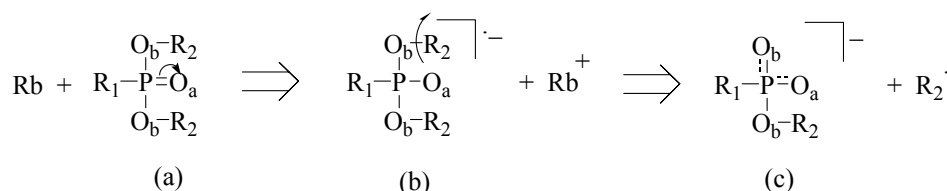
$$\text{RRF} = 0.930 - 0.076(\pm 0.031)(\text{I}_1 + \text{I}_2) \quad (25)$$

$n = 28 \quad r = 0.703 \quad SE = 0.053 \quad F = 25.38$

Amongst all, Eqs. (16) and (18) suggest the necessity of electron rich carbon (C_d) of C_d–O_b

(CCOC) in the formation of desired responding species. This probably operates by making the cleavage of C_d–O_b bond, *i.e.*, O–R₂ bond more facile at a later stage.

Collectively, the above discussion on the RRF of organophosphonates favors the polarization of (d–p)π electrons of P=O_a to O_a with the extraction of an electron from the alkali atom (Rb) as ‘the most probable path’ for the formation of the characteristic responding species in the GC–TID (Scheme 1).



Scheme 1. It is proposed that in GC–TID the organophosphonates (a) extract an electron from the alkali atom and form the radical anions (b) which in a subsequent step may undergo homolytic cleavage of O_b–R₂ to generate characteristic responding species for identification in the detector.

According to the previous report [2], one possibility is that on the approach of an alkali atom, the reaction may proceed by the homolytic fission of O_b–C_d bond of P–O_b–C_d (Figure 1b) to generate phosphonyl radical, which in a subsequent step extracts an electron from another alkali atom to form the negatively charged ion. The other suggested possibility is that the compound captures an electron and simultaneously cleaves the O_b–C_d (Figure 1b) bond in a homolytic manner to generate the desired responding species.

While both these possibilities are centered around the cleavage of O_b–C_d of P–O_b–C_d (Figure 1b) as the principal path for the initiation and formation of responding species [2], the equations obtained here only favor polarization of (d–p)π bond of P=O_a for the initiation of the reaction as discussed above. However, our results support the other observations of previous report [2] such as, the influence of global properties, alkyl chain and alkyl branching etc. in deciding the RF (or RRF) of the compounds under study.

4 CONCLUSIONS

The equations in Table 3 represent the collective information derived here and address the influence of different structural features of organophosphonates in the formation of GC–TID responding species which, in turn, gets reflected in the quantum of RF. Of all the variables included in the study, total refractivity (TREF), Molecular weight (mass) (MW) and thermodynamic properties, *e.g.*, enthalpy of vaporization (ENTH) of the compounds are found to be important in correlating their response factor. For the GC–TID detection of organophosphonates, the results obtained here preferred the polarization of (d–p)π bond of P=O_a for the initiation of the reaction to generate the characteristic responding species.

Also, the study suggests that the influence of R₁ and R₂ alkyls on the RRF is two-fold – one in terms of their steric bulk and the other in terms of the branching effect of R₁ and/ or R₂ groups. It further suggests that the restrictions imposed due to the branching effect are approximately in the same order for both R₁ and R₂. Recent simulation experiments involving a model adsorbent phase and a drug highlighted the importance of liquid chromatographic indices of some compounds in predicting their albumin binding affinity [20]. In similar lines, as all the equations presented in Table 3 are highly significant, we hope that they serve the predictive aspect by way of providing the estimates of GC–TID response factors of new compounds.

Acknowledgment

This is CDRI Communication No. 6422.

5 REFERENCES

- [1] B. Kolb and J. Bischoff, New Design of a Thermionic Nitrogen and Phosphorus Detector for GC. *J. Chromatogr. Sci.* **1974**, 12, 625–629.
- [2] U. V. R. V. Saradhi, M. V. S. Suryanarayana, A. K. Gupta, R. P. Semwal and B. Jayaram, Studies on the Response Mechanism of Thermionic Detection to Organophosphorus Compounds. *J. Chromatography A* **2001**, 911, 63–73.
- [3] D. Rogers and A. J. Hopfinger, Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships, *J. Chem. Inf. Comput. Sci.* **1994**, 34, 854–866.
- [4] H. Kubinyi, Variable selection in QSAR studies. 1. An evolutionary algorithm, *Quant. Struct.–Act. Relat.* **1994**, 13, 285–294.
- [5] Y. S. Prabhakar, A Combinatorial Approach to the Variable Selection in Multiple Linear Regression: Analysis of Selwood *et al* Data set –A Case Study, *QSAR and Combinatorial Science*, **2003**, 22, 583–595.
- [6] D. T. Stanton and P. C. Jurs, Computer–Assisted Prediction of Gas Chromatographic Retention Indices of Pyrazines. *Anal. Chem.* **1989**, 61, 1328–1332.
- [7] C. G. Georgakopoulos, J. C. Kiburis and P. C. Jurs, Prediction of Gas Chromatographic Relative Retention Times of Stimulants and Narcotics. *Anal. Chem.* **1991**, 63, 2021–2024.
- [8] C. G. Georgakopoulos, O. G. Tsika and J. C. Kiburis, Prediction of Gas Chromatographic Relative Retention Times of Anabolic Steroids. *Anal. Chem.* **1991**, 63, 2025–2028.
- [9] T. F. Woloszyn and P. C. Jurs, Quantitative Structure–Relationship Studies of Sulfur Vesicants. *Anal. Chem.* **1992**, 64, 3059–3063.
- [10] A. R. Katritzky, E. S. Ignatchenko, R. A. Barcock, V. S. Lobanov and M. Karelson, Prediction of Gaschromatographic Retention Times and Response Factors Using a General Quantitative Structure–Property Relationship Treatment. *Anal. Chem.* **1994**, 66, 1799–1807.
- [11] T. Ivanciuc and O. Ivanciuc, Quantitative Structure–Retention Relationship Study of Gas Chromatographic Retention Indices for Halogenated Compounds, *Internet Electron. J. Mol. Des.* **2002**, 1, 94–107, <http://www.biochempress.com>. Also see references cited therein.
- [12] HyperChem Version 5.0, Hypercube, Inc., U.S.A.
- [13] I. Moriguchi, Y. Kanada and K. Komatsu, Van der Waals volume and Related Parameters for Hydrophobicity in Structure–Activity Studies. *Chem. Pharm. Bull.* **1976**, 24, 1799–1806.
- [14] S. P. Gupta and Y. S. Prabhakar, Importance of van der Waals Volume in QSAR Studies for Drugs. *J. Sci. Ind. Res.* **1985**, 44, 189–198.
- [15] Insight II, version 2.3.0. Biosym Technologies Inc, 9685 Scranton Road, San Diego, California, U.S.A. **1993**.
- [16] A. K. Ghose and G. M. Crippen, Atomic Physicochemical Parameters for Three Dimensional Structure–Directed Quantitative Structure–Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions, *J. Chem. Inf. Comput. Sci.* **1987**, 27, 21–35.
- [17] V. N. Viswanadhan, A. K. Ghose, G. R. Revankar and R. K. Robins, Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure–Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for An Automated Superimposition of Certain Naturally Occuring Nucleoside Antibiotics, *J. Chem. Inf. Comput. Sci.* **1989**, 29, 163–172.

- [18] P. Armitage and G. Berry, *Statistical Methods in Medical Research 2nd Edition*, Blackwell Scientific Publications, Oxford, **1990**, pp.296–357.
- [19] A. C. Rencher and F. C. Pun, Inflation of R^2 in best subset regressions, *Technometrics*. **1980**, 22, 49–53.
- [20] T. Hanai, R. Miyazaki, E. Kamijima, H. Homma, and T. Kinoshita, Computational Prediction of Drug–Albumin Binding Affinity by Modeling Liquid Chromatographic Interactions, *Internet Electron. J. Mol. Des.* **2003**, 2, 702–711, <http://www.biochempress.com>.

Biographies

Yenamandra S. Prabhakar is assistant director in the Medicinal Chemistry Division, Central Drug Research Institute, Lucknow, India. After obtaining a Ph.D. degree in chemistry (theoretical aspects of drug design) from the Birla Institute of Technology and Science, Pilani, India, Dr. Prabhakar undertook postdoctoral research in ethno biology with Professor P. Pushpangadan at the Regional Research Laboratory, Jammu. Dr. Prabhakar's current research interests include development of newer QSAR software and ways of addressing tetrahedral centers in modeling studies