Three-dimensional pharmacophore modelling studies on antagonists of endothelin receptor ET_A

V. S. Babu[1], V. N. Balaji[1,*], Vellarkad N. Viswanadhan[1], Ming Fai Chan[2], Samiron Phukan[1,3] and Shashidhar N. Rao[4,5]

[1]Jubilant Biosys Ltd, #96, Industrial Suburb, 2nd Stage, Yeshwantpur, Bangalore 560 022, India
[2]Accugent Laboratories Inc, 5933 Sea Lion Place, Suite 105, Carlsbad, CA 92010-6624, USA
[3]Present address: Lupin Limited, Genesis Campus, 1st Floor, Phase II, Hinjewadi, Pune 411 057, India
[4]Schrodinger Inc, 120 West 45th Street, New York, NY 10036, USA
[5]Present address: Tripos International Inc, 1699 South Hanley Road, St Louis, Missouri 63144, USA

We present 3D pharmacophore modelling and 3D quantitative structure–activity relationship (QSAR) studies on a collection of *N*-(isoxazolyl) sulphonamide endothelin (ET) receptor antagonists spanning nearly five orders of magnitude in their ET receptor antagonist activity. The alignments needed for the development of 3D QSAR models were obtained using the methods of pharmacophore generation and alignment in the Schrödinger module PHASE. Training and corresponding test sets for the 3D QSAR models were derived by binning the activities into five segments and choosing various percentages randomly within each bin. ET_A receptor antagonists training set correlation coefficients (R^2) and mean square errors ranged from 0.7 to 0.97 and 0.2 to 0.6 respectively. The test set predictions were characterized by Q^2 values ranging from 0.3 to 0.55, whereas the Pearson-*R* factors and root mean square errors varied from 0.6 to 0.75 and 0.2 to 0.8 respectively. Our studies clearly demonstrate that we have derived various models that are predictive for the test sets of ET receptor antagonists. Furthermore, we have employed data scrambling studies of the 3D pharmacophore models of ET_A receptor antagonists to validate the QSAR models.

Keywords: Antagonist, endothelin receptor, three-dimensional pharmacophore.

ENDOTHELIN (ET) family members are recognized as potent, endogenous vasoconstrictors. There are at least three known endothelin receptors, ET_A, ET_{B1} and ET_{B2}, all of which are G protein-coupled receptors whose activation results in elevation of intracellular-free calcium[1–6]. Three well-known ET receptor antagonists are ambrisentan, sitaxsentan (selective for ET_A) and bosentan (which affects both ET_A and ET_B)[7–16]. The compound sitaxsentan is a sulphonamide class ET receptor antagonist and was marketed in Europe under the brand name THELIN for pulmonary arterial hypertension. The rationale for benefit compared to bosentan, a nonselective ET blocker, is negligible inhibition of the beneficial effects of ET_B stimulation, such as nitric oxide production and clearance of ET from circulation. Bosentan is used to treat pulmonary hypertension by blocking the action of ETs that would otherwise promote vasoconstriction and lead to high blood pressure. The chemistry and biology of ET receptor antagonists have been widely studied[17–25]. Due to the G-coupled receptor nature of ET molecules, their structures are not solved to date. Consequently, the design of ET receptor antagonists has been directed by ligand-based approaches by experimental[17–25] and theoretical methods[26].

In light of their biological significance, we have undertaken molecular modelling studies on a collection of sulphonamide ET_A receptor antagonists, in which we have built 3D pharmacophore models common to all the active antagonists. These models have been employed to rationalize the measured biological activities of not only active antagonists, but also those of inactive and weakly active antagonists. The models developed for the receptor have been analysed to identify potential areas of selectivity in the hyperspace of 3D pharmacophores that led to the discovery of sitaxsentan and bosentan. We have subsequently used these models to filter for potential ET_A receptor antagonists from a database of commercially available compounds. Such compounds could serve as templates for the design of next-generation ET receptor antagonists as potential cardiovascular agents.

The dataset of *N*-(isoxazolyl) sulphonamide ET receptor antagonists used in this study was derived from Chan *et al.*[25] (Supplementary material will be available from the authors upon request). The two-dimensional preliminary structures of these compounds were built using ChemDraw and exported as 2D coordinates in mol (structure-data SD) files. The molecules were subsequently prepared using the Schrödinger ligand preparation tool, LigPrep, inside MAESTRO[27], by retaining the input tautomeric forms. None of the molecules has a chiral carbon, which implied that stereochemistry-related expansion of the molecular structures did not take place during their LigPrep treatment.

The prepared ligands were then passed through the 'PHASE' workflow[28] to generate common feature pharmacophore models, which in turn were used to generate 3D quantitative structure–activity relationship (QSAR) models. In this workflow, conformations were generated using the ligand torsion search method with default settings (maximum number of conformations = 1000, post-processing minimization refinement cycles = 50, energy cut-off = 10 kcal mol^{-1}, minimum distance cut-off for redundant conformations = 2.0). Feature sites (hydrogen-bond donors, acceptors, hydrophobe, ring aromatic, positive and negative centres) corresponding to all the conformations of all the compounds were generated using the default feature definitions.

*For correspondence. (e-mail: vnbalaji@jubilantbiosys.com)

The dataset of 73 compounds was divided into training and test sets. Towards this end, all these compounds were initially classified into five bins based on activities corresponding to the following ranges: $4 < pIC_{50} = -\log(IC_{50}) < 5$, $5 < pIC_{50} < 6$, $6 < pIC_{50} < 7$, $7 < pIC_{50} < 8$ and $8 < pIC_{50} < 9$. Training sets were constructed by choosing a percentage (50, 60 and 70) of the total number of compounds in each bin randomly. These random selections were done three times to select potentially different training sets. Correspondingly, the remaining 50%, 40% and 30% of the compounds in each bin were deemed to constitute corresponding test sets. This will ensure that the actives and inactives are reasonably distributed uniformly in each of the training and corresponding test sets. To clarify further, a random choice over the entire set may result in training and test sets that are skewed with respect to activity, i.e. a training set may end up with all active compounds with the corresponding test set dominated by inactive compounds. Such divisions will not yield predictive models and hence the activity binning approach was adopted.

In all the training sets, compounds with pIC_{50} values greater than 6.2 were deemed to be active for the purposes of developing common feature pharmacophores. These were developed with five features with the condition that all the active compounds in a training set contained the variants. Each variant can consist of any combination of the five features. For example, a variant can have four acceptors and a hydrophobe, whereas another can have one each of hydrophobe, ring aromatic, hydrogen-bond donor, hydrogen-bond acceptor and a negative ionic moiety. The pharmacophores were then scored and fine-tuned for overlap by the inactive compounds using default settings, which has been described previously[28]. The alignments of all the ligands to the top-ranked pharmacophores were used to generate QSAR models, setting the number of partial least square (PLS) factors 2–4. Predicted activities of test set compounds were plotted against their experimental activities, and the relevant statistics were computed. A test set compound is deemed to be an outlier if its predicted activity (pIC_{50}) differs from its experimental activity by more than 1 unit.

The PHASE methodology[28] allows the identification and construction of 3D pharmacophores for a collection of active molecules in a training set of compounds with biological activities, by exhaustively searching their multi-dimensional hypotheses spaces. The hypotheses in turn are collections of pharmacophoric features (which may be hydrogen-bond donors, hydrogen-bond acceptors, hydrophobes, aromatic rings, positive and negative ionic and ionizable centres) contained in the active molecules and characterized by their 3D arrangements. PHASE identifies the 3D pharmacophores as 'common feature hypotheses' and ranks them on the basis of how well the alignment of molecules correlates the activities of active and inactive molecules in the training and external test sets. These correlations are derived using the occupancy bits of various atom types in the 3D space encompassing all the aligned molecules, as independent descriptors. Since such descriptors are much larger in number than the number of observations (ET_A receptor antagonism) which constitute the dependent descriptors, PLS regression techniques[29] are employed to derive 3D QSAR models. These enable the extraction of key (latent) factors in the large collection of variables that account for most of the variance in the observed data, while eliminating redundancy in the descriptors[29].

The correlation between experimental and predicted activities is critical in identifying QSAR models, which can be used in a predictive manner. The goodness of such models is measured in terms of correlation coefficient (R^2) and cross-validated correlation coefficient (Q^2). Ideally both these values should be 1 for perfect models, whereas values away from 1 indicate less predictivity. However, given the expected errors from experimental measurements and from a practical point of view, values of $R^2 > 0.7$ and $Q^2 > 0.4$ would be useful in making reliable predictions on test sets of compounds. The Pearson-R factor is the square root of R^2, and would required to be above 0.8 in case of predictive models.

It may also be noted that PHASE does not consider (by default) the energy differences associated with different conformations so long as they all belong to an ensemble with energies below a certain cut-off value (10 kcal/mol in our studies). All the conformations in such an ensemble are treated equally in the generation of common-feature hypotheses. Since the conformations are generated using a comprehensive force field such as OPLS2007, they are free of any steric hindrance. Also, since the pharmacophores are being generated only from active compounds, it is assumed that they would not be having steric clashes with the receptor residues in their binding sites.

Although we have conducted the pharmacophore and 3D QSAR analyses for nine sets of training and test sets, for each percentage split, we report only the results corresponding to the best test set predictions. Thus, only three sets of results are discussed, corresponding to training set percentages of 50, 60 and 70 (Supplementary material will be available from the authors upon request). In light of the somewhat limited size of the total dataset, there are non-trivial overlaps between the members of various training sets. The analyses of their compositions show that the percentages of overlaps between the three pairs of divisions (50 : 50 and 60 : 40), (50 : 50 and 70 : 30) and (60 : 40 and 70 : 30) are 28, 45 and 58 respectively. The statistics associated with the 3D QSAR models developed on the basis of the alignments to top-ranked pharmacophores is listed in Table 1. As indicated earlier, R^2 and Q^2 values of greater than 0.7 and 0.4 respectively, indicate the robustness of the models in terms of their applicability towards predictions of test set activities. Also listed in Table 1 are the values of standard deviation and

Table 1. Statistics associated with 3D QSAR models for ET_A receptor antagonists

ET_A Set	No. of PLS factors	SD	R^2	F	P	RMSE (training)	RMSE (test)	Q^2	Pearson-R
50 : 50 Division	2	0.3449	0.8865	133	8.58E-17	0.3306	0.8488	0.289	0.6249
	3	0.2484	0.9429	182	1.41E-20	0.2346	0.7752	0.4069	0.6824
	4	0.1954	0.9657	225	6.02E-23	0.1818	0.8099	0.3527	0.6596
60 : 40 Division	2	0.5443	0.7043	48.8	1.42E-11	0.5254	0.7215	0.5097	0.721
	3	0.4728	0.7823	47.9	2.62E-13	0.4508	0.7171	0.5158	0.7215
	4	0.412	0.8388	50.7	6.06E-15	0.3879	0.687	0.5555	0.7492
70 : 30 Division	2	0.4572	0.7798	85	1.69E-16	0.4436	0.7813	0.4428	0.7096
	3	0.3565	0.8689	104	9.55E-21	0.3423	0.7679	0.5104	0.7278
	4	0.2975	0.9106	117	1.65E-23	0.2826	0.8192	0.4931	0.695
70 : 30 Division	2	0.723	0.449	19.6	6.13E-07	1.07		0.044	0.312
(data scrambled)	3	0.672	0.534	18	6.42E-08	1.06		0.061	0.313
	4	0.635	0.59	16.8	1.50E-08	1.14		–0.081	0.188

SD, Standard deviation; R^2, Square of the correlation coefficient or the training set Pearson-R value; RMSE, Root mean square error; Q^2, Square of the cross-validated correlation coefficient (Q); Pearson-R, Correlation coefficient for the test set of compounds; F, Measure of the statistical significance of the overall fit of the activity data to the model; P, P-factor or the P-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. The lower the P-value, the less likely the result is if the null hypothesis is true, and consequently more 'significant' the result is, in the sense of statistical significance, and PLS, Partial least square.

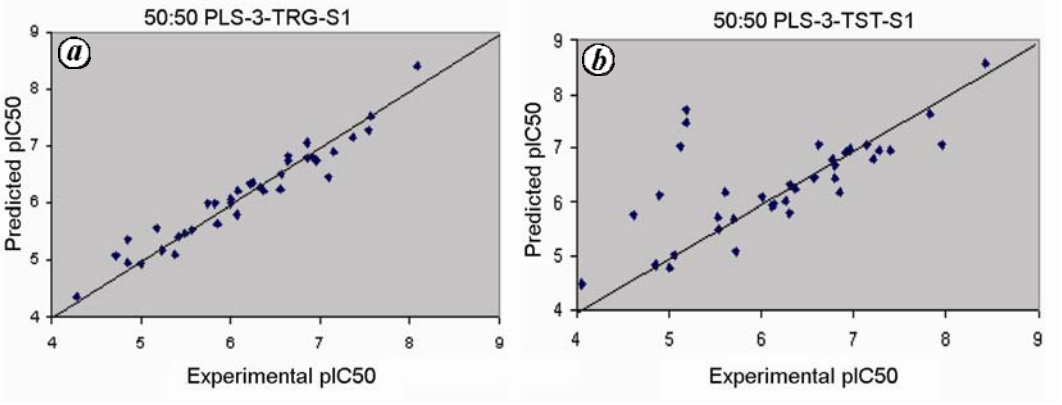


Figure 1. Correlation plots for training (***a***) and test (***b***) set compounds between experimental and predicted pIC_{50} values based on models derived with 50 : 50 division of ET_A receptor antagonism data using three partial least square (PLS) factors.

root mean square error (RMSE) of predictions, which should ideally be 0.0, but are typically less than 1.0 for good working models. In general, given the experimental errors, the RMSE values are about 2–3 times larger than the standard deviations. The scatter plots of predicted activities versus experimental activities of the training and test set compounds of the models are shown in Figures 1–3.

In the cases of 50 : 50 and 70 : 30 divisions, the test set predictions are better correlated with experimental data when the models with three and four PLS factors are employed in comparison to the predictions using models with two PLS factors. This improved performance with higher number of PLS factors is to be expected as the variance in the descriptors relative to activity is better explained by a larger number of them. However, it may be noted that the usage of excessive number of PLS factors can lead to over-fitting of the training set and consequently larger errors in test set predictions. In the case of 60 : 40 division, increasing number of PLS factors leads to models with progressively better predictions for test set compounds. This is further illustrated in the tables of predicted values and the correlation plots corresponding to all the three PLS factors (2, 3 and 4) shown in the supplementary material (Appendix 1; Supplementary material will be available from the authors upon request). The PLS factors themselves are detailed in Appendix 2 of the supplementary material.

The best value of Q^2 (0.56) is obtained for models generated with 60 : 40 division, with corresponding values of Pearson-R and RMSE being 0.749 and 0.687 respectively. The best values of Q^2 for models with 70 : 30 and 50 : 50 divisions are 0.51 and 0.41 (with three PLS factors). It may be noted that these values generally indicate the good predictive nature of the models we have developed. However, as would normally be expected, some of the compounds (e.g. compound 55 with pIC_{50} of 7.56) are

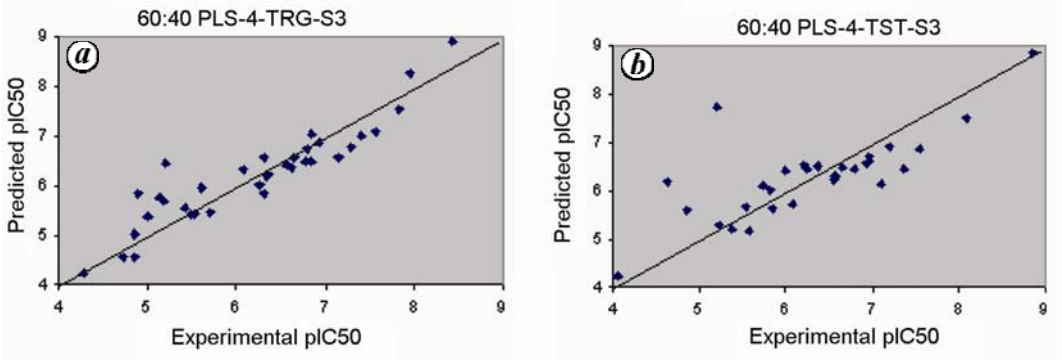


Figure 2. Correlation plots for training (***a***) and test (***b***) set compounds between experimental and predicted pIC_{50} values based on models derived with 60 : 40 division of ET_A receptor antagonism data using four PLS factors.

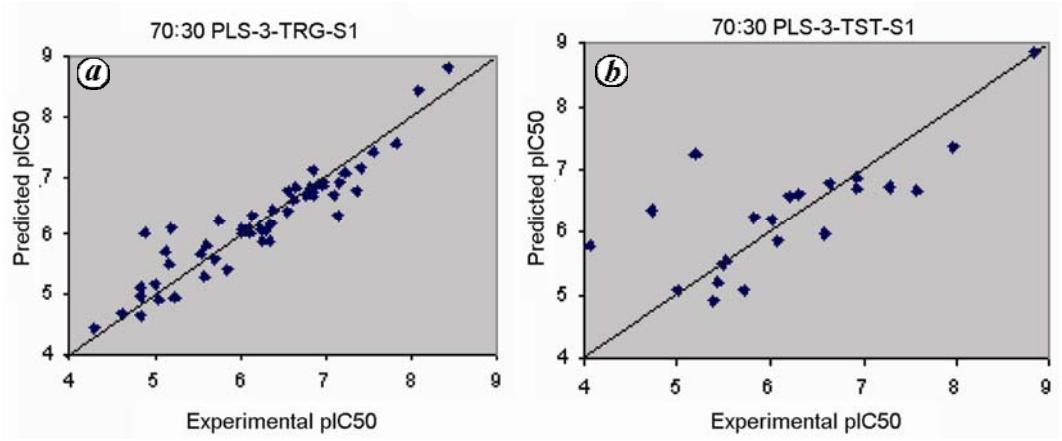


Figure 3. Correlation plots for training (***a***) and test (***b***) set compounds between experimental and predicted pIC_{50} values based on models derived with 70 : 30 division of ET_A receptor antagonism data using three PLS factors.

Table 2. Test set compounds predicted as outliers by various models of ET_A receptor antagonists

ET_A Set	No. of PLS factors	Outlier numbers
50 : 50 Division	2	15, 18, 27, 28, 40, 49
	3	15, 27, 28, 40, 49
	4	15, 27, 28, 40, 49
60 : 40 Division	2	15, 27, 28, 40, 49, 64
	3	22, 27, 28, 49
	4	27, 28, 49
70 : 30 Division	2	15, 18, 27, 28, 40, 67
	3	15, 18, 27, 28, 67
	4	15, 18, 27, 67

under-predicted and some mildly active compounds are over-predicted. As stated earlier, this can come about due to either more than normal inaccuracies in experimental data or due to the inaccuracies of representations of the pharmacophore features in the molecules.

Table 2 lists the compounds that are predicted as outliers by various models developed from all the three divisions of the training and test sets. The number of outliers varies between 3 and 6 for most of the models. In the case of the models derived from the two PLS factors from all the three divisions, there are six (nearly 20%) outliers. The models derived from 60 : 40 and 70 : 30 divisions (four PLS factors) are characterized by the least number of outliers. Compounds 27 and 28 seem to be predicted as outliers by a majority of the models studied. These compounds with four substituents are weakly active against ET_A receptor (pIC_{50} = 5.19), but are predicted to be considerably more active by most models, indicating possible disruption of the binding mode (e.g. due to a steric clash) not captured by the models. Outliers in test set predictions can come about for a number of reasons, some of which are as follows: (1) greater than expected or normal inaccuracies in experimental measurements. This is significant in light of the fact that we have defined the outlier as any compound where the predicted activity is more than one log unit away from the corresponding experimental activity. (2) Presence of features which do not clearly fall under one of the six defined feature types in PHASE. For example, chlorine atom is typically characterized as a hydrophobic group while in actuality, depending on its environment, it could potentially have characteristics of a hydrogen-bond acceptor. This type of scenario may also be true in case of tri-fluoromethyl

moiety. (3) Occurrence of similar molecules in training set as outliers. In such case, a prudent way would be to re-develop models that are constructed without the training set outliers.

The 3D QSAR models developed for ET_A antagonism data clearly demonstrate reasonable predictability of experimental data in test set compounds. The alignment of a couple of the active ligands to the highest scored 3D pharmacophore is shown in Figure 4. We have also shown the visualization of the QSAR model in terms of the favourable (blue cubes) and unfavourable regions (red cubes) where substitutions lead to enhancement and degradation of biological activities respectively (Figure 5). Each of these cubes may represent a composite of different pharmacophore feature types (hydrogen-bond donor, hydrogen-bond acceptor, positive and negative ionic functions, hydrophobic and aromatic functions) or due to a single one. Thus, they help distinguish the activities of different compounds on the basis of the nature of the substituent, a fact that is helpful to chemists in the design of molecules with desired biological activities. It may be noted that the conformational energies of the reference ligand conformations from which all the highly scored 3D pharmacophores are derived for the ET_A receptor antagonists, are within 5 kcal/mol of their respective global minima.

To further confirm the validity of these models, we have developed 3D QSAR models after we randomly scrambled the biological activity data. In this approach, the pIC_{50} data for ET_A antagonism were scrambled across the entire set of molecules (Supplementary material will be available from the authors upon request). This was followed by a 70 : 30 division of the 'scrambled data set' into training and test sets respectively, and subsequent development of scrambled 3D QSAR models with two, three and four PLS factors, as described in the text. We chose not to examine other ratios of training and test sets while developing the 3D QSAR models for the scrambled data set, since the purpose of this exercise is to establish that a predictive model cannot be obtained when the biological data are scrambled (improper assignment of structure–activity relationship). In that sense, the 70 : 30 ratio represents an extreme case since scrambling with a smaller dataset of training compounds is likely to lead to worse correlation between the predicted activities and experimental observations. The common features pharmacophore that was used for the alignment of the test and training set compounds consisted of three hydrogen-bond acceptors, a donor and a hydrophobe, with a poor selectivity value of 1.2, which contrasts with the values higher than 2 obtained for models generated with straight (unscrambled) activity data.

As is to be expected, we found that these scrambled models had Q^2 values less than 0.1 for test set predictions, whereas the R^2 values for training set predictions were less than 0.6 (Supplementary material will be available from the authors upon request). These poor R^2 values are to be expected as the relative alignments of the training set molecules against a set of common features do not correlate with the real active site pharmacophore. The

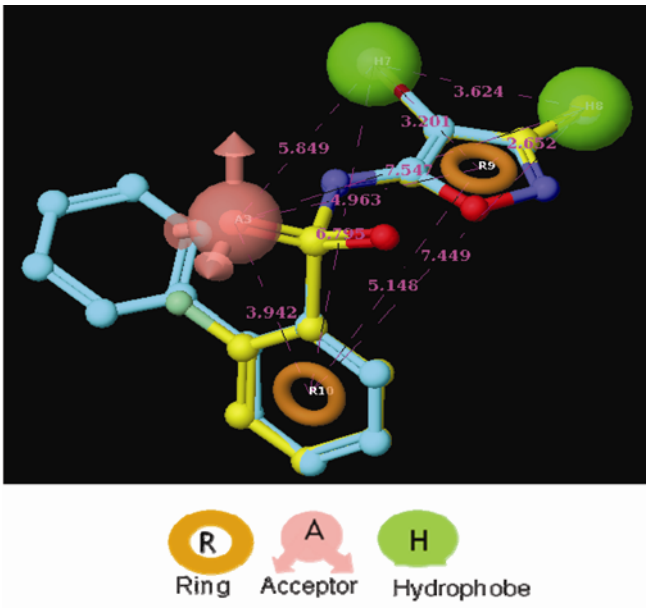


Figure 4. Alignments of active training set ET_A receptor antagonists. The 3D pharmacophore model of ET_A receptor antagonism data were derived with 60 : 40 division and four PLS factors. The 'aromatic' feature is exemplified by the light brown donut-like ring, while the hydrogen-bond acceptor feature is represented by the magenta-coloured sphere. The green-coloured spheres correspond to hydrophobic moieties. Note that the acceptor feature is characterized by arrows that represent potentially the directions of lone pairs.

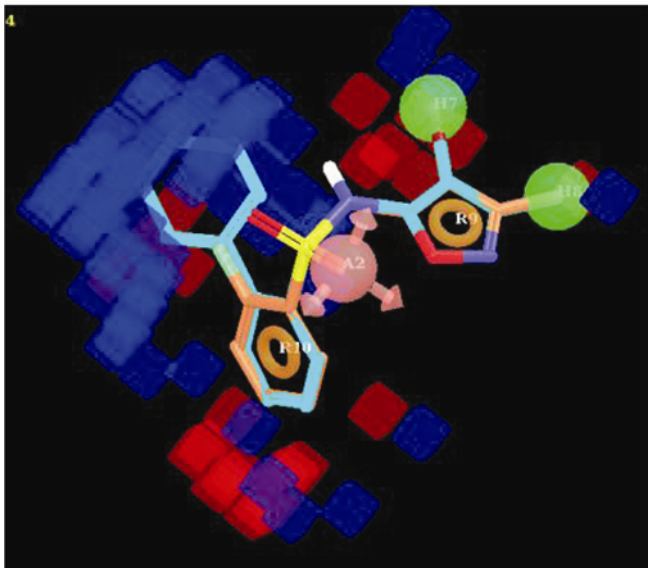


Figure 5. Schematic representation of the QSAR model derived on the basis of the alignment of training set compounds to the top-ranked pharmacophore model shown in Figure 5. The blue and red cubes correspond to regions where substituents lead to enhancement or deterioration of the biological activities respectively.

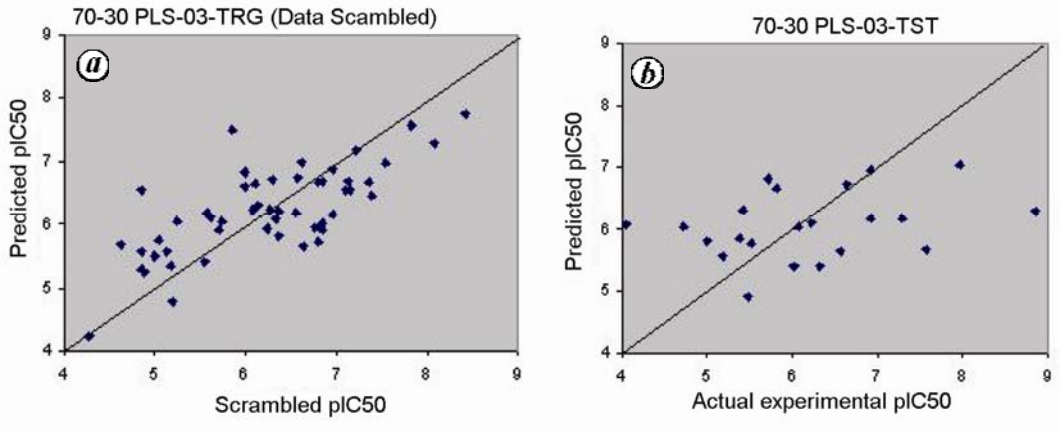


Figure 6. Correlation plot between experimental pIC_{50} and corresponding predicted data developed using 'scrambled' 70 : 30 training (***a***) and test (***b***) sets based on models developed with three PLS factors.

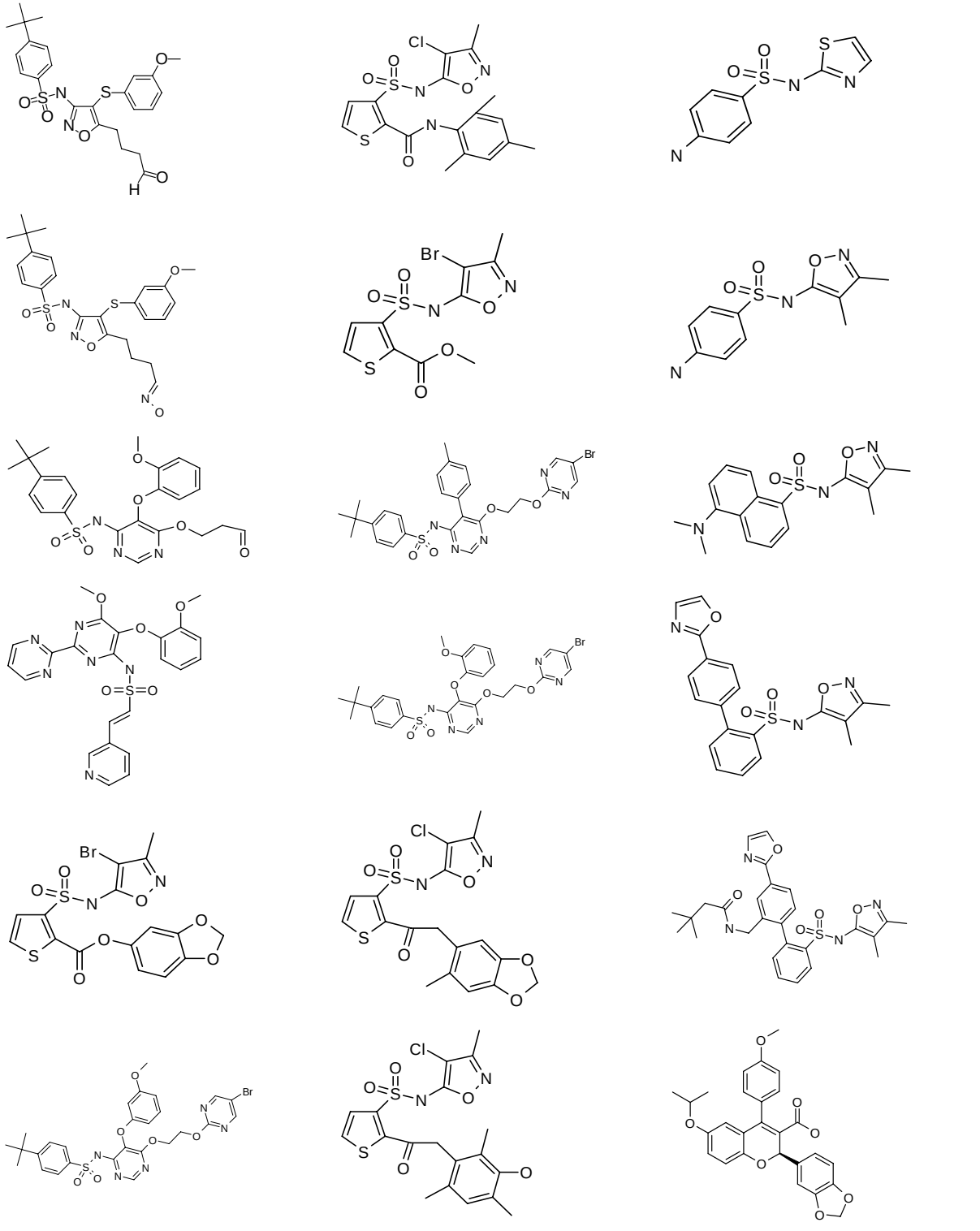
correlation plots corresponding to these models are shown in Figure 6. In these plots, the training set plot has the experimental activity that is from the scrambled data (Supplementary material will be available from the authors upon request). The plot for the test set data is against the actual compound activities. Hence, it is not surprising that the test set compounds are predicted poorly. These observations lend credence to the predictive strength of the models developed with straight activity data without being deemed 'ACCIDENTALLY PREDICTIVE'.

How do our 3D pharmacophore and 3D QSAR models compare with the CATALYST generated 3D pharmacophore models for ET_A receptor antagonists reported by Funk *et al.*[26], since the previous study has employed a different chemical class of compounds. The specific details of the pharmacophore models are naturally different. The primary difference lies in the presence of a negative ionic feature in the models reported by Funk *et al.*[26]. This ionic feature is absent in our models since our training and test set compounds do not contain any negatively ionizable functionality. Many of our top-ranked models contain two aromatic ring features instead of one, as in the previous study[26]. Also, the hydrogen-bond acceptor feature in our models is nearer to the pair of hydrophobic features than in the CATALYST models[26]. Another key difference between the two sets of model lies in the energy penalties paid by the training set active compounds to overlap with the pharmacophores. As reported in table 2 of Funk *et al.*[26], it appears that the fitting to the top-ranked CATALYST models was achieved by active molecule conformers that had energies >5 kcal/mol above the global minimum. By contrast, in our models, the conformations of active training set compounds that correspond to the derived hypotheses are within 5 kcal/mol of the global minimum. However, the most critical difference comes about in the realm of predictions of the activities of test set compounds. The previous study[26] had obtained large error factors (>10, corresponding to a deviation of more than one log unit) in the predictions of activities of nine (out of 30) test set compounds (30%). Also, they do not present any statistics related to the test set predictions and hence it is hard to judge specifically the performances of such models. Moreover, our studies have exhaustively searched the hypotheses space for all feasible models and consequently are able to identify models that show fewer outliers and closer correlation with experimental data on the test set compounds.

To realize the utility of the pharmacophore models developed using active members of various training sets, we have used a few of them to carry out enrichment studies, wherein a collection of 18 ET_A receptor antagonists (Table 3) was seeded in a decoy data set of 8000 ligands derived from public domain and Jubilant databases. The decoys were filtered for drug-like qualities ($200 <$ molecular weight < 600, number of hydrogen-bond donors < 5, number of hydrogen-bond acceptors < 12, cLogP < 8 and maximum number of rotatable bonds = 14). Table 4 illustrates the results of the enrichment studies in terms of the percentage of actives retrieved in the top 1–6% of the database compounds rank-ordered by the fitness score to the pharmacophore. It is not entirely surprising that due to the chemical diversity exhibited by the active antagonists, none of the three pharmacophores is able to retrieve all the actives in the screen. Pharmacophores derived from 60 : 40 and 70 : 30 divisions of the training set recover a maximum of 13 (72.22%) of the active compounds. There is an overlap in most of the actives recovered from the three pharmacophores. However, two compounds recovered by the 60 : 40 and 70 : 30 pharmacophores are not recovered through the screen against the 50 : 50 pharmacophore. Thus, a total of 13 compounds out of 18 (72.22%) are collectively recovered from the screen of the three pharmacophores.

The enrichment values are computed relative to the random recovery of actives from the entire database (0.22%). Thus, the results in Table 4 demonstrate that

Table 3. ET_A antagonists used for enrichment study



50% of the actives can be recovered in the top 1% of the database with decent enrichment values. The worst rank of the actives is 571, which corresponds to less than 7% of the entire database. Thus, in a drug-discovery scenario, the pharmacophores can be gainfully employed to quickly screen a large database and extract the top 7% of the ranks and expect to recover about 50–70% of the actives. Considering that only 3D pharmacophores are employed, this recovery rate is a useful first step in the lead discovery workflow, wherein the output from pharmacophore

Table 4. Results obtained from enrichment study in terms of the actives recovered in the top 1–6% of the database molecules ranked on fitness scores. Data are shown for the top-ranked pharmacophore models obtained using the 50 : 50, 60 : 40 and 70 : 30 ratios of training set to test set compounds. ENR refers to the enrichment values in these fragments of the database. Also shown is the rank of the last active compound recovered in a screen

	Top 1%	Top 2%	Top 3%	Top 4%	Top 5%	Top 6%	Active ligands retrieved
50 : 50 Pharmacophore							
No. of actives	9	9	11				07, 15, 11, 06, 05, 16, 12, 17, 14, 02, 03
% True actives	50	50	61.11				
ENR	50	25	20.3				
Rank of last active			321 (11)				
60 : 40 Pharmacophore							
No. of actives	9	9	9	9	10	13	07, 15, 16, 05, 11, 06, 12, 17, 14, 02, 04, 09, 03
% True actives	50	50	50	50	55.5	72.22	
ENR	50	25	16.7	12.5	11.1	12.1	
Rank of last active						571 (13)	
70 : 30 Pharmacophore							
No. of actives	9	9	9	9	10	13	07, 15, 16, 05, 11, 06, 12, 17, 14, 02, 04, 09, 03
% True actives	50	50	50	50	55.5	72.22	
ENR	50	25	16.7	12.5	11.1	12.1	
Rank of last active						571 (13)	

searches can be subjected to further treatment by more sophisticated treatment of structure-based methodologies such as docking, free energy calculations.

We have presented 3D pharmacophore modelling and 3D QSAR studies on a collection of *N*-(isoxazolyl) sulphonamide ET receptor antagonists spanning nearly five orders of magnitude in their ET receptor antagonist activity. Test set statistics clearly demonstrates the strong predictive nature of the models; the results of test set predictions are only moderately sensitive to the choice of the training and test sets, with generally better predictions for datasets with slightly larger training sets than test sets. We have used these 3D pharmacophore models in the enrichment and preliminary virtual screening studies, wherein active ET_A antagonists seeded in a random database of drug-like decoy molecules are recovered within the top 7% of the ranks based on pharmacophore alignment score. They offer new avenues for the discovery of novel chemical scaffolds for the development of ET receptor antagonists.

1. Liu, P., Hopfner, R. L., Xu, Y. J. and Gopalakrishnan, V., Vasopressin-evoked $[Ca^{2+}]i$ responses in neonatal rat cardiomyocytes. *J. Cardiovasc. Pharmacol*., 1999, **34**(4), 540–546.
2. Batra, V. K., Hopfner, R. L., Gopalakrishnan, V. and McNeill, J. R., Pneumadin-evoked intracellular free Ca^{2+} responses in rat aortic smooth muscle cells: effect of dexamethasone. *Biochem. Pharmacol*., 1999, **58**(1), 177–182.
3. Iwamuro, Y. *et al.*, Activation of two types of Ca^{2+}-permeable nonselective cation channel by endothelin-1 in A7r5 cells. *Br. J. Pharmacol*., 1998, **124**(7), 1541–1549.
4. Thorlin, T., Eriksson, P. S., Persson, P. A., Aberg, N. D., Hansson, E. and Rönnbäck, L., Delta-opioid receptors on astroglial cells in primary culture: mobilization of intracellular free calcium via a pertussis sensitive G protein. *Neuropharmacology*, 1998, **37**(3), 299–311.
5. Pollock, D. M., Keith, T. L. and Highsmith, R. F., Endothelin receptors and calcium signaling. *FASEB J*., 1995, **9**(12), 1196–1204.
6. Masaki, T., The discovery, the present state, and the future prospects of endothelin. *J. Cardiovasc. Pharmacol*. (*Suppl. 5*), 1989, **13**, S1–S4.
7. Kabunga, P. and Coghlan, G., Endothelin receptor antagonism: role in the treatment of pulmonary arterial hypertension related to scleroderma. *Drugs*, 2008, **68**(12), 1635–1645.
8. MacIntyre, I. M., Dhaun, N., Goddard, J. and Webb, D. J., Sitaxsentan sodium for pulmonary hypertension. *Drugs Today* (*Barcelona*), 2008, **44**(8), 585–600.
9. Cheng, J. W., Ambrisentan for the management of pulmonary arterial hypertension. *Clin. Ther*., 2008, **30**(5), 825–833.
10. Tempe, D. K., Datt, V. and Datta, D., Bosentan for the treatment of portopulmonary hypertension. *Ann. Card. Anaesth*., 2008, **11**(2), 139–140.
11. Newman, J. H., Kar, S. and Kirkpatrick, P., Ambrisentan. *Nature Rev. Drug Discov*., 2007, **6**(9), 697–698.
12. Dhaun, N. *et al.*, The pharmacokinetic profile of sitaxsentan, a selective endothelin receptor antagonist, in varying degrees of renal impairment. *Br. J. Clin. Pharmacol*., 2007, **64**(6), 733–737.
13. O'Callaghan, D. S. and Gaine, S. P., Sitaxsentan: an endothelin-A receptor antagonist for the treatment of pulmonary arterial hypertension. *Int. J. Clin. Pract*., 2006, **60**(4), 475–481.
14. Vatter, H. and Seifert, V., Ambrisentan, a non-peptide endothelin receptor antagonist. *Cardiovasc. Drug Rev*., 2006, **24**(1), 63–76.
15. Wu, C. *et al.*, Discovery, modeling, and human pharmacokinetics of N-(2-Acetyl-4,6-dimethylphenyl)-3-(3,4-dimethylisoxazol-5-ylsulfamoyl)-thiophene-2-carboxamide (TBC3711), a second generation, ET_A selective, and orally bioavailable endothelin antagonist. *J. Med. Chem*., 2004, **47**(8), 1969–1986.
16. Wu-Wong, J. R., Sitaxsentan (ICOS-Texas Biotechnology). *Curr. Opin. Invest. Drugs*, 2001, **2**(4), 531–536.
17. Kanda, Y. *et al.*, Synthesis and structure-activity relationships of potent and orally active sulfonamide ET_B selective antagonists. *Bioorg. Med. Chem*., 2001, **9**(4), 897–907.
18. Harada, H. *et al.*, Ethenesulfonamide and ethanesulfonamide derivatives, a novel class of orally active endothelin-A receptor antagonists. *Bioorg. Med. Chem*., 2001, **9**(11), 2955–2968.
19. Raju, B., Okun, I., Stavros, F. and Chan, M. F., Search for surrogates: A study of endothelin receptor antagonist structure activity relationships. *Bioorg. Med. Chem. Lett*., 1997, **7**(7), 933–938.

