An extended study on quantitative structure-antitrypanosomal activity relationships of sesquiterpene lactones

Njogu Mark Kimani [1], Josphat Matasyoh [2], Marcel Kaiser [3,4], Mauro S. Nogueira[1], Gustavo H. G. Trossini [5] and Thomas J. Schmidt [1]

[1]Institute of Pharmaceutical Biology and Phytochemistry (IPBP), University of Münster, PharmaCampus Corrensstraße 48, Münster D-48149, Germany
[2]Department of Chemistry, Egerton University, P.O. Box 536, Egerton 20115, Kenya
[3]Swiss Tropical and Public Health Institute (Swiss TPH), Socinstr. 57, Basel CH-4051, Switzerland
[4]University of Basel, Petersplatz 1, Basel CH-4003, Switzerland
[5]Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, Av. Lineu Prestes, 580, 05508-000 São Paulo, Brazil

Introduction

In continuation of a previous quantitative structure-activity relationship (QSAR) study on the antitrypanosomal activity of 69 sesquiterpene lactones (STLs) towards *Trypanosoma brucei rhodesiense* (*Tbr*) [1], the causative agent of the East African form of human African trypanosomiasis, QSAR models for a much larger and more diverse set (n=130; Figure 1) of such natural products were established in this study using three different complementary methodologies.

Detailed QSAR analyses using various complementary approaches yielded models with good internal and external predictive ability. For a set of compounds as chemically diverse as the one under study, the models exhibited good coefficients of determination (R^2) ranging from 0.71 to 0.85, as well as internal (leave-one-out Q^2 values ranging from 0.62 to 0.72) and external validation coefficients (P^2 values ranging from 0.54 to 0.73)

QSAR modelling

Three different QSAR approaches, all using linear regression modelling: (1) "Classical" descriptor-based QSAR using a genetic algorithm to select the most relevant variables, i.e. the same approach as in our previous study [1], (2) indicator variables deduced from pharmacophore features obtained from a 3D alignment of the most active molecules as applied in [2] and (3) hologram QSAR (HQSAR) based on fragments extracted from the molecular structure as used, e.g., in [3] were employed in this study.

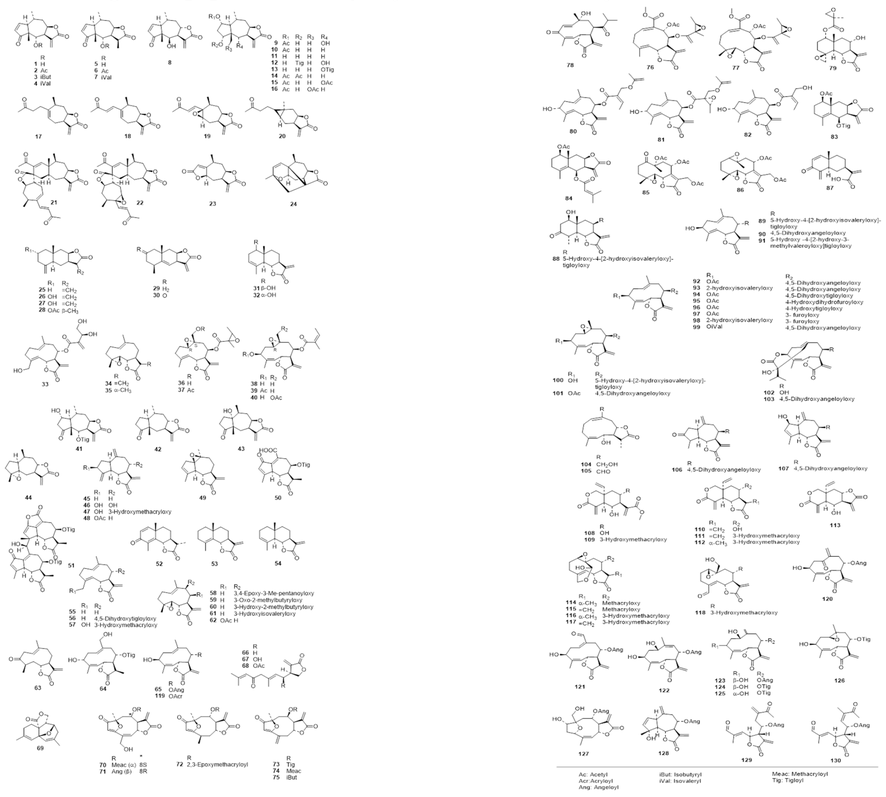

Figure 1: Structures of STLs used in the present QSAR analysis

In "Classical" descriptor-based QSAR, two models were developed; model 1 developed with the lowest force field energy structure, and model 2 with Boltzmann-weighted descriptor averages in order to include the aspect of molecular flexibility.

$pIC_{50}(Tbr) = 11.9040 + 0.0255\ ASAP4 + 0.0270\ \ ENONCS - 9.1834\ FASA^{+} + 2.1450\ \ npr1 - 1.6691\ vsurf\_CW2 + 0.6144\ \ vsurf\_ID2$
$(n = 87;\ R^2 = 0.72;\ Q^2 = 0.67;\ RMSE = 0.46;\ RMSEP = 0.51;\ model1)$

$pIC_{50}(Tbr) = -1.4803 + 0.0169\ ASAP4 + 0.0135\ ENONCS + 12.3791\ FASA^{-} - 2.1775\ vsurf\_CW2\ \ - 1.1349\ vsurf\_EWmin3 - 1.5861\ vsurf\_HB8$
$(n = 90;\ R^2 = 0.71;\ Q^2 = 0.67;\ RMSE = 0.47;\ RMSEP = 0.50;\ model2)$

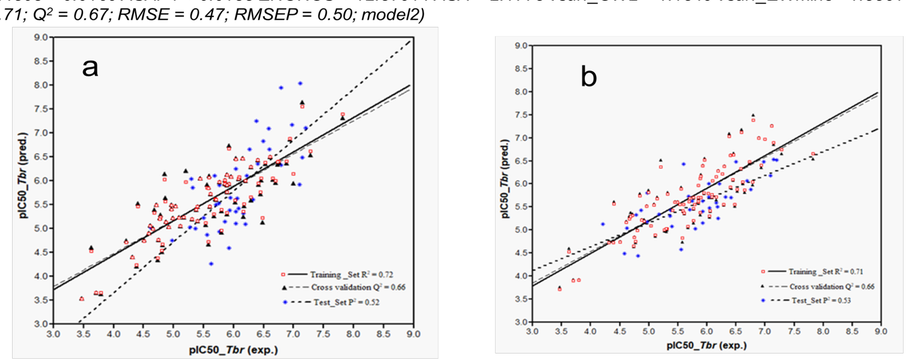

Figure 2: Plot of predicted versus experimental pIC50 values as obtained from the QSAR; (a) model 1 (b) model 2. NB: Three outliers, **32**, **107** and **119**, were removed from the test set

In the pharmacophore based model, each of the STLs were aligned with the superposed structures of the four most active compounds as template (see figure 3) and manually checked for the presence of various pharmacophore features. The indicator variables generated, 1 if a certain feature is present and 0 when otherwise, together with the global alignment score, S, were subjected to PLS analysis to develop model 3. In model 4, features F3 and F5, and F7 and F8 were combined to form features F3-5 and F7-8, respectively. The indicator variables were then combined with the classical descriptors used in model 1 and subjected to regression analysis to yield model 5.

$pIC_{50}(Tbr) = 1.8693 + 0.5940\ F1 - \ 0.1728\ F2 + 0.1552\ F3 + 0.2672\ F5 + 0.2714\ F7 + 0.3563\ F8 + 0.4939\ F9 - 0.2190\ S$
$(n = 86;\ R^2 = 0.73;\ Q^2 = 0.66;\ RMSE = 0.42;\ RMSEP = 0.48;\ model\ 3)$

$pIC_{50}(Tbr) = 1.46744 + 1.4336\ F1 - 0.6970\ F2 + 0.4715\ F3\text{-}5 + 1.0902\ F7\text{-}8 + 1.3065\ F9 - 0.0233\ S$
$(n = 86;\ R^2 = 0.71;\ Q^2 = 0.66;\ RMSE = 0.44;\ RMSEP = 0.48;\ model\ 4)$

$pIC_{50}(Tbr) = 5.3407 + 1.1855\ F1 + 0.6010\ F3\text{-}5 + 0.8498\ F7\text{-}8 + 1.3077\ F9 - 6.3704\ FASA+ + 3.6760\ FASA\_H$
$(n = 86;\ R^2 = 0.76;\ Q^2 = 0.72;\ RMSE = 0.40;\ RMSEP = 0.44;\ model\ 5)$

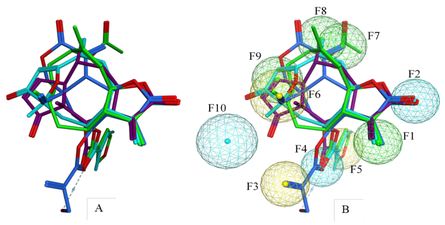

Figure 3: A: Alignment of four most active compounds as obtained by the flexible alignment function of MOE based on pharmacophoric atom properties; violet: **2**, light blue: **73**, green: **97**, blue: **111**. B: Common pharmacophore features of **2**, **73**, **97** and **111**. Pharmacophore spheres: blue: hydrogen bond acceptor; yellow: hydrophobe; green: reactive carbon of α, β or α, β, γ, δ unsaturated carbonyl groups. The features F1-F10 are numbered as used in the QSAR modelling.

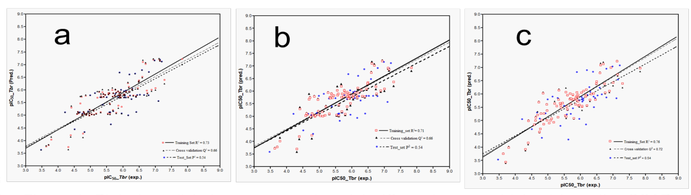

Figure 4: Plot of predicted versus experimental pIC_{50} values as obtained from the QSAR models 3 (a), 4 (b) and 5 (c)

HQSAR models were developed with the default SYBYL fragment size default settings (4-7 atoms) and all standard options of hologram length (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 and 401). The three best models are shown in bold in Table 1. The fragment distinction was then maintained and the fragment size was varied to determine the statistical influence of the variation on the output. For the three best fragment distinctions (Table 1) the fragment sizes: 2-5 atoms, 3-6 atoms, 5-8 atoms and 6-9 atoms were evaluated.

Table I: HQSAR models with different fragment size variations (n =87)

F_{dist}	F_{size} (Atoms)	Q^2	SEV	R^2	SEE	HL	PC	SEEP	P^2
A/C/H/Ch	2-5	0.55	0.61	0.74	0.46	199	5	-	-
	3-6	0.59	0.58	0.80	0.40	71	6	-	-
	4-7	**0.63**	**0.55**	**0.85**	**0.36**	**71**	**6**	**0.39**	**0.66**
	5-8	0.59	0.58	0.85	0.35	199	6	-	-
	6-9	0.52	0.62	0.89	0.30	401	6	-	-
A/C	2-5	0.56	0.60	0.75	0.46	199	5	-	-
	3-6	0.59	0.58	0.81	0.39	401	6	-	-
	4-7	**0.62**	**0.56**	**0.84**	**0.37**	**401**	**6**	**0.36**	**0.73**
	5-8	0.57	0.59	0.82	0.38	401	5	-	-
	6-9	0.57	0.58	0.83	0.38	257	5	-	-
A/C/Ch	2-5	0.59	0.58	0.76	0.44	199	5	-	-
	3-6	0.62	0.56	0.81	0.40	71	6	-	-
	4-7	**0.62**	**0.56**	**0.81**	**0.39**	**199**	**5**	**0.38**	**0.68**
	5-8	0.58	0.56	0.84	0.36	199	5	-	-
	6-9	0.60	0.58	0.84	0.36	353	5	-	-

F_{dist}: fragment distinction; F_{size}: fragment distinction; HL: hologram length; PC: number of PLS principal components; SEV: standard error of validation; SEE: standard error of estimation; SEEP: standard error of estimation of test set; A: atom; C: connection; H: hydrogen atom and Ch: chirality

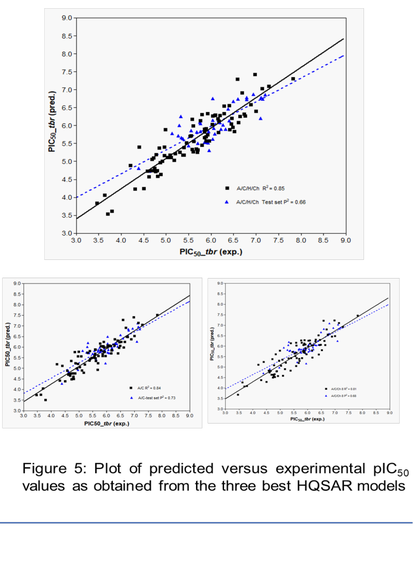

Figure 5: Plot of predicted versus experimental pIC_{50} values as obtained from the three best HQSAR models

Conclusions

- All the QSAR models have very comparable statistics (R^2 values: 0.71 - 0.85, Q^2 values: 0.62 - 0.72 and P^2 values: 0.54 - 0.73).
- The models further corroborate the fact that the antitrypanosomal activity of STLs is very much dependent on the presence and relative position of reactive enone or dienone groups within the molecular structure, but modulated by their hydrophilic/hydrophobic property and molecular shape; which confirms and extends previous QSAR studies.

References

1. Schmidt, T. J. *et al*. *Antimicrob. Agents Chemother.* **2014**, *58*, 325–332.
2. Schomburg, C. *et al*. *Eur. J. Med. Chem.* **2013**, *63*, 313–320.
3. Trossini, G.H.G. *et al*. *molecules.* **2014**, *19*, 10546-10562.

Acknowledgement

This work is performed within the framework of ResNetNPND: www.resnetnpnd.org