

1 Article

# 2 A 3D-QSAR study on the antitrypanosomal and 3 cytotoxic activities of steroid alkaloids by 4 comparative molecular field analysis

5 Charles O. Nnadi,<sup>1,2</sup> Julia B. Althaus,<sup>1</sup> Ngozi J. Nwodo,<sup>2</sup> Thomas J. Schmidt<sup>1,\*</sup>6 <sup>1</sup> Institute of Pharmaceutical Biology and Phytochemistry (IPBP), University of Münster, PharmaCampus  
7 Corrensstraße 48, D-48149, Münster, Germany; E-mail: [charles.nnadi@unn.edu.ng](mailto:charles.nnadi@unn.edu.ng)8 <sup>2</sup> Department of Pharmaceutical and Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of  
9 Nigeria Nsukka, 410001 Enugu State Nigeria; E-mail: [ngozi.nwodo@unn.edu.ng](mailto:ngozi.nwodo@unn.edu.ng)10 \* Correspondence: [thomschm@uni-muenster.de](mailto:thomschm@uni-muenster.de); Tel.: +49-251-83-33378

11

12 **Abstract:** As part of our research for new leads against human African trypanosomiasis (HAT), we  
13 report on a 3D-QSAR study for antitrypanosomal activity and cytotoxicity of aminosteroid-type  
14 alkaloids recently isolated from the African medicinal plant *Holarrhena africana* A. DC.  
15 (Apocynaceae), some of which are strong trypanocides against *Trypanosoma brucei rhodesiense* (*Tbr*)  
16 with low toxicity against mammalian cells. Fully optimized 3D molecular models of seventeen  
17 congeneric *Holarrhena* alkaloids were subjected to a comparative molecular field analysis (CoMFA).  
18 CoMFA models were obtained for both, the anti-*Tbr* and cytotoxic activity data. Model performance  
19 was assessed in terms of statistical characteristics ( $R^2$ ,  $Q^2$  and  $P^2$  for partial least squares (PLS)  
20 regression, internal cross-validation (leave-one-out) and external predictions (test set), respectively,  
21 as well as the corresponding SDEP and  $F$ -values). With  $R^2=0.99$ ,  $Q^2=0.83$  and  $P^2=0.79$  for anti-*Tbr*  
22 activity and  $R^2=0.94$ ,  $Q^2=0.64$ ,  $P^2=0.59$  for cytotoxicity against L6 rat skeletal myoblasts, both models  
23 were of good internal and external predictive power. The regression coefficients of the models  
24 representing the most prominent steric and electrostatic effects on anti-*Tbr* and for L6 cytotoxic  
25 activity were translated into contour maps and analyzed visually, allowing suggestions for possible  
26 modification of the aminosteroids to further increase the antitrypanosomal potency and selectivity.  
27 Very interestingly, the 3D-QSAR model established with the *Holarrhena* alkaloids also applied to the  
28 antitrypanosomal activity of two aminocycloartane-type compounds recently isolated by our group  
29 from *Buxus sempervirens* L. (Buxaceae), which indicates that these structurally similar natural  
30 products share a common SAR and, possibly, mechanism of action with the *Holarrhena* steroids. This  
31 3D-QSAR study has thus resulted in plausible structural explanations of the antitrypanosomal  
32 activity and selectivity of aminosteroid- and aminocycloartane-type alkaloids as an interesting new  
33 class of trypanocides and may represent a starting point for lead optimization.

34 **Keywords:** aminosteroid, aminocycloartane, alkaloid, 3D-QSAR, CoMFA, antitrypanosomal  
35 activity, cytotoxicity.  
36

## 37 1. Introduction

38 Understanding the structure-activity relationships (SARs) of bioactive molecules is a crucial  
39 prerequisite for lead optimization and further drug development. We recently reported on  
40 aminosteroid-type alkaloids from West African *Holarrhena africana* A. DC. (a synonym of *H. floribunda*  
41 (G. Don) T. Durand and Schinz) as a novel class of antitrypanosomal compounds with considerable  
42 activity against *Trypanosoma brucei rhodesiense* (*Tbr*) and low cytotoxicity which may represent a  
43 promising starting point towards new drugs against human African trypanosomiasis (HAT) [1]. In  
44 our first report, we also described some elementary structure-activity relationships, already  
45 perceived when comparing the structures and biological potency of the compounds. The mechanism