20. Raju, B., Wu, C., Castillo, R., Okun, I., Stavros, F. and Chan, M. F., 2-Aryloxycarbonylthiophene-3-sulfonamides: highly potent and ET_A selective endothelin receptor antagonists. *Bioorg. Med. Chem. Lett*., 1997, **7**(16), 2093–2098.
21. Morimoto, H. *et al.*, Modifications and structure-activity relationships at the 2-position of 4-sulfonamidopyrimidine derivatives as potent endothelin antagonists. *Bioorg. Med. Chem. Lett*., 2002, **12**(1), 81–84.
22. Wu, C. *et al.*, Endothelin antagonists: substituted mesitylcarboxamides with high potency and selectivity for ET(A) receptors-1. *J. Med. Chem.*, 1999, **42**(22), 4485–4499.
23. Wermuth, C. G., Selective optimization of side activities: another way for drug discovery. *J. Med. Chem.*, 2004, **47**(6), 1303–1314.
24. Chan, M. F. *et al.*, The discovery and structure-activity relationships of nonpeptide, low molecular weight antagonists selective for the endothelin ET(B) receptor. *Bioorg. Med. Chem*., 1998, **6**, 2301–2316.
25. Chan, M. F. *et al.*, Halogen substitution at the isoxazole ring enhances the activity of *n*-(isoxazolyl) sulfonamide endothelin antagonists-1. *Bioorg. Med. Chem. Lett.*, 1996, **6**(20), 2393–2398.
26. Funk, O. F., Kettmann, V., Drimal, J. and Langer, T., Chemical function based pharmacophore generation of endothelin-A selective receptor antagonists. *J. Med. Chem*., 2004, **47**(11), 2750–2760.
27. MAESTRO (version 8.5), A graphical user Interface for Schrödinger suite, developed and marketed by Schrodinger LLC, New York, USA.
28. Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E. and Friesner, R. A., PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput. Aided Mol. Des*., 2006, **20**, 647–671.
29. Tobias, R., *An Introduction to Partial Least Squares Regression*; http://support.sas.com/techsup/technote/ts509.pdf; Abdi, H., *Partial Least Squares (PLS) Regression*; http://www.utdallas.edu/~herve/Abdi-PLS-pretty.pdf; Hasegawa, K. and Funatsu, K., *SAR QSAR Environ Res.*, 2000, **11**(3–4), 189–209.

Received 14 June 2010; revised accepted 23 March 2011

Forest dynamics and carbon stocks in Rio Doce State Park – an Atlantic rainforest hotspot

Thiago Metzker[1,]*, Tereza C. Spósito[2], Mariana T. F. Martins[2], Marise B. Horta[2] and Queila S. Garcia[2]

[1]PPG-ECMVS, Ecology, Conservation and Wildlife Management and [2]Department of Botany, ICB–UFMG, CP:486, CEP:31270-901, Belo Horizonte, Minas Gerais, Brazil

Tropical forests are significant biodiversity epicentres and work as important modulators of climatic change. In natural cycles, carbon stocks are regulated by vegetation dynamics through the rates of mortality, recruitment and turnover. This study was conducted in five permanent monitoring plots (1 ha each: total 5 ha) in the Rio Doce State Park (RDSP), the most continuous Atlantic Forest remaining in Minas Gerais, Brazil and considered one of the world's hotspots. The aboveground biomass ranged from 201 Mg/ha in the primary forest to 92 Mg/ha in the secondary forest. The recruitment rate (1.8) was higher than the mortality rate (1.1); however, the average diameter of dead trees was higher than that of the recruited trees. Notwithstanding this result, the internal diametric increment (ingrowth) in RDSP was compensated by the biomass loss of dead trees, producing positive growth in the annual biomass and increasing their carbon stocks by 1.0 Mg C/ha/yr. Interest in conservation and management of tropical forests has intensified since the 2007 IPCC results. Thus, our study provides data that will help calibrate the use of international methodologies in local projects using data previously collected with a standardized methodology.

Keywords: Aboveground biomass, carbon stock, permanent monitoring plots, tropical forest.

TROPICAL forests store millions of tonnes of carbon and account for approximately 40% of all terrestrial carbon in terrestrial vegetation[1]. In natural cycles, the biomass stock ensures a balance in the carbon cycle through dynamic exchanges with the atmosphere. Forest dynamics is controlled by the balance between turnover, death and recruitment rates – essential components of net primary production (NPP)[2].

Recent studies suggest that turnover and growth rates of tropical forests have increased in the last few decades due to stimulation of forest growth caused by increase in CO_2 concentrations in the atmosphere[2–6]. However, Feeley[7] demonstrated that growth rates for forests in Panama and Malaysia have declined due to a decrease in precipitation and an increase in the minimum daily temperature.

Changes in tropical forest pattern dynamics can result in consequences for the biodiversity and economy of the planet. Biomass and biodiversity loss in forests due to burnout and deforestation, for example, affects the natural regime of the hydrologic cycle. Tropical forests play an important role in the control of rain precipitation through the release of volatile organic compounds (VOCs). VOCs function like cloud condensation nuclei and effectively control precipitation and cloud condensation formations[8]. Without this system of water replenishment, the whole hydrologic cycle can be compromised, resulting in a process of forest desertification.

The Atlantic Forest, a diverse tropical forest located in the Brazilian coast and inland, has high levels of biodiversity, endemism and environmental impacts. This forest is one of the world's biodiversity hotspots[9], and has only 8% of its original area. There are strong economic and social demands to use the remnant lands despite the effort

*For correspondence. (e-mail: thiagobio@ufmg.br)