

Structure–activity relationship analysis of cationic 2-phenylbenzofurans as potent anti-trypanosomal agents: a multivariate statistical approach

Sarvesh Kumar Paliwal · Ankita Narayan Verma · Shailendra Paliwal

Received: 16 October 2010 / Accepted: 14 April 2011 / Published online: 21 May 2011
© Springer-Verlag 2011

Abstract As part of an effort to establish a structure–activity relationship of diamidines against African trypanosomes, a quantitative correlation between molecular structure and anti-trypanosomal activity of 2-phenylbenzofuran derivatives was attained using classical quantitative structure–activity relationship (QSAR) descriptors and 3D similarity indices. A good model was obtained on the basis of classical descriptors; however, the model derived using descriptors based on similarity indices neither complemented the classical descriptors nor were significantly predictive. The best QSAR model with chemical descriptors that showed good correlative and predictive ability with $r = 0.91$, $r^2 = 0.82$, and $r_{cv}^2 = 0.80$ was developed using stepwise multiple linear regression analysis (MLR) and a comparable partial least squares analysis (PLS) model with $r_{cv}^2 = 0.79$ was also obtained. The QSAR models revealed that a substituent steric descriptor (Verloop B_1 parameter) and geometrical (moment of inertia 3 length) and hydrophobic ($\log P$) descriptors of the whole molecule have significant impact on anti-trypanosomal activity of the compounds. The best QSAR models were validated by the leave one out technique. To further confirm the predictive power of the models, an external set of molecules was used which was not part of the training set. The high agreement between experimental and predicted inhibitory values, obtained in the validation procedure, indicates the good quality of the derived QSAR models. In addition to QSAR analysis Lipinski's rule

was also applied to the series under consideration and newly designed molecules in order to check the drugability of the compounds; no violation of this rule was found. Hence 2-phenylbenzofuran has tremendous potential to yield orally active anti-trypanosomal agents.

Keywords QSAR · MLR · PLS · Similarity indices · Human African trypanosomiasis

Introduction

Protozoal infections such as malaria, trypanosomiasis, and leishmaniasis are a major health menace and are coincident with much of the world's most acute poverty in developing countries with estimated worldwide deaths of over one million people per year [1]. According to the World Health Organization (WHO) [2], more than 60 million people are at risk of trypanosomiasis, with an estimated 0.3–0.5 million infected people out of which only a small number are subject to active surveillance or have access to health centers where reliable diagnosis and treatment are available [3].

Human African trypanosomiasis (HAT), also called African sleeping sickness, is a re-emerging infectious vector-borne parasitic disease that is found only in the intertropical regions of Africa because of the ecology of the insect vector. Although there are many species of trypanosomes, only two, belonging to the *brucei* group, are infectious to humans: *Trypanosoma brucei gambiense*, found in West and Central Africa, leads to a chronic form of the disease, whereas *Trypanosoma brucei rhodesiense*, found in East and Central Africa, leads to a more virulent and acute condition. The disease is caused by infection with haemoflagellates of the *Trypanosoma brucei* subspecies,

S. K. Paliwal (✉) · A. N. Verma
Department of Pharmacy, Banasthali University,
Banasthali, Rajasthan, India
e-mail: paliwalsarvesh@yahoo.com; paliwalsarvesh@gmail.com

S. Paliwal
Department of Pharmacy, L.L.R.M. medical college,
Meerut, UP, India

which are introduced into the human bloodstream via the bite of infected tsetse flies [4] of the genus *Glossina*. After an insect bite, parasites replicate at the site of infection, producing a local inflammatory reaction, then spread to the regional lymph nodes. From there, the disseminate throughout the host, eventually becoming established in the central nervous system (CNS). African sleeping sickness is currently classified as a “neglected disease” because there are no effective and safe therapies; the treatment, once the parasite has crossed the blood–brain barrier, relies on the use of two clinical medicines that are melarsoprol and eflornithine [5, 6]. However, these drugs have several side effects like myocardial damage, hypertension, exfoliative dermatitis, and reactive encephalopathy, and in addition, resistance has been reported [7–12]. Therefore, there is clearly a vital need for a new generation of drugs [13].

The synthesis of the aromatic diamidines pentamidine and furamidine (Fig. 1) and the discovery of their broad antiparasitic activity are of interest for the development of anti-trypanosomal compounds [14–18], but the mechanism of action of the drugs is not well understood [19]. It has been postulated that diamidines may exert their biological activity by binding to the minor groove of DNA at the AT-rich site which leads to the inhibition of the DNA-dependent enzyme or direct inhibition of transcription [20–25]. Therefore the DNA minor groove has emerged as an important target site for the development of synthetic therapeutic agents. Although pentamidine has had significant clinical success, its toxicity, lack of oral availability, emergence of pentamidine-resistant organisms, and use of furamidine as an alternative to pentamidine emphasize the essential need to develop additional drugs for the treatment of HAT.

Over the last 15 years there has been a revival of drug research and development regarding neglected parasitic diseases, and a number of drug development projects are currently ongoing. However, discovering new lead

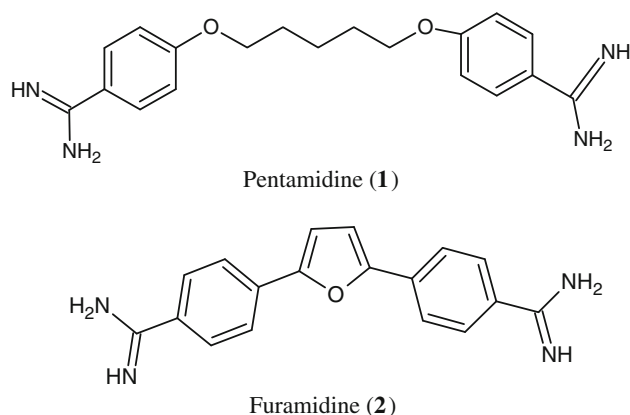


Fig. 1 Structure of key dicationic anti-trypanosomal agents

compounds and optimizing existing lead structures with anti-trypanosomal activity remain a crucial step to sustain the progress achieved to date. The use of quantitative structure–activity relationships (QSAR), since their advent in 1962, has become increasingly helpful in understanding many aspects of chemical–biological interactions in drug and other scientific research [26]. With a properly designed set of congeners, carefully tested in almost any biological system, it has become easy to derive a QSAR by a steadily increasing number of computerized approaches. QSAR models, mathematical equations relating chemical structure to their biological activity, give information that is useful for drug design and medicinal chemistry [27]. An exhaustive literature review revealed that very few attempts have been made to build a QSAR model in the field of HAT, exemplified by the work of Athri et al. [28]; hence we endeavored to develop a robust QSAR model for 2-phenylbenzofuran derivatives and use it in the design of some new orally bioavailable anti-trypanosomal compounds.

The objective of the present multiple QSAR investigations was to develop 3D QSAR models based on similarity indices and 2D QSAR models based on classical descriptors. Firstly, 3D QSAR models derived from similarity data using multiple linear regression analysis (MLR) are discussed. A major benefit of 2D compared to 3D QSAR methods is that the former neither requires conformational search nor structural alignment; moreover, 2D QSAR methods are easily automated and can even be adapted to the task of database searching or virtual screening [29]. This consideration led us to further develop MLR and partial least squares analysis (PLS) models with classical descriptors.

Results and discussion

For all the compounds listed in Table 10, the shape, electronic, refractivity, lipophilicity, and combined similarity matrices were computed. Each matrix was subjected to PCA. Table 1 lists the seven principal components (PCs) having a variation up to 90%. MLR analysis was performed with these seven PCs derived from the similarity matrix. Out of these seven PCs only four PCs (PC1, PC2, PC3, and PC6) were involved in the model development (Eq. 1). Although the r^2 value of Eq. 1 was considerable, the high s value and low f and r_{cv}^2 values suggest a poor predictivity of the developed model.

$$\begin{aligned}
 Y &= -0.219 \times X^1 + 0.490 \times X^2 - 0.283 \times X^3 \\
 &\quad - 0.317 \times X^6 + 1.369 \\
 s \text{ value} &= 0.50, f \text{ value} = 17.40, r = 0.82, \\
 r^2 &= 0.67, r_{cv}^2 = 0.54
 \end{aligned}
 \tag{1}$$

Table 1 Total variance and eigenvalues of PCs from similarity matrix

| Principal components | Total variance explained | Eigenvalue |
|----------------------|--------------------------|------------|
| PC1 | 0.319173 | 70.218 |
| PC2 | 0.530886 | 46.577 |
| PC3 | 0.647551 | 25.6663 |
| PC4 | 0.742101 | 20.8011 |
| PC5 | 0.816368 | 16.3387 |
| PC6 | 0.867206 | 11.1842 |
| PC7 | 0.89623 | 6.38542 |

where X^1 = principal component 1, X^2 = principal component 2, X^3 = principal component 3, and X^6 = principal component 6.

All our efforts to develop a predictive model from PCs of similarity data did not yield a robust QSAR model so we concluded that the data under consideration are unsuitable for generating a QSAR from PCs of similarity matrices. Hence we decided to perform an MLR analysis using the original set of physicochemical parameters (classical descriptors) as independent variables and the inhibitory activity of anti-trypanosomal agents as dependent variables. Initial regression analysis with more than 250 descriptors showed the poor internal predictive ability of the developed model, as revealed by a low r_{cv}^2 value (-0.298), that could be due to a very high number of descriptors. So to find a set of most significant descriptors, data reduction was carried out employing the technique described in the “[Materials and methods](#)” section. The regression model developed using the three most significant descriptors obtained after

data reduction was poor (Eq. 2) as revealed by its low r , r^2 , r_{cv}^2 , s , and f values.

$$Y = 2.330 \times X^1 - 0.519 \times X^2 - 0.466 \times X^3 - 0.294$$

$$s \text{ value} = 0.501, f \text{ value} = 26.374, r = 0.818,$$

$$r^2 = 0.669, r_{CV}^2 = 0.467 \quad (2)$$

where X^1 = Verloop B_1 (substituent 4), X^2 = moment of inertia 3 length (whole molecule), X^3 = log P (whole molecule).

According to general statistical standards, only r^2 values of at least 0.80 can be accepted [30]; the r^2 value for Eq. 2 is 0.669. So to further improve the quality of the model we decided to work in two areas: first, the confirmation of the statistical quality of the descriptors entering the model; second, the identification of outliers.

Out of the three descriptors selected in Eq. 2, two showed high correlation with the biological activity (Table 2), whereas the moment of inertia 3 length shows moderate correlation (0.357) with biological activity; however, the intercorrelation coefficients of all three parameters were very small (<0.16), so the combination of the three descriptors can provide a good fit. The t test values, jackknife standard error (SE), and covariance SE values (Table 3) were significant for all three descriptors which confirms the importance of each selected descriptor.

After confirming the statistical quality of the descriptors we attempted to find the possible outliers. Five outliers (**3**, **13**, **23**, **28**, and **42**) were identified on the regression graph of the recalculated equation because they were plotted far away from the regression line; this disparity means that their observed biological activity was higher than that predicted by the QSAR model, or may point toward experimental or even a typographical error. One by one

Table 2 Correlation matrix of the most relevant descriptors used in the QSAR development

| | Verloop B_1 (sub. 4) X^1 | Moment of Inertia 3 length (whole molecule) X^2 | Log P (whole molecule) X^3 |
|--|---------------------------------|--|-----------------------------------|
| Verloop B_1 (sub. 4) X^1 | 1 | | |
| Moment of inertia 3 length (whole molecule) X^2 | 0.166223 | 1 | |
| Log P (whole molecule) X^3 | -0.052886 | 0.164208 | 1 |
| sub. substituent Log(1/ IC_{50}) | 0.475182 | -0.357948 | -0.733344 |

Table 3 Jackknife SE, covariance SE, and t test values for the selected descriptors

| Descriptors | Jackknife SE ^a | Covariance SE ^b | t value ^c |
|---|---------------------------|----------------------------|------------------------|
| Verloop B_1 (sub. 4) X^1 | 0.368282 | 0.460297 | 4.7883 |
| Moment of inertia 3 length (whole molecule) X^2 | 0.0318809 | 0.141784 | -3.59992 |
| Log P (whole molecule) X^3 | 0.0143173 | 0.0650384 | -7.30993 |

^a An estimate of the SE of each regression coefficient derived from a jackknife procedure on the final regression

^b Gives an estimate of the SE of each regression coefficient derived from the covariance matrix

^c Measures the significance of each variable included in the final model

deletion of the outliers resulted in the formation of the following four equations (3–6):

$$Y = 2.612 \times X^1 - 0.678 \times X^2 - 0.492 \times X^3 - 0.158 \quad (3)$$

$$Y = 2.751 \times X^1 - 0.607 \times X^2 - 0.477 \times X^3 - 0.717 \quad (4)$$

$$Y = 2.422 \times X^1 - 0.635 \times X^2 - 0.491 \times X^3 + 0.079 \quad (5)$$

$$Y = 2.840 \times X^1 - 0.568 \times X^2 - 0.487 \times X^3 - 0.981 \quad (6)$$

where X^1 = Verloop B_1 (sub. 4), X^2 = moment of inertia 3 length (whole molecule), and X^3 = $\log P$ (whole molecule).

The statistics of these equations are listed in Table 4. The statistics of Eq. 6 were better than those of Eqs. 3–5 as revealed by a high r_{cv}^2 value of 0.80 that explains 80% of the variance in biological activity. The lower value of s (0.36) and the high value of f (54.90) indicate a good internal predictive ability of the developed model. The model also exhibited a square correlation coefficient of 0.82 and a high correlation coefficient of 0.91 between descriptors and anti-trypanosomal activity. Because QSAR based on PCs from similarity matrices did not yield promising results, the QSAR model with high statistical significance (Eq. 6) developed using the original classical descriptors was selected for further analysis.

To further confirm the soundness and predictive ability of the model, PLS analysis was performed using the same data set. For a well-defined problem, both MLR and PLS should generate comparable results [31]. The results of the PLS as shown in Eq. 7 were also evaluated on the basis of r_{cv}^2 of the model.

$$Y = 2.667 \times X^1 - 0.483 \times X^2 - 0.514 \times X^3 - 0.852$$

$$r_{cv}^2 = 0.7933, \text{ fraction of variance explained} = 0.825 \quad (7)$$

where X^1 = Verloop B_1 (sub. 4), X^2 = moment of inertia 3 length (whole molecule), and X^3 = $\log P$ (whole molecule).

The external predictive ability of QSAR models (MLR and PLS) was also checked using test sets of compounds that were excluded during the model development. All the compounds in the test set were treated in a manner analogous to the compounds in the training set. The observed

and predicted values of the training and test set of compounds (Tables 5, 6) show that the prediction using the derived QSAR equation was very close to the observed values. Figures 2 and 3 plot the observed activity versus predicted activity for the training and test set compounds.

It is clear from the r , r^2 , r_{cv}^2 , s , and f values that the 2D QSAR analysis (MLR and PLS) performed using the original classical descriptor is statistically better than the QSAR analysis using similarity indices. Thus, these results demonstrate that both MLR and PLS are highly competitive QSAR techniques as applied to this data set.

Equations 6 and 7 indicate that Verloop B_1 , moment of inertia 3 length, and $\log P$ are the main independent factors determining the biological activity. The negative coefficients of the $\log P$ and moment of inertia 3 length show that decreasing the $\log P$ and moment of inertia 3 length values can enhance the anti-trypanosomal activity, whereas the positive coefficient of the Verloop B_1 shows that increasing the Verloop B_1 (sub. 4) values can enhance the anti-trypanosomal activity. The two parameters $\log P$ and moment of inertia 3 length are for the whole molecule, whereas the Verloop B_1 parameter is for the substituent $R^{4'}$. The selected parameter $\log P$, which plays a fundamental role in biochemical processes and influences the fate of a molecule in the binding environment, exhibits the largest negative correlation (-0.733) with biological activity. Thus, a substituent which decreases the hydrophobicity leads to a better fit of the molecule. Probably, the compounds under consideration penetrate deeply into the groove and fit snugly between the walls of the groove. Their amidines form hydrogen bonds with a carbonyl functionality present on the thymine and/or imine group of adenine on the edges of the bases on the floor of the groove. Hence, the hydrophilic substituents in the molecule influence the interaction between the aromatic diamidine and the minor groove of the double helix DNA and selective inhibition of kinetoplast DNA synthesis.

The Verloop parameters [32–34] are a set of multidimensional steric descriptors defining a box that can be used to characterize the shape and volume of the substituent. These parameters are very important in explaining the steric influence of substituents on the interaction between organic compounds and macromolecular drug receptors. The Verloop B_1 parameters describe the width of the substituent in the direction perpendicular to the length of the substituent. In the present study the Verloop B_1 of substituent $R^{4'}$ exhibits the second highest positive correlation (0.475) with biological activity which means that a larger volume of the substituent at $R^{4'}$ is important for a favorable interaction with the receptor. For a chemical compound to interact with an enzyme or a receptor it has to approach and then bind to a binding site. The bulk, size, and shape of the compound have an influence on this

Table 4 Details of equations obtained after deletion of outliers

| Outliers | Equation | r | r^2 | r_{cv}^2 | s value | f value |
|-----------------------|----------|------|-------|------------|-----------|-----------|
| 3 and 13 | (3) | 0.86 | 0.75 | 0.60 | 0.43 | 37.055 |
| 3, 13, and 23 | (4) | 0.88 | 0.77 | 0.73 | 0.41 | 42.26 |
| 3, 13, 23, and 42 | (5) | 0.88 | 0.79 | 0.75 | 0.39 | 43.99 |
| 3, 13, 23, 42, and 28 | (6) | 0.91 | 0.82 | 0.80 | 0.36 | 54.90 |

Table 5 Actual and predicted activity of training set compounds

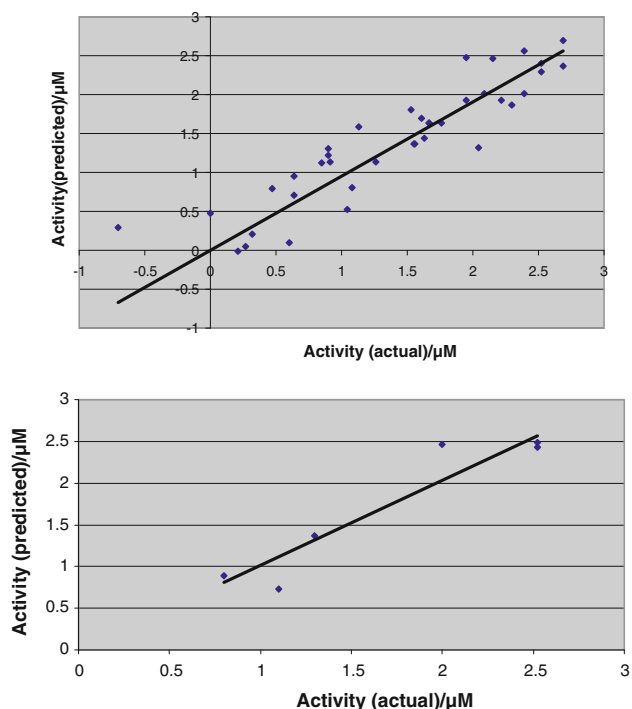
| Compounds | Actual activity (μM) | Predicted activity (μM) | | Residuals values | |
|-----------|-----------------------------------|--------------------------------------|--------|------------------|-------|
| | | MLR | PLS | MLR | PLS |
| 2 | 0.9 | 1.310 | 1.22 | -0.41 | -0.32 |
| 4 | 0.6 | 0.093 | 0.009 | 0.50 | 0.59 |
| 5 | 0.21 | -0.015 | -0.029 | 0.22 | 0.23 |
| 6 | 0.91 | 1.135 | 1.06 | -0.22 | -0.15 |
| 7 | 0.64 | 0.946 | 0.87 | -0.30 | -0.23 |
| 8 | 1.61 | 1.693 | 1.71 | -0.08 | -0.10 |
| 9 | 0.32 | 0.210 | 0.23 | 0.10 | 0.087 |
| 11 | 2.15 | 2.469 | 2.43 | -0.31 | -0.28 |
| 12 | 0.85 | 1.119 | 1.06 | -0.26 | -0.21 |
| 14 | 1.67 | 1.629 | 1.66 | 0.04 | 0.006 |
| 15 | 0.27 | 0.049 | 0.08 | 0.22 | 0.181 |
| 16 | 0.47 | 0.795 | 0.84 | -0.32 | -0.37 |
| 17 | 2.22 | 1.925 | 2.00 | 0.29 | 0.21 |
| 18 | 1.04 | 0.522 | 0.58 | 0.51 | 0.45 |
| 19 | 0.9 | 1.223 | 1.30 | -0.32 | -0.40 |
| 21 | 1.55 | 1.36 | 1.29 | 0.18 | 0.25 |
| 22 | 2.39 | 2.01 | 1.98 | 0.37 | 0.40 |
| 24 | 1.63 | 1.44 | 1.39 | 0.18 | 0.23 |
| 26 | 2.52 | 2.29 | 2.32 | 0.22 | 0.19 |
| 27 | 1.13 | 1.58 | 1.55 | -0.45 | -0.42 |
| 29 | 2.69 | 2.37 | 2.41 | 0.31 | 0.27 |
| 31 | 1.56 | 1.37 | 1.44 | 0.18 | 0.11 |
| 32 | 2.69 | 2.69 | 2.65 | -0.006 | 0.03 |
| 33 | 1.67 | 1.63 | 1.56 | 0.03 | 0.10 |
| 34 | 2.09 | 2.01 | 1.97 | 0.07 | 0.11 |
| 35 | 2.3 | 1.87 | 1.90 | 0.42 | 0.39 |
| 36 | -0.002 | 0.47 | 0.48 | -0.47 | -0.48 |
| 37 | 0.64 | 0.71 | 0.80 | -0.07 | -0.16 |
| 38 | 2.52 | 2.40 | 2.42 | 0.11 | 0.098 |
| 39 | 1.26 | 1.14 | 1.10 | 0.11 | 0.15 |
| 41 | 1.95 | 2.48 | 2.51 | -0.53 | -0.56 |
| 42 | 1.76 | 1.64 | 1.66 | 0.11 | 0.09 |
| 44 | 2.39 | 2.56 | 2.54 | -0.17 | -0.15 |
| 45 | 2.04 | 1.32 | 1.27 | 0.71 | 0.76 |
| 46 | 1.95 | 1.93 | 1.91 | 0.010 | 0.039 |
| 47 | 1.53 | 1.80 | 1.85 | -0.27 | -0.32 |
| 48 | -0.7 | 0.29 | 0.33 | -0.99 | -1.03 |
| 49 | 1.08 | 0.81 | 0.88 | 0.26 | 0.19 |

process because a bulky substituent either acts like a shield and hinders the ideal interaction between the chemical and receptor, or it may help to orientate a drug properly for maximum receptor binding and increase activity. Increasing the shape and volume of the substituent at the R^{4'} position orients the molecule in such a way that there is maximum binding and hence increases the anti-trypanosomal activity of the phenylbenzofuran.

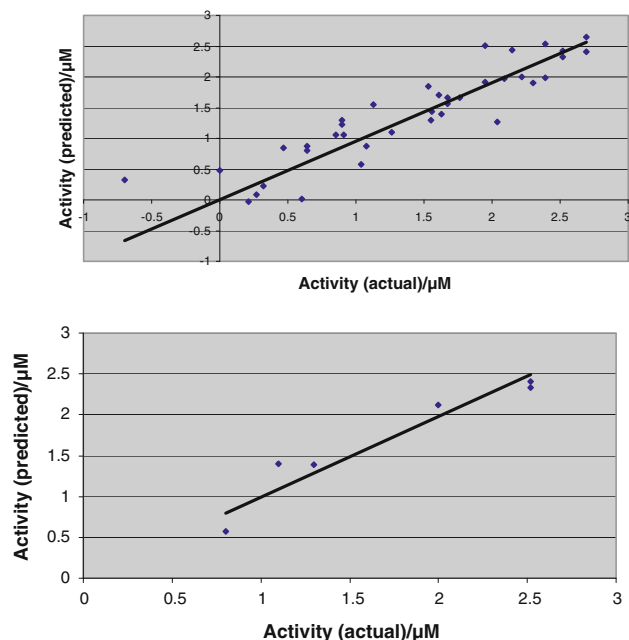
The negative contribution of the moment of inertia 3 length, which is a more advanced structural molecular descriptor derived from the three-dimensional coordinates of the atomic nuclei and atomic masses and/or atomic radii in the molecule, to the biological activity shows that an overall substitution pattern that decreases the shape and mass distribution of the whole molecule leads to an increase in anti-trypanosomal activity.

Table 6 Actual and predicted activity of test set compounds (MLR and PLS)

| Compounds | Actual activity (μM) | Predicted activity (μM) | |
|-----------|-----------------------------------|--------------------------------------|------|
| | | MLR | PLS |
| 1 | 2.52 | 2.49 | 2.41 |
| 10 | 1.1 | 0.73 | 1.40 |
| 20 | 2.52 | 2.43 | 2.33 |
| 25 | 2 | 2.46 | 2.12 |
| 30 | 0.8 | 0.89 | 0.57 |
| 40 | 1.3 | 1.37 | 1.39 |

**Fig. 2** Plot of actual versus predicted activity of training set and test set compounds derived from MLR analysis

During analysis of the results we noticed that there is a mismatch between the steric and geometrical descriptors, because a positive contribution of Verloop B_1 accounts for the presence of bulkier groups at the $R^{4'}$ position, whereas the negative contribution of a geometrical descriptor (moment of inertia 3 length) accounts for decreasing the bulk of the whole molecule. These findings clearly suggest that while designing new molecules care must be taken so that only those functional groups that have optimal width limits can be substituted at $R^{4'}$ position to facilitate the interaction with the receptors. However, these functional groups should not exert any steric effect on the orientation of the whole molecule that may cause a negative influence on the proper orientation of the molecule towards the

**Fig. 3** Plot of actual versus predicted activity of training and test set compounds derived from PLS analysis

receptor site. Therefore, bulk and shape of the compounds should be increased to an optimal level.

This study clearly indicates that optimum bulkier groups and hydrophilic substituents increase the anti-trypanosomal activity of the benzofuran. This fact is clear from diamidine compounds **26**, **29**, **32**, and **38** of the series, which had low values of $\log P$ and moment of inertia 3 length and high values of the Verloop B_1 parameter (Table 7) and exhibit anti-trypanosomal activity equal or superior to that of pentamidine ($IC_{50} = 3 \mu\text{M}$) owing to the presence of an optimum bulky group at the $R^{4'}$ position as compared to other compounds in the series, whereas the presence of di(*N*-isopropyl)amidines and diimidazolines, which are bulkier than diamidines, leads to activity less than pentamidine. Diamidines **29** and **32**, bearing methoxy and hydroxy groups in the $2'$ -position of the phenyl ring, were the most potent compounds in the series, with anti-trypanosomal IC_{50} values of $2 \mu\text{M}$. Not only the size of the groups, but also their position is important for activity. The most active compounds of the series **26**, **29**, and **30** have an amidine group at R^1 ; changing the position of amidine from R^1 to R^2 and increasing the bulk by replacing the amidine with (*N*-isopropyl)amidines drastically reduces the activity as illustrated by compound **48**. The most favorable position, which gives maximum activity, is the presence of an amidine group at 5.

After analysis of these results we are in a position to propose that steric, geometrical, and hydrophobic factors are the important determinants for anti-trypanosomal activity of the drugs binding to the AT sequence of the

Table 7 Calculated values of various descriptors included in the model development

| Compounds | Verloop B_1 | Moment of inertia 3 length | Log P |
|-----------|---------------|----------------------------|---------|
| 2 | 2.12 | 3.05 | 4.12 |
| 4 | 1.99 | 3.28 | 5.58 |
| 5 | 1.65 | 2.97 | 4.17 |
| 6 | 2.10 | 3.24 | 4.16 |
| 7 | 1.99 | 3.02 | 4.16 |
| 8 | 1.87 | 3.13 | 1.79 |
| 9 | 1.90 | 3.88 | 4.12 |
| 11 | 2.11 | 2.94 | 1.79 |
| 12 | 2.12 | 3.37 | 4.12 |
| 14 | 1.87 | 3.22 | 1.79 |
| 15 | 1.87 | 4.03 | 4.12 |
| 16 | 1.81 | 3.57 | 2.75 |
| 17 | 2.11 | 4.13 | 1.54 |
| 18 | 2.13 | 4.68 | 3.87 |
| 19 | 2.10 | 4.50 | 2.50 |
| 21 | 2.13 | 3.27 | 3.84 |
| 22 | 2.11 | 3.16 | 2.46 |
| 24 | 2.14 | 3.40 | 3.55 |
| 26 | 2.15 | 3.70 | 1.54 |
| 27 | 2.41 | 4.23 | 3.87 |
| 29 | 2.22 | 3.87 | 1.54 |
| 31 | 2.14 | 4.42 | 2.50 |
| 32 | 2.10 | 2.75 | 1.51 |
| 33 | 2.24 | 3.33 | 3.84 |
| 34 | 2.08 | 3.02 | 2.46 |
| 35 | 1.88 | 3.09 | 1.51 |
| 36 | 1.91 | 3.71 | 3.84 |
| 37 | 1.78 | 3.81 | 2.46 |
| 38 | 2.15 | 3.46 | 1.54 |
| 39 | 2.13 | 3.61 | 3.87 |
| 41 | 2.24 | 3.77 | 1.54 |
| 43 | 2.12 | 3.86 | 2.50 |
| 44 | 2.10 | 2.99 | 1.51 |
| 45 | 2.15 | 3.43 | 3.84 |
| 46 | 2.07 | 3.11 | 2.46 |
| 47 | 1.92 | 3.40 | 1.51 |
| 48 | 1.91 | 4.02 | 3.84 |
| 49 | 1.80 | 3.73 | 2.46 |

DNA minor groove. The anti-trypanosomal activity of 2-phenylbenzofuran derivatives increases upon addition of an optimal bulkier group, which helps the molecule to orient in a manner to match the width of the DNA minor groove (minor groove widths in A:T regions have been reported in the range of 3–4 Å [35]). This complementary structure enhances the surface contacts and aligns the groups for favorable interactions with DNA. The QSAR

model reveals that the presence of lipophilic groups on the molecules prevents the formation of hydrogen bonds with the AT base pair on the floor of the DNA minor groove, resulting in reduced anti-trypanosomal activity. Hence, the presence of a hydrophilic group facilitates the anti-trypanosomal activity. So on the basis of the present study new compounds were designed with higher potency than the existing benzofurans.

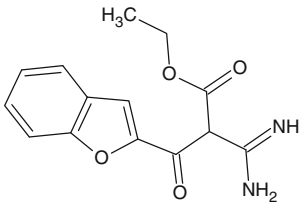
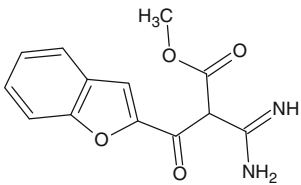
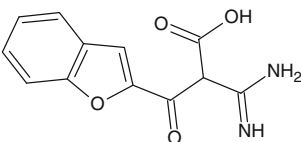
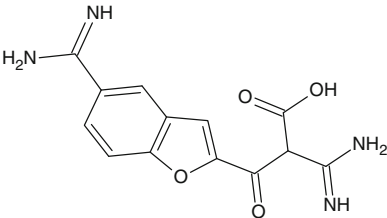
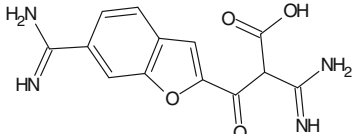
Design of new compounds with high predicted anti-trypanosomal activity

On the basis of the above discussion about the three parameters contributing to anti-trypanosomal activity in Eq. 6, we found that three parameters are very important and exhibit greater correlation with $-\log IC_{50}$; moreover, one of them relates to substituent $R^{4'}$ whereas the other two relate to the structure of the whole molecule. Therefore, by modifying the $R^{4'}$ substituents to improve the values of the three parameters, we have theoretically designed five new compounds **50–54** with higher anti-trypanosomal activities (in the range 3.16–3.80) than the given 2-phenylbenzofuran derivatives. The structures of the five new compounds and the values of the three parameters calculated using the same methods are given in Table 8, along with the $-\log IC_{50}$ values predicted using Eq. 6. It is clear from the table that the $R^{4'}$ structure is modified to reduce its moment of inertia 3 length and log P value and to increase the Verloop B_1 value. It is apparent from the structures of the new compounds that replacement of the substituted phenyl ring at the 2-position of the benzofuran nucleus (Fig. 4) with other bulkier groups enhances the activity as does the presence of an amidine group (cf. **53** and **54** versus **50–52**); therefore the presence of the amidine group is necessary for anti-trypanosomal activity. The values of the three parameters suggest that all the five compounds were successfully designed according to the results obtained from the present QSAR study. Such results further indicate that our model established using the QSAR studies is significant and predictive, and that the consideration of the molecular design is also reasonable.

Lipinski rule of five

Lipinski's "rule of five" is a heuristic approach for predicting drug-likeness stating that molecules with a molecular weight greater than 500, log P greater than 5, more than 5 hydrogen bond donors, and more than 10 hydrogen bond acceptors have poor absorption or permeation [36]. This rule describes only the molecular properties related to the pharmacokinetics of molecules, i.e., the absorption, distribution, metabolism, and excretion (ADME) of bioactive compounds in a higher organism.

Table 8 Structure, predicted biological activity, and calculated values of various descriptors of designed compounds

| Structures of the designed molecules | Verloop B_1 (sub. 4) | Moment of inertia 3 length (whole molecule) | Log P (whole molecule) | Biological activity (μM) |
|--|------------------------|--|-----------------------------|--|
|  <p style="text-align: center;">50</p> | 2.273 | 3.477 | 0.575 | 3.16 |
|  <p style="text-align: center;">51</p> | 2.200 | 3.360 | 0.232 | 3.20 |
|  <p style="text-align: center;">52</p> | 2.127 | 2.954 | 0.201 | 3.25 |
|  <p style="text-align: center;">53</p> | 2.274 | 3.271 | -0.622 | 3.86 |
|  <p style="text-align: center;">54</p> | 2.126 | 3.187 | -0.622 | 3.52 |

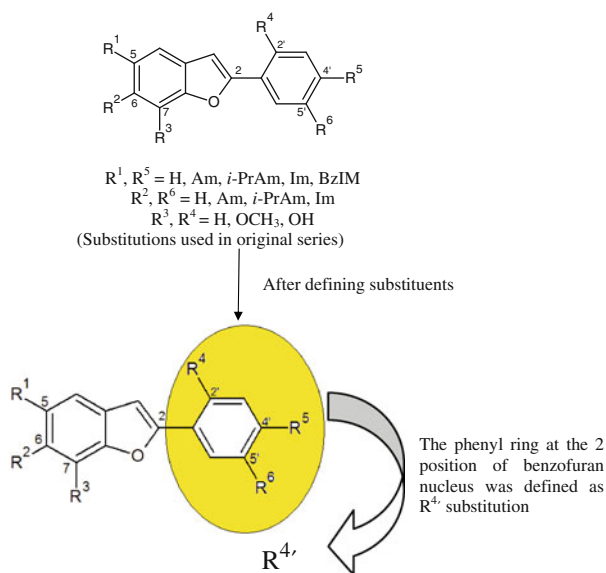


Fig. 4 Lead structure and the substitution pattern of 2-phenylbenzofuran used in the original series and in the QSAR study

There is no consideration of the pharmacodynamic aspects of the molecules, which deal with drug action on the body or on microorganisms and other parasites within or on the body. The parameters included in Lipinski's rule of five were calculated for the series under consideration and the designed compounds and are summarized in Table 9. The results clearly indicate that there is no violation of Lipinski's rule and it is highly likely that all the designed compounds will have favorable pharmacokinetics profiles.

Conclusion

The DNA minor groove is an important target site for enzymes and transcription control proteins, and it has been a particularly attractive target site for the development of synthetic agents for therapeutic purposes and for sequence-selective recognition of DNA. All the results discussed indicate that PCA based on similarity matrices and classical descriptors does not give significant results whereas by using original classical steric, geometrical, and hydrophobic descriptors a very robust model has been derived that also possesses very powerful predictive ability. The model was validated by standard statistical means to check how it reproduces and explains the differences in the experimentally known activity data. Detailed structural investigation revealed that the anti-trypanosomal activity is predominantly explained by the substituent size, shape, and hydrophobicity of the whole molecule and provided insights into how modulation of the steric bulkiness of the substituents and hydrophobicity of whole molecule could

be useful to optimize the anti-trypanosomal activity and hence improve the observed biological activity. Thus, the model reported in the present study will be helpful in the development of new compounds with improved efficacy and oral bioavailability.

Materials and methods

Biological activity data and 3D structure generation

The structures and anti-trypanosomal activities of 49 2-phenylbenzofuran derivatives described previously [37] (Table 10) were used for the present study. The trypanosomal inhibitory activity of the compounds in the series were expressed as $\log(1/IC_{50})$ values to obtain linear relationships in equations, where IC_{50} refers to experimentally determined micromolar (μM) concentration of the compounds required to inhibit 50% of the trypanosomal activity. The structures of all the molecules were drawn using the TSARTM 3D version 3.3 software and were converted into high quality 3D structures employing CORNIA which is a rule-based method developed by Gasteiger, Rudolph, and Sadowski [38] for the conversion of 2D molecules into their 3D structure. The created 3D models were cleaned up and subjected to energy minimization using the COSMIC module of TSAR. The COSMIC force field calculates molecular energies by summing the bond length, bond angle, torsion angle, van der Waals, and coulombic terms for all appropriate sets of atoms. The lowest energy structure of each molecule was used to calculate the molecular descriptors. Charges were calculated using the derive charge -2 option, which is an empirical method. It calculates charges considering the inductive effect in the saturated molecule via the atomic electronegativity, polarizability, and Hückel molecular orbital calculations for π systems through the appropriate coulomb and resonance integrals.

Data set preparation

In the original series the 2-phenylbenzofuran moiety was taken as a template with six substituents $R^1, R^2, R^3, R^4, R^5,$ and R^6 each attached to this template by a single bond (Fig. 4). Certainly there are many ways to represent a molecule as a template with many substituents. So after some experimenting we chose a substituent-defining scheme in which the number of substituents was four as depicted in Fig. 4. The substituents were defined using the define substituent option in the TSAR worksheet's toolbar. All the molecules had the same number of substituents and all the substituents were numbered $R^1, R^2, R^3,$ and $R^{4'}$ (Table 10), where the substituted phenyl ring at the 2

Table 9 ADME properties of the 2-phenylbenzofuran derivatives and designed compounds

| Compounds | ADME weight | ADME hydrogen bond acceptor | ADME hydrogen bond donor | ADME log <i>P</i> |
|-----------|-------------|-----------------------------|--------------------------|-------------------|
| 2 | 362.52 | 3 | 2 | 4.1268 |
| 4 | 426.5 | 3 | 2 | 5.58 |
| 5 | 326.38 | 3 | 2 | 4.1788 |
| 6 | 378.46 | 3 | 2 | 4.1665 |
| 7 | 378.46 | 3 | 2 | 4.1665 |
| 8 | 278.34 | 3 | 2 | 1.799 |
| 9 | 362.52 | 3 | 2 | 4.1268 |
| 11 | 278.34 | 3 | 2 | 1.799 |
| 12 | 362.52 | 3 | 2 | 4.1268 |
| 14 | 278.34 | 3 | 2 | 1.799 |
| 15 | 362.52 | 3 | 2 | 4.1268 |
| 16 | 330.42 | 3 | 2 | 2.753 |
| 17 | 308.37 | 4 | 2 | 1.5463 |
| 18 | 392.55 | 4 | 2 | 3.8741 |
| 19 | 360.45 | 4 | 2 | 2.5003 |
| 21 | 378.52 | 4 | 3 | 3.8424 |
| 22 | 346.42 | 4 | 3 | 2.4686 |
| 24 | 394.52 | 5 | 4 | 3.558 |
| 26 | 308.37 | 4 | 2 | 1.5463 |
| 27 | 392.55 | 4 | 2 | 3.8741 |
| 29 | 308.37 | 4 | 2 | 1.5463 |
| 31 | 360.45 | 4 | 2 | 2.5003 |
| 32 | 294.34 | 4 | 3 | 1.5146 |
| 33 | 378.52 | 4 | 3 | 3.8424 |
| 34 | 346.42 | 4 | 3 | 2.4686 |
| 35 | 294.34 | 4 | 3 | 1.5146 |
| 36 | 378.52 | 4 | 3 | 3.8424 |
| 37 | 346.42 | 4 | 3 | 2.4686 |
| 38 | 308.37 | 4 | 2 | 1.5463 |
| 39 | 392.55 | 4 | 2 | 3.8741 |
| 41 | 308.37 | 4 | 2 | 1.5463 |
| 43 | 360.45 | 4 | 2 | 2.5003 |
| 44 | 294.34 | 4 | 3 | 1.5146 |
| 45 | 378.52 | 4 | 3 | 3.8424 |
| 46 | 346.42 | 4 | 3 | 2.4686 |
| 47 | 294.34 | 4 | 3 | 1.5146 |
| 48 | 378.52 | 4 | 3 | 3.8424 |
| 49 | 346.42 | 4 | 3 | 2.4686 |
| 50 | 274.3 | 5 | 1 | 0.5753 |
| 51 | 260.27 | 5 | 1 | 0.2328 |
| 52 | 246.24 | 5 | 2 | 0.2011 |
| 53 | 288.29 | 6 | 3 | -0.6223 |
| 54 | 288.29 | 6 | 3 | -0.6223 |

position of benzofuran (Fig. 4) was taken as the 4th substitution. The molecules of the series were divided randomly [39, 40] into training set and test set. The training set was used to build linear models so that an accurate

relationship could be found between structure and biological activity. The test set of six molecules was not used to develop the regression model but served to check the predictive power of the developed model.

Table 10 Structures, $\log(1/IC_{50})$ values, and various substitutions defined by TSAR for MLR and PLS analysis

| No. | R ¹ | R ² | R ³ | R ^{4'} | Log 1/IC ₅₀ |
|-----|----------------|----------------|----------------|-----------------|---------------------------|
| 1 | | H | H | | 2.52 |
| 2 | | H | H | | 0.90 |
| 3 | | H | H | | 0.72 |
| 4 | | H | H | | 0.60 |
| 5 | | H | H | | 0.21 |
| 6 | | H | H | | 0.91 |
| 7 | | H | H | | 0.64 |
| 8 | | H | H | | 1.61 |

Descriptor generation

The numerical descriptors are responsible for encoding important features of the structure of the molecules. Both classical and similarity based descriptors were calculated for each compound in the training set, using the TSAR software.

Similarity indices

Similarity indices represent a quantitative measure of the similarity between two molecules on the basis of their size, shape, electronic distribution, lipid solubility, water solubility, or chemical reactivity [41]. Molecular similarity was introduced as a concept by Carbo et al. [42]. The use of

Table 10 continued

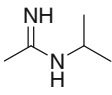
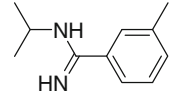

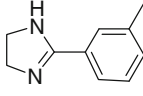

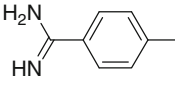
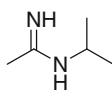
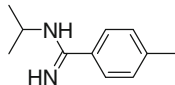

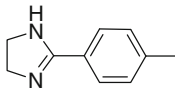

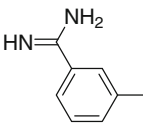
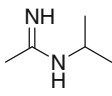
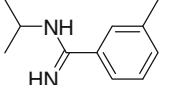

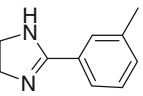

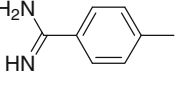
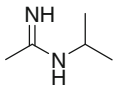
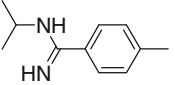
| No. | R ¹ | R ² | R ³ | R ⁴ | Log 1/IC ₅₀ |
|-----|---|---|------------------|---|---------------------------|
| 9 |  | H | H |  | 0.32 |
| 10 |  | H | H |  | 1.10 |
| 11 | H |  | H |  | 2.15 |
| 12 | H |  | H |  | 0.82 |
| 13 | H |  | H |  | 0.49 |
| 14 | H |  | H |  | 1.67 |
| 15 | H |  | H |  | 0.27 |
| 16 | H |  | H |  | 0.47 |
| 17 |  | H | OCH ₃ |  | 2.22 |
| 18 |  | H | OCH ₃ |  | 1.04 |

Table 10 continued


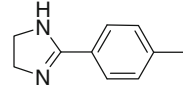

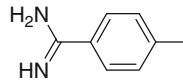
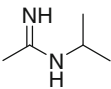
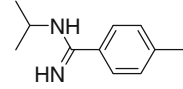

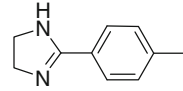

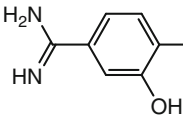
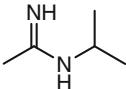
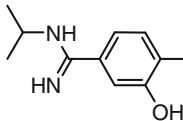

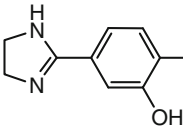

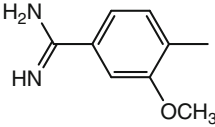
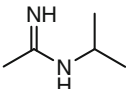
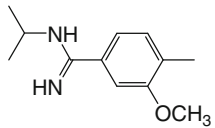
| No. | R ¹ | R ² | R ³ | R ⁴ | Log 1/IC ₅₀ |
|-----|---|----------------|------------------|---|---------------------------|
| 19 |  | H | OCH ₃ |  | 0.90 |
| 20 |  | H | OH |  | 2.52 |
| 21 |  | H | OH |  | 1.55 |
| 22 |  | H | OH |  | 2.39 |
| 23 |  | H | OH |  | 1.74 |
| 24 |  | H | H |  | 1.63 |
| 25 |  | H | OH |  | 2.00 |
| 26 |  | H | H |  | 2.52 |
| 27 |  | H | H |  | 1.13 |

Table 10 continued


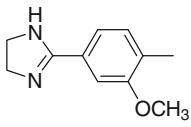

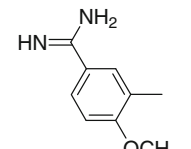

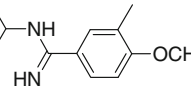

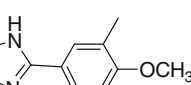

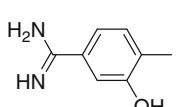

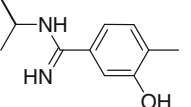

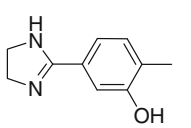

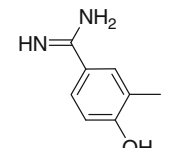

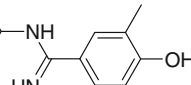
| No. | R ¹ | R ² | R ³ | R ⁴ | Log 1/IC ₅₀ |
|-----|---|----------------|----------------|---|------------------------|
| 28 |  | H | H |  | 0.60 |
| 29 |  | H | H |  | 2.69 |
| 30 |  | H | H |  | 0.80 |
| 31 |  | H | H |  | 1.56 |
| 32 |  | H | H |  | 2.69 |
| 33 |  | H | H |  | 1.67 |
| 34 |  | H | H |  | 2.09 |
| 35 |  | H | H |  | 2.30 |
| 36 |  | H | H |  | -0.002 |

Table 10 continued


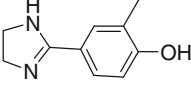

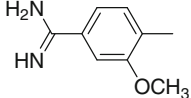
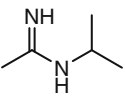
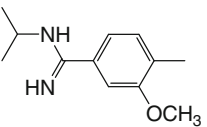

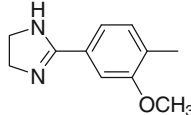

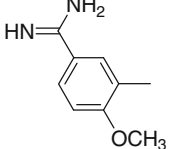
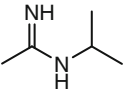
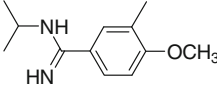

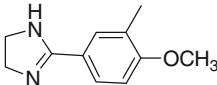

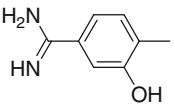
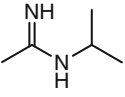
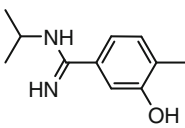

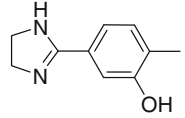

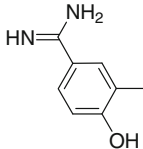
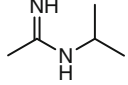
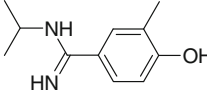

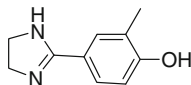
| No. | R ¹ | R ² | R ³ | R ⁴ | Log 1/IC ₅₀ |
|-----|---|---|----------------|---|---------------------------|
| 37 |  | H | H |  | 0.64 |
| 38 | H |  | H |  | 2.52 |
| 39 | H |  | H |  | 1.26 |
| 40 | H |  | H |  | 1.30 |
| 41 | H |  | H |  | 1.95 |
| 42 | H |  | H |  | 0.031 |
| 43 | H |  | H |  | 1.76 |
| 44 | H |  | H |  | 2.39 |
| 45 | H |  | H |  | 2.04 |
| 46 | H |  | H |  | 1.95 |

Table 10 continued

| No. | R ¹ | R ² | R ³ | R ⁴ | Log 1/IC ₅₀ |
|-----|----------------|---|----------------|---|---------------------------|
| 47 | H |  | H |  | 1.53 |
| 48 | H |  | H |  | -0.70 |
| 49 | H |  | H |  | 1.08 |

similarity as a 3D QSAR tool was introduced by Good et al. [43] and has been used by several other groups [44, 45]. The molecular similarity indices were computed with the ASP similarity program in the TSAR software. Among the two approaches (a grid-based method and Gaussian approximation), the Gaussian approximation was used for calculating similarity indices because it closely mirrors that of the grid-based calculations but is much faster. A Gaussian approximation based $N \times N$ similarity matrix was constructed and subjected to data reduction techniques [46].

Classical descriptors

Initially, 267 classical descriptors belonging to structural, geometrical, electronic, and hydrophobic classes were generated for both whole molecules and substituents of compounds in the training set. The values of descriptors for every compound were checked to ensure that the value of each descriptor was calculated for each structure and that there is sufficient variation in these values [47]. The descriptors for which values were not calculated or had constant value for every structure in the data were discarded. In order to study the data patterns and to reduce the data redundancy a data reduction was performed.

Variable selection (data reduction)

In order to select the suitable classical and similarity based descriptors for MLR and PLS analysis, Pearson's correlation matrix and principal component analysis (PCA) were performed on the larger number of descriptor pool.

Input variable selection using PCA

PCA [48, 49] is a multivariate data compression method, which groups the correlated variables and replaces the original descriptors by a new set called principal components (PCs), onto which the data is projected. These PCs contain most of the variability in the data set, are completely uncorrelated, and are built as a simple linear combination of original variables. PCA was performed on both the similarity matrix and classical descriptors and the data were subjected to multivariate statistical techniques.

Input variable selection by generating correlation matrix for classical descriptors

The classical descriptors were selected by correlating each descriptor with one another using correlation matrix. The correlation terms involved in the correlation matrix indicate the extent of co-linearity. A term close to 1 indicates high co-linearity, whereas a value below 0.5 indicates that no co-linearity exists between the two parameters [50]. Among the highly intercorrelated parameters the one that showed low correlation with the biological activity (IC_{50} value) was discarded whereas the other was kept. This process was repeated for each and every set of two consecutive parameters and finally descriptors that were highly correlated with the biological activity but did not have any correlation among each other were retained. Six descriptors, Verloop B_1 (sub. 4), moment of inertia 3 length (whole molecule), dipole moment Y components (sub. 3), dipole moment Z component (sub. 3), log P (whole molecule), and the sum of E-state values (whole molecule)

were retrieved that had good correlation with the biological activity and had minimum intercorrelation among themselves. For choosing the most significant descriptors (out of six) contributing to the anti-trypanosomal activity of 2-phenylbenzofuran derivatives a deletion process was performed in which each variable in the model was omitted in turn and the models were generated using the remaining parameters. Hence after deleting dipole moment Y component (sub. 3), dipole moment Z component (sub. 3), and the sum of E-state indices (whole molecule), the best generated linear model of MLR included Verloop B_1 (sub. 4), moment of inertia 3 length (whole molecule), and log P (whole molecule), indicating their major influence on biological activity.

Model development

After selecting the necessary subsets of descriptors from PCA and original descriptors, statistical models were generated for the training set compounds. The relationship between selected structural parameters and biological activities was quantified by MLR and PLS implemented in TSAR 3.3. Values for F -to-enter and F -to-leave were set to 4. Identification of outliers present in the series is important as they may lead to a poor result. Data points that cannot be described using the QSAR equations are referred to as outliers [51]. Calculating and plotting the residuals and standardized residuals helped to find the possibility of having outliers in our data set. Five compounds (3, 13, 23, 28, and 42) with higher residual values were plotted away from the regression line and were identified and deleted as outliers.

MLR analysis

MLR calculates an equation describing the relationship between a single dependent y variable that is the biological data and several explanatory independent x variables that are physicochemical parameters. The relationship between biological activities expressed as $\log(1/IC_{50})$ and PCs and a reduced set of original descriptors was analyzed statistically by fitting the data to regression equations consisting of various combinations of these parameters. The best model was selected on the basis of various statistical parameters such as correlation coefficient (r), coefficient of determination (r^2), standard deviation (SD), sequential Fisher test (F), and test for statistical significance (t).

PLS

PLS analysis calculates equations describing the relationship between one or more dependent variables and a group of explanatory variables. PLS linear regression is a recent

technique that generalizes and combines features from PCA and multiple regressions. PLS is a method suitable for overcoming the problems in MLR related to multicollinear or overabundant descriptors. To check the robustness and predictive ability of the models generated using MLR, PLS analysis was performed.

Model validation

QSAR model validation is an important part of understanding statistically robust models capable of making accurate and reliable predictions of biological activities of new compounds not present in the data set. The model was validated on the basis of various statistical parameters such as fraction of variance (r^2), which gives information about the goodness of fit of a model (for a predictive QSAR model r^2 value should be greater than 0.60 [52, 53]); cross-validation test (r_{cv}^2 ; for a very robust model its value should be greater than 0.50); standard deviation (s), which is an absolute measure of quality of fit of the model (the smaller the value of s , the better is the QSAR model); and Fischer statistics (f), which is a measure of the level of statistical significance of the regression model (the larger the value, the greater is the probability that the QSAR model is significant). Overfitting of the model was also checked in order to validate the given model. Generally a given model overfits if it includes more descriptors than required. In order to avoid overfitting in our model a small number of descriptors was used for the model development.

Acknowledgments The authors pay sincere thanks to Prof. Aditya Shastri, Vice Chancellor, Banasthali University, Rajasthan, India for providing necessary computational facilities for the completion of the study in a convenient manner.

References

1. Barrett MP (1999) Lancet 353:1113
2. WHO Expert Committee (1998) Control and surveillance of African Trypanosomiasis. In: WHO Technical Report Series. World Health Organization, Geneva, vol 881. <http://www.who.int>; <http://www.who.int/tdr>
3. Enanga RJS, Burchmore ML, Barrett MP (2002) Cell Mol Life Sci 59:845
4. Welburn SC, Maudlin I (1999) Parasitol Today 15:399
5. Barrett MP (2000) Curr Opin Infect Dis 13:647
6. Nichols AC, Yielding L, Agbe SAO (2000) J Parasitol 86:177
7. Keiser J, Burri C (2000) Acta Tropica 74:101
8. Keiser J, Ericsson O, Burri C (2000) Clin Pharmacol Ther 67:478
9. Pepin J, Milord F (1994) Adv Parasitol 33:1
10. Legros D, Fournier C, Etchegorry MG, Maiso F, Szumilin E (1999) Bull Soc Pathol Exot 92:171
11. Legros D, Evans S, Maiso F, Enyaru JCK, Mbulamberi D (1999) Trans R Soc Trop Med Hyg 93:439
12. Barrett MP, Fairlamb AH (1999) Parasitol Today 15:136
13. Renslo AR, McKerrow JH (2006) Nat Chem Biol 2:701

14. Tidwell RR, Boyken DW (2003) In: Demeunynck M, Baelly C (eds) Small molecule DNA and RNA binders: from synthesis to nucleic acid complexes, vol 2. Wiley, Weinheim, p 414
15. Faerlamb AH (2003) Trends Parasitol 19:488
16. Bouteille B, Oukem O, Bisser S, Dumas MB (2003) Fundam Clin Pharmacol 17:171
17. Singh H, Sivakumar RJ (2004) Infect Chemother 10:307
18. Jannin J, Cattand P (2004) Curr Opin Infect Dis 17:565
19. Donkor IO, Huang TL, Tao B, Rattendi D, Lane S, Vargas M, Goldberg B, Bacchi C (2003) J Med Chem 46:1041
20. Wilson WD, Nguyen B, Tanious FA, Mathis A, Hall JE, Stephens CE, Boykin DW (2005) Curr Med Chem 5:389
21. Soeiro MNC, De Souza EM, Stephens CE, Boykin DW (2005) Expert Opin Invest Drugs 14:957
22. Dykstra CC, McClernon DR, Elwell LP, Tidwell RR (1994) Antimicrob Agents Chemother 38:1890
23. Bailly C, Dassonneville L, Carrascol C, Lucasl D, Kumar A, Boykin DW, Wilson WD (1999) Anti Cancer Drug Des 14:47
24. Fitzgerald DJ, Anderson JN (1999) J Biol Chem 274:27128
25. Henderson D, Hurley LH (1995) Nat Med (NY) 1:525
26. Paliwal SK, Pal M, Siddiqui AA (2009) Med Chem Res 19:475
27. Kurup A, Grag R, Carini DJ, Hansch C (2001) Chem Rev 101:2727
28. Athri P, Wenzler T, Ruiz P, Brun R, Boykin DW, Tidwell RR, Wilson WD (2006) Bioorg Med Chem 14:3144
29. Tropsha A, Cho SJ, Zheng W (1999) "New tricks for an old dog": development and application of novel QSAR methods for rational design of combinatorial chemical libraries and database mining. In: Parrill AL, Reddy MR (eds) Rational drug design: novel methodology and practical applications, ACS Symposium Series 719. American Chemical Society, Washington, DC, p 198
30. Yao SW, Lopes VHC, Fernández F, García-Mera X, Morales M, Rodríguez-Borges JE, Cordeiroa MNDS (2003) Bioorg Med Chem 11:4999
31. Cramer RD (1993) Perspect Drug Discov Des 1:269
32. Verloop A, Tipker HW (1976) Development and application of new steric substituent parameters in drug design. In: Ariens EJ (ed) Drug design, vol 7. Academic, New York, p 165
33. Verloop A, Tipker J (1976) Pestic Sci 7:379
34. Verloop A, Tipker JA (1977) Comparative study of new parameters in drug design. In: Keverling BJA (ed) Biological activity and chemical structure. Elsevier, Amsterdam, p 63
35. Neidle S (2001) Nat Prod Rep 18:291
36. Lipinski CA, Lombardo F, Dominy DW, Feeney PJ (2001) Adv Drug Deliv Rev 46:3
37. Bakunov SA, Bakunova SM, Wenzler T, Barszcz T, Werbovets KA, Brun R, Tidwell RR (2008) J Med Chem 51:6927
38. Sadowski J, Gasteiger (1993) J Chem Rev 93:2567
39. Wu W, Walczak B, Massart DL, Heuerding S, Erni F, Last IR, Prebble KA (1996) Chemometr Intell Lab Syst 33:35
40. Yasri A, Hartsough DJ (2001) Chem Inf Comput Sci 41:1218
41. Burt C, Richards GWJ (1990) Comput Aided Mol Des 4:231
42. Carbo R, Leyda L, Arnau M (1980) Int J Quantum Chem 17:1185
43. Good AC, So SS, Richards WG (199) J Med Chem 36:433
44. Good AC, Peterson SJ, Richards WG (1993) J Med Chem 36:2929
45. Burt G, Huxley P, Richards WG (1990) J Comput Chem 11:1139
46. Kubinyi H, Hamprecht FA, Mietzner T (1998) J Med Chem 41:2553
47. Si HZ, Wang T, Zhang KJ, Hu ZD, Fan BT (2006) Bioorg Med Chem 14:4834
48. Adams MJ (1995) Chemometrics in analytical spectroscopy. Royal Society of Chemistry, Cambridge
49. Nosov VN (1990) Computational biometrics. Moscow University Press, Moscow (in Russian)
50. Rameshwar N, Krishna K, Ashok KB, Parthasarathy T (2006) Bioorg Med Chem 14:319
51. Kim D, Hong S, Lee D (2006) Int J Mol Sci 7:485
52. Golbraikh A, Tropsha A (2002) J Mol Graphics Model 20:269
53. Tropsha A, Gramatica P, Gombar VK (2003) QSAR Comb Sci 22:69