46 of action and biological target for the antitrypanosomal activities of these compounds still unknown,  
47 quantitative structure-activity relationships (QSAR) studies may represent the first step towards a  
48 better understanding of their promising bioactivity. QSAR studies play an important role in ligand-  
49 based drug discovery, design and further development [2, 3]. Comparative molecular field analysis  
50 (CoMFA) represents the earliest and most widely applied three dimensional (3D-) QSAR method [4].  
51 It employs statistical techniques to correlate biological endpoints with molecular features in terms of  
52 calculated interaction energies with virtual probes and interactive graphics can be used subsequently  
53 to interpret the results. Notably, the CoMFA method, originally developed with a set of steroids [5],  
54 has frequently been applied to numerous steroids and various biological endpoints [6-13], so that it  
55 appeared straightforward to apply it to the present set of alkaloids with pregnane and pregnene  
56 scaffolds. The compounds under study can roughly be grouped into (a) androstane and androst-5-  
57 ene derivatives with 3-amino- and 17 $\beta$ -acetyl substituents, (b) pregn-5-ene derivatives with a 3-amino  
58 substituent and an additional amino group connecting C-20 and C-18 forming a pyrrolidine ring and  
59 (c) pregnene derivatives with a C-3- or C-7-oxo substituent and a C-18/C-20 pyrrolidine or pyrroline  
60 ring similar to (b). The comparatively low flexibility within the steroid nucleus of these molecules  
61 makes their alignment, which is a most crucial step in 3D-QSAR, particularly in CoMFA, quite  
62 straightforward. The structural congenericity of the compounds in our data set furthermore makes it  
63 likely that they act by a common single target mediated mechanism. The corresponding endpoint  
64 data originate from a congruent series of determinations in the same laboratory and cover a  
65 considerable range of potency (2.5 log units in case of the antitrypanosomal activity). At the same  
66 time, the variations in substitution of the steroid core and a reasonable number of molecules in our  
67 set make it appear feasible to capture the most relevant structure-activity associations using a 3D-  
68 QSAR method such as CoMFA, which could, in future studies, be used to attempt predictions for  
69 (semi-) synthetic modifications to further optimize the activity profile.

## 70 2. Results and Discussion

### 71 2.1. Modeling and alignment of molecular structures

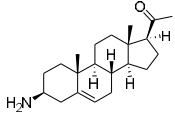
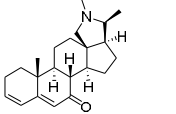
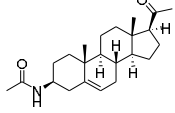
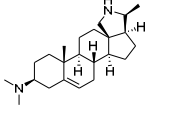
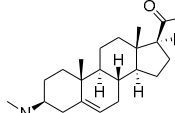
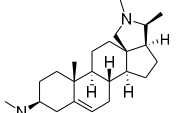
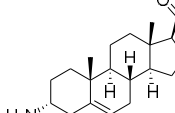
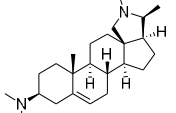
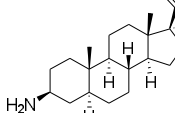
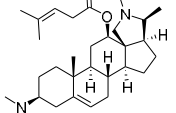
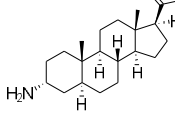
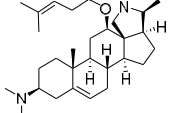
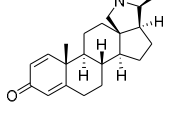
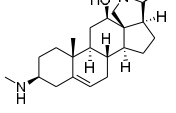
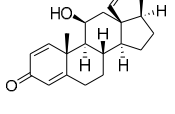
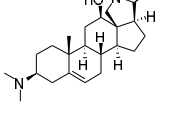
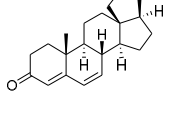
72 The structures and activity data of the *Holarrhena* alkaloids under study against *Tbr* and L6 rat  
73 skeletal myoblasts are presented in Table 1. The data were divided into a training set to generate the  
74 CoMFA models and a test set for subsequent external validation.

75 An optimized molecular model of each compound was generated using the Molecular  
76 Operations Environment (MOE [14]). The lowest energy conformer of the most active molecule (**3**,  
77 Figure 1A) served as the template scaffold for atom-by-atom alignment of all steroids. The atoms of  
78 the steroid nucleus marked in Figure 1A were used as alignment points for the compound  
79 superposition, which is shown in Figure 1B.

80

81

**Table 1.** Chemical structures and activities of steroid alkaloids used for 3D QSAR studies

Compounds	pIC <sub>50</sub> (Tbr)	pIC <sub>50</sub> (L6)	Compounds	pIC <sub>50</sub> (Tbr)	pIC <sub>50</sub> (L6)
	6.3958	5.2928		5.1326*	4.3800*
	5.3159	4.8135		6.7781	4.2972
	7.1249*	5.6057*		6.7781*	4.5618
	6.4320	4.7993*		6.3778	4.2130
	6.1726*	4.7707		6.9245	4.8444
	5.9190	4.7603		5.7807	4.0103*
	4.8282*	n.t		5.0856	3.7426
	5.0624	4.1816		4.6798	3.9045
	4.7568	3.8897			

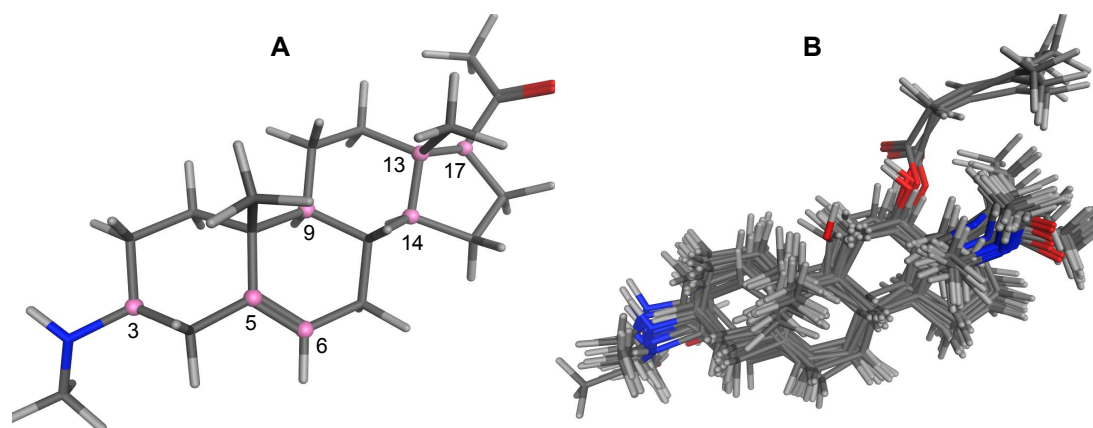
\*Test set compounds; pIC<sub>50</sub> = -log(IC<sub>50</sub>) as used in the CoMFA analysis; n.t = not tested

82

83

84

85



86

87

**Figure 1.** A: 3D model of **3** used as alignment template. Selected atoms (marked) were used as matching points for the superposition. B: Superposed 3D structures of all steroid alkaloids under study.

88

## 2.2. CoMFA modelling

89

90

91

92

93

94

95

96

97

98

99

The molecular structures were exported to the software *Open3DQSAR* [15]. After calculation of molecular interaction fields (MIFs) with a steric and an electrostatic probe and various data pre-treatment steps, PLS regression with up to five latent variables (PLS components) was performed to model linear relationships between the differences in the molecules' MIF energies with the changes in the trypanocidal (*Tbr*) and cytotoxic (L6) activities. The resulting models were cross-validated using the leave-one-out (LOO) method and the optimum number of components in the models for each activity selected based on coefficients of determination ( $Q^2$ ) of the cross validation predictions. The predictive ability of the models was furthermore assessed by activity predictions for the test sets of compounds excluded during CoMFA model generation. The general statistical parameters of the two resulting models are summarized in Table 2; More details on each model and the predicted  $pIC_{50}$  values of all compounds in both models are reported as Supplementary Materials (Tables S1-S3).

100

### 2.2.1. PLS and model statistics for anti-*Tbr* activities of steroid alkaloids

101

102

103

104

105

106

107

The best model developed for trypanocidal (*Tbr*) activity comprised an optimum number of 3 latent variables (PLS components, PCs) and confirmed a strong correlation between the variations in the MIFs and the in vitro activities of the steroid alkaloids. The PLS regression yielded a coefficient of determination,  $R^2$  of 0.99, a coefficient of determination,  $Q^2$  of 0.83 for the leave-one-out internal cross validation and a coefficient of determination,  $P^2$  of 0.79 for the predictions of the test set compounds' activities. The model statistics are summarized in Table 2 and a plot of predicted versus measured  $pIC_{50}$  values is presented in Figure 2A.

108

### 2.2.2. PLS and model statistics for L6 cytotoxic activities of steroid alkaloids

109

110

111

112

113

114

115

The CoMFA model for cytotoxic activities of steroid alkaloids on L6 rat skeletal myoblasts also yielded a strong PLS correlation with  $R^2 = 0.94$ ,  $Q^2 = 0.64$ . For this model only two PLS components were required, since no further significant increase of  $Q^2$  was observed in a three-component model. The predictive ability of this two-component model gave a coefficient of determination  $P^2$  of 0.59. The somewhat lower statistical quality of this regression model in comparison with the one for anti-*Tbr* activity may be explained by the more narrow range of biological data ( $\approx 1.9$  log units). For model statistics see Table 2 and for a plot of predicted versus measured activity data see Figure 2B.

116

117

118

119

120

121

**Table 2.** Statistics of the CoMFA models for trypanocidal (*Tbr*) and cytotoxic (L6 cells) activities. PC is the number of latent variables (PLS components, PCs) in each model.

Model statistics	Anti- <i>Tbr</i> , PC = 3	L6 cytotoxic, PC = 2
$R^2 \pm$ SDEC	$0.995 \pm 0.056$	$0.940 \pm 0.111$
$Q^2 \pm$ SDEP	$0.83 \pm 0.33$	$0.64 \pm 0.28$
$P^2 \pm$ SDEP	$0.79 \pm 0.51$	$0.59 \pm 0.42$
F-ratio	482.639	70.452
Model predictive regression	$y = 0.621x + 2.206$	$y = 0.674x + 1.420$

122

123

124

125

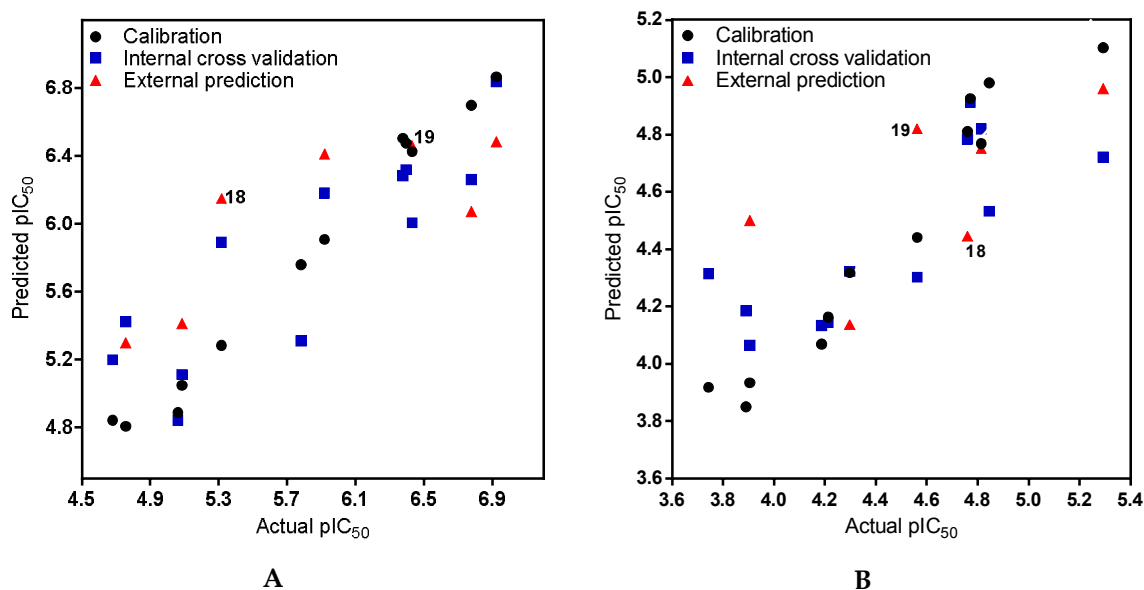
$R^2$  = non-cross validated coefficient of determination;  $Q^2$  = coefficient of determination for leave-one-out internal cross-validation;  $P^2$  = coefficient of determination for test set predictions;  $F$  = Fisher value; SDEC = standard deviation error in calculation; SDEP = standard deviation error in prediction; PC = number of latent variables (PLS components);  $x$  = actual  $pIC_{50}$ ;  $y$  = predicted  $pIC_{50}$ .

126

127

128

Most importantly, the high  $F$ -values and  $SDEP < P^2$  for both models indicate that the possibility of chance correlation in the models is small. More so, the closeness of the coefficients of determination for internal ( $Q^2$ ) validation and the external ( $P^2$ ) predictions confirms the robustness of the models.



129

130

131

132

133

134

**Figure 2.** Scatter plots of predicted versus actual activity of steroid alkaloids against *Tbr* (A) and L6 rat skeletal myoblasts (B). Data result from the PLS models with a 3 (A) and a 2 (B) components and represent non-cross validated (black circles), internal predictions by LOO cross validation (blue squares) and test set predictions (red triangles). The positions of the *Buxus* alkaloids 18 and 19 (see section 2.4) are marked. For numerical values of all compounds see Table S3, Supplementary Materials.

135

### 2.3. Analysis of the CoMFA contour maps

136

137

138

139

The regression coefficients of the final CoMFA models for the steroid alkaloids were translated into contour maps for both steric and electrostatic effects on antitrypanosomal (*Tbr*) activity and cytotoxic activity on L6 cells (see Figures 3 and 4, respectively). The contours reflect properties of a hypothetical common receptor binding site where variations of steric and electrostatic features of the

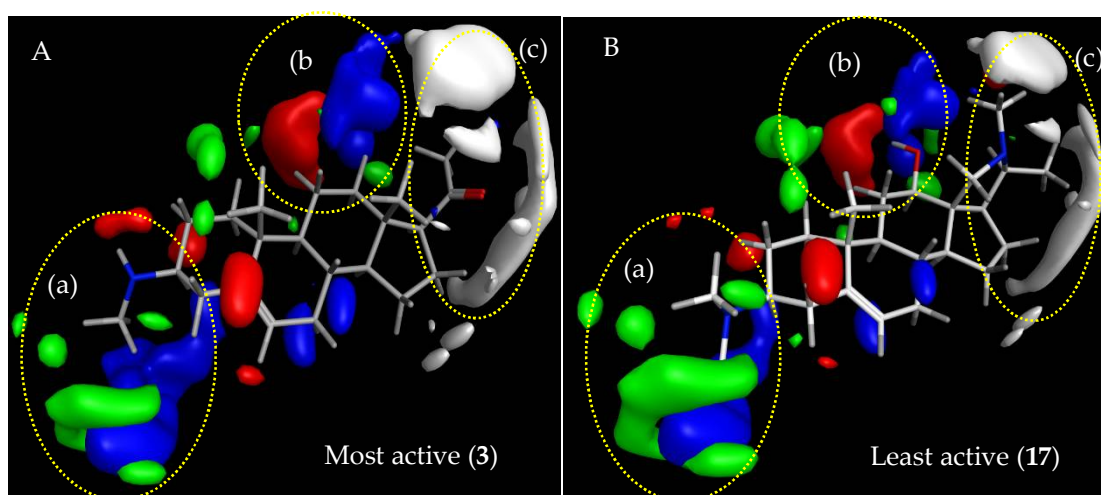
140 molecules affect most significantly the antitrypanosomal and cytotoxic activities of the compounds  
141 and give some general insight into the nature of the putative common receptor-ligand binding region.

### 142 2.3.1. CoMFA model for antitrypanosomal activity of steroid alkaloids

143 Contributions by steric effects are observed mainly in the vicinities of the C-3 amino group  
144 (region (a) in Figure 3, where a steric interaction tends to increase activity) and around the amino  
145 group of the pyrroline/pyrrolidine rings and the C-17 $\beta$ -acetyl or C-20 methyl groups (region (c) in  
146 Figure 3) where steric interaction tends to decrease activity. Region (a) accounts for interactions of  
147 methyl groups bound to the amino substituents in position C-3 where monomethylation significantly  
148 enhances the activity (compare **3** with **1** and **5**). On the other hand, methyl substituents on the  
149 pyrroline/ pyrrolidine ring of the weaker trypanocides (**17** shown in Figure 3B) protrude into the  
150 region (c) of sterically unfavourable interaction and obviously reduce this activity.

151 Prominent electrostatic effects are observed around the C-3 amino position (Figure 3, blue area  
152 in region (a)) and in the vicinity of the C-11/C-12 position of the steroid skeleton (Figure 3, red and  
153 blue areas in region (b); The N-atom of the C-18-N-C-20 bridge may also interact with the latter but  
154 the distance is relatively large and so that this is obviously of less importance since the activity is not  
155 strongly influenced by this structure element)). Electrostatic interaction with a positive charge on the  
156 putative receptor is favorable for antitrypanosomal activity mainly in the vicinity of the amino groups  
157 at C-3 while it can be either detrimental or favorable in the C-11/C-12 and C-18-N-C-20 area (red and  
158 blue areas in region (b), respectively).

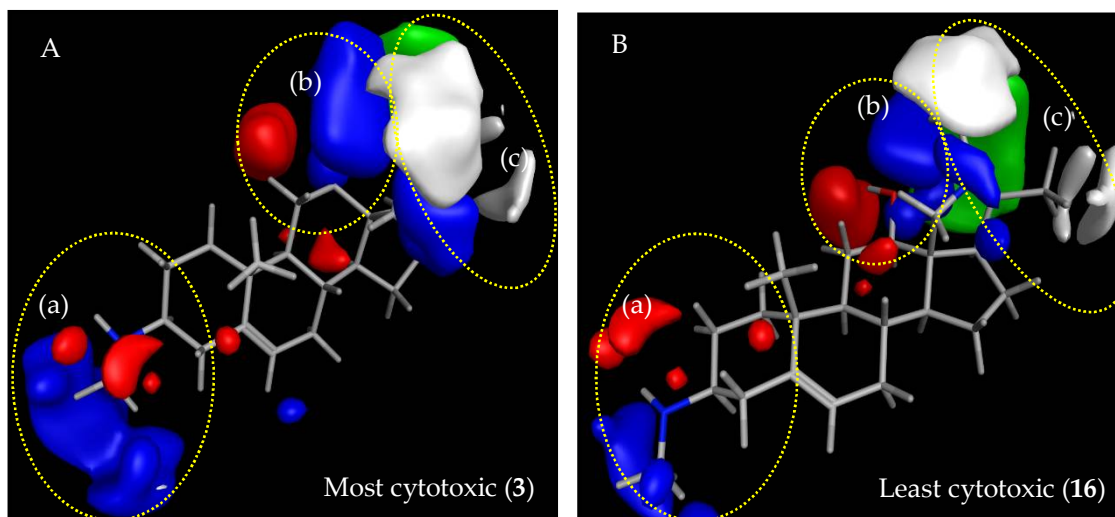
159



160

161 **Figure 3.** Steric and electrostatic contour maps representing the CoMFA model for anti-*Tbr* activity.  
162 Compounds shown are the strongest (A; **3**) and weakest (B; **17**) trypanocides in the series. Green and white  
163 regions, respectively, indicate areas where steric interactions increase and decrease activity. Blue and red  
164 regions denote enhancing and detrimental electrostatic effects with the positively charged probe,  
165 respectively.

166



167

168

169

170

171

172

**Figure 4.** Steric and electrostatic contour maps representing the CoMFA model for cytotoxic activity on L6 rat skeletal myoblasts. Compounds shown are the strongest (A; **3**) and weakest (B; **16**) cytotoxins in the series. Green and white regions, respectively, indicate areas where steric interactions increase and decrease activity. Blue and red regions denote enhancing and detrimental electrostatic effects with the positively charged probe, respectively.

173

### 2.3.2. CoMFA model for L6 cytotoxic activity of steroid alkaloids

174

175

176

177

178

179

180

The steric contour map of the L6 cytotoxic activities (Figure 4) is completely governed by interactions in the C-17 region (region (c) in Figure 4). The additional pyrrole/pyrrolidine rings of **7-17** protrude with their methyl substituents to the white (sterically unfavourable) region while the C-17 $\beta$ -acetyl groups of **1-6** only protrude into the green (sterically favourable) contour in this region suggesting that this substituent enhances cytotoxicity. Similarly, the 12-O-(4'-methyl-3'-pentenoyl) ester chain of **14** and **15** protrudes to the green region with the  $\Delta^{3,4}$ -double bond partially touching the sterically favourable region (not shown).

181

182

183

184

185

186

187

188

189

The electrostatic contour maps for the L6 cytotoxic activities of the steroid alkaloids resemble those in the model for anti-*Tbr* activity with an enhancing influence of region (a) near the C-3 amino group. The amino groups at position C-3 interact with the blue contour region so that they have a similar effect on cytotoxicity as on antitrypanosomal activity while the amino groups of the pyrrole/pyrrolidine rings are somewhat more remote from the respective blue contour in region extending in regions (b) and (c) where interactions with electropositive charge would favour the cytotoxic activity. The hydroxy group of the C-12 alcohols **16/17** is located between a red (unfavourable) and a blue (favourable) region while the ester oxygens of their counterparts **14/15** protrude more into the blue region (not shown).

190

191

### 2.3.3. Comparison of CoMFA models and considerations on antitrypanosomal selectivity of steroid alkaloids

192

193

194

195

196

197

198

199

200

201

As becomes clear from the previous sections, the CoMFA models for the two activities under study are relatively similar indicating that the putative biological receptor sites are related but differ enough to warrant considerable selectivity in some compounds. As already pointed out, in region (a), interaction of the C-3 amino group with a positive charge (blue region) on the receptor site is favored for both activities, which confirms our earlier postulation [1] that such substituents at the C-3 position are required for strong antitrypanosomal activity of steroid alkaloids against *Tbr* but also confer cytotoxicity. The green contour in region (a), observed only in the *Tbr* model, Figure 3, indicates that there might be a favorable steric interaction with the methyl groups of C-3-methylamino groups but the higher activity of such methylamines could equally be caused by their increased basicity, which appears even more likely since activity increases in the order -NH<sub>2</sub><-

202 N(CH<sub>3</sub>)<sub>2</sub>-NHCH<sub>3</sub> and not with the number of methyl groups. The lower anti-*Tbr* activity and  
203 cytotoxicity of the acetamide **2** compared with **1** and **3** must be attributed to a much weaker  
204 electrostatic interaction of the non-basic amide nitrogen in this region. Along the same lines, the  
205 necessity for electrostatic interactions with a basic group in region (a) for high anti-*Tbr* activity is also  
206 reflected in the comparatively low activities of **7-10**, none of which bears an amino group in this  
207 region. Compounds **7-9** are 3-oxosteroids and **10** is a 7-oxosteroid with only a hydrogen at C-3. Thus,  
208 the influences on activity of interactions in region (a) on anti-*Tbr* and cytotoxic activity appear to  
209 parallel each other suggesting that this position could not easily be further optimized to increase the  
210 selectivity towards *Tbr*.

211 While interactions in region (b) of both models are rather similar and may thus not yield any  
212 good points of attack for selectivity, the two models indicate that region (c) may represent a viable  
213 target in this respect. Interestingly, cytotoxic activity is most prominently influenced by both steric  
214 and electrostatic interactions in region (c) (Figure 4) comprising the pyrroline/pyrrolidine ring of **11-17**  
215 and the C-17 $\beta$ -acetyl substituent of **1-6**. The higher cytotoxic activity observed in compounds **1-6**  
216 in comparison with the pentacyclic diamines **11-17**, according to the model, may be explained by a  
217 significant steric contribution of the C-17 $\beta$ -acetyl group whose methyl group points towards a green  
218 area of favorable steric interaction in the L6 model, alongside a favorable electrostatic interaction of  
219 the keto oxygen with a blue electropositive region, which is larger than in the *Tbr* model and extends  
220 in both regions (b) and (c). An electrostatic interaction within this region should also be possible for  
221 the amino nitrogen in the pyrroline/pyrrolidine ring connecting C-18 and C-20 of **11-17**, which might  
222 hence tend to increase the cytotoxic activity of the compounds but this is apparently outmatched by  
223 the stronger steric effect of the methyl substituents at C-20 and the nitrogen atom, which point into  
224 an extended white area of unfavorable Van der Waals interaction. Nevertheless, it could be  
225 speculated that replacement of the nitrogen between C-18 and C-20 by, e.g., a methylene group, could  
226 further decrease the cytotoxicity because the mentioned electrostatic effect in this case would not be  
227 possible. A white contour area of detrimental steric interaction is also observed in the anti-*Tbr* model  
228 but is apparently of less influence. The pyrrolidine/pyrroline structure element in region (c) and, in  
229 particular, the methyl groups attached to it thus appear to reduce the cytotoxicity to a greater extent  
230 than antitrypanosomal activity and thereby to increase the selectivity of the alkaloids bearing this  
231 structural feature. While the C-17 $\beta$ -acetyl substituent of the pregnane/pregnene derivatives (**1-6**) as  
232 well as the additional ring in (**7-17**) is obviously well accommodated, this region of detrimental steric  
233 effect in the anti-*Tbr* model appears to be more sensitive to the differences in methylation at the C-  
234 18/20 nitrogen bridge. This is in line with the higher activity of **11** compared with **13**. These two  
235 compounds differ by the presence and absence of a methyl substituent on the pyrrolidine nitrogen in  
236 **13** and **11**, respectively. The unmethylated **11** yielded an IC<sub>50</sub> value against *Tbr* of 0.17  $\mu$ M while that  
237 of the methylated pyrrolidine **13** was 0.42  $\mu$ M indicating that this *N*-methyl group is responsible for  
238 an unfavourable steric interaction in region (c). Taking into account that compounds with a  
239 monomethylamino group at C-3 (region (a)) show a higher anti-*Tbr* activity than the respective  
240 dimethylamines, it may hence be hypothesized, that a derivative of the C-3-monomethylamine **12**  
241 without the *N*-methylation at the pyrrolidine nitrogen would be even more active since **12** already as  
242 such has an IC<sub>50</sub> of only 0.166  $\mu$ M. At the same time, the increase in cytotoxicity upon removal of this  
243 methyl group is not significant so that a further increased selectivity can be expected for *N*<sup>18</sup>-  
244 demethyl-**12**.

#### 245 2.4. Application of the CoMFA model to cycloartenoid alkaloids from *Buxus sempervirens*

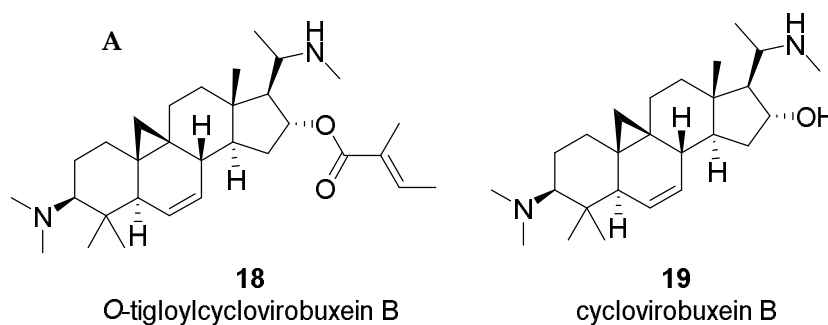
246 In a recent study aimed at compounds with antiplasmodial activity in European Box (*Buxus*  
247 *sempervirens* L., Buxaceae), we isolated the aminocycloartane-type alkaloid *O*-tigloylcyclovirobuxein  
248 **B (18)** which showed considerable in vitro activity against *Plasmodium falciparum* (*Pf*; IC<sub>50</sub>=0.92  $\mu$ M)  
249 and some activity, yet at a much lower level, against *Tbr* (IC<sub>50</sub>=3.7  $\mu$ M) [16]. In continuation of this  
250 study, a small amount of the unesterified alcohol, cyclovirobuxein B (**19**), was also isolated from this  
251 plant and tested for antiplasmodial as well as antitrypanosomal activity [17]. Surprisingly, this



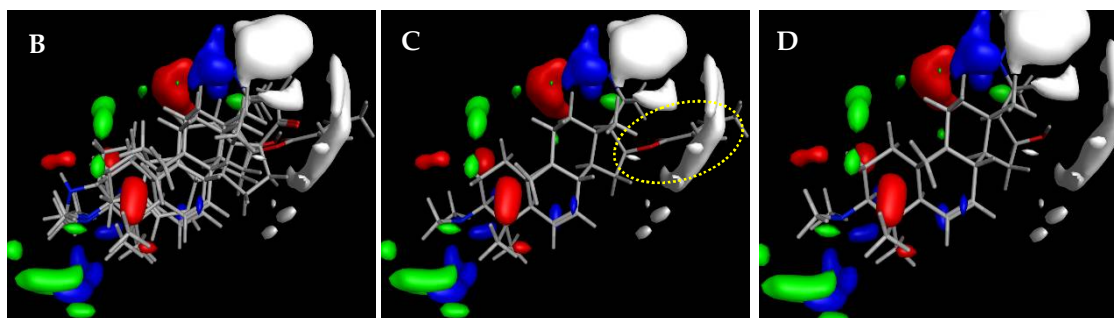
252 alkaloid displayed a very strong activity against *Tbr* with an IC<sub>50</sub> value of only 0.16 μM while the  
253 antiplasmodial activity was unchanged (0.98 μM) in comparison with the ester.

254 In view of the structural similarity (see Figure 6A, B) between these cycloartenoid alkaloids and  
255 the steroids from *Holarrhena*, it was straightforward to assume that these compounds might address  
256 the same molecular target in *Tbr*. Molecular models of the two *Buxus*-alkaloids were therefore  
257 prepared and aligned with compound **3** in the same manner as described above. Their anti-*Tbr*  
258 activity was then predicted with the CoMFA model. The predicted pIC<sub>50</sub> values (6.15 and 6.46 for **18**  
259 and **19**, respectively) were in reasonable agreement with the experimental values (5.43 and 6.79, i.e.  
260 **18** was predicted much less active than **19**) which is also reflected in their positions in the scatter plots  
261 of Figure 2). Furthermore, their interactions with the CoMFA contour maps of the QSAR model for  
262 anti-*Tbr* activity match the SAR described above very well (Figures 6B, C). In particular, the bulky  
263 tigloyl ester group of **18** explains its low antitrypanosomal activity very well since it collides with the  
264 sterically unfavorable white contour in region (c) of the model (Figure 6C) while there is no such  
265 unfavourable clash in case of the strong trypanocide **19**, which shows similar properties as the more  
266 active *Holarrhena* alkaloids. These results thus strongly support the hypothesis that the *Buxus*  
267 alkaloids indeed affect the trypanosomes by the same mechanism of action as the *Holarrhena*-  
268 aminosteroids. Further studies aimed at isolation and activity testing of a larger variety of *Buxus*  
269 alkaloids in order to allow establishment of a refined joint 3D-QSAR have, therefore, been initiated.

270  
271  
272  
273  
274



275



276

277 **Figure 6.** A: Structures of two *Buxus sempervirens* alkaloids, **18** and **19**. B: 3D-models of **18** and **19**  
278 superposed with **3**, C: O-tigloylcyclovirobuxein B (**18**) and C: cyclovirobuxein B (**19**). B-D show the  
279 contour maps of the CoMFA model for anti-*Tbr* activity (compare Figure 3) and the sterically  
280 unfavorable interaction of **18** explaining its low activity is marked.

281

282

### 283 3. Materials and Methods

#### 284 3.1. Data set

285 A set of 17 steroid alkaloids, all isolated from *H. africana* [1], were used for the 3D-QSAR  
286 computational studies of trypanocidal (*Tbr*) activities. The data set was divided into training (12  
287 compounds) and test (5 compounds) sets based on random selection. Sixteen of the alkaloids, used  
288 also for 3D-QSAR of L6 cytotoxic activities, were also divided into training (12 compounds) and test  
289 (4 compounds) sets. For the 3D QSAR, the molar (M) IC<sub>50</sub> values for *T. brucei rhodesiense* and L6  
290 cytotoxicity were converted to pIC<sub>50</sub> (Table 2), which is the negative decadic logarithmic value of the  
291 IC<sub>50</sub> (-log<sub>10</sub>IC<sub>50</sub>) as the target (dependent variable). The structures and IC<sub>50</sub> values of all the compounds  
292 used as both training and test sets are shown in Table 1.

#### 293 3.2. Building of molecular models

294 All 3D structures were built from fragments in the molecular operating environment, MOE [14].  
295 Initial three-dimensional molecular models of all compounds were energy minimized with the  
296 MMFF94x force field and then submitted to a low-mode dynamic (LMD) conformational search using  
297 default settings of MOE. The resulting conformers (within an energy window of 3 kcal/mol from the  
298 global minimum) were energy minimized using the semi-empirical Austin Model 1 (AM1)  
299 Hamiltonian (MOPAC module of MOE) and the conformers with lowest AM1 energy were used for  
300 QSAR studies.

#### 301 3.3. Alignment procedure

302 All structures were aligned with the lowest energy conformer of compound 3 which showed the  
303 highest antitrypanosomal (IC<sub>50</sub>, *Tbr* 0.075 μM) and cytotoxic (IC<sub>50</sub> L6 2.48 μM) activities. The selected  
304 atoms of the steroid skeleton (carbons 3, 5, 6, 9, 13, 14, and 17, see Figure 1) were used as matching  
305 points for the superposition. All the superposed molecules and their corresponding pIC<sub>50</sub> values were  
306 then converted to *Open3DQSAR*-compatible format for 3D-QSAR studies.

307

#### 308 3.4. Comparative Molecular Field Analysis (CoMFA)

##### 309 3.4.1. Data pretreatment

310 The CoMFA study was performed with the open-sourced software *Open3DQSAR* [15]. The  
311 aligned ligand assembly was automatically enclosed in a grid box exceeding the largest molecule by  
312 5 Å in each direction and a 1 Å mesh step size was chosen for the molecular field grids. Steric  
313 (Lennard-Jones potential) and electrostatic (Coulombic potential) molecular interaction fields (MIFs)  
314 were computed with *Open3DQSAR* using MMFF94 van der Waals parameters and charges. The steric  
315 interaction field was computed using an sp<sup>3</sup> carbon atom as probe while the electrostatic interaction  
316 field was computed using a volume-less probe with +1 charge. Training set MIF data were pre-filtered  
317 by setting an energy (Van der Waals and electrostatics) cutoff at ±30 kcal/mol; variables having a  
318 standard deviation below 2.0 were discarded to minimize noise and accelerate the regression analysis  
319 [18]. Furthermore, block unscaled weighting was applied to both steric and electrostatic fields to give  
320 them the same importance in the PLS model.

##### 321 3.4.2. PLS regression and model validation

322 The regression analysis of CoMFA field energies was performed using partial least squares (PLS)  
323 regression to correlate the descriptors (i.e. the MIF energies) with the pIC<sub>50</sub> data by extracting five  
324 latent variables (PLS components, PCs). In order to improve the model, smart region definition was  
325 performed on the aligned molecules to reduce the dependency from grid-to-molecule reciprocal  
326 orientations. The internal validation was performed for the 5 PCs by leave-one-out (LOO) cross-

327 validation method and the performance expressed as the coefficient of determination  $Q^2$  for the  
328 correlation between experimental and predicted  $pIC_{50}$  data of the training set.

329 The external predictive power of the developed CoMFA model was assessed by predicting the  
330 activities of the test set molecules, which were excluded during model development. The structural  
331 preparation of test set molecules as well as alignment and MIF calculation was the same as for the  
332 training set molecules. The activity of the test set was predicted by using the model derived from the  
333 training set where the predictive power of each PLS is expressed as the coefficient of determination  
334  $P^2$  for the correlation between experimental and predicted  $pIC_{50}$  data of the test set. The number of  
335 significant PLS components for each model was selected on the basis of the increase in  $Q^2$  observed  
336 when adding a further constituent (see data in Table S1, Supplementary Material). In case of the  
337 model for anti-*Tbr* activity, 3 PCs were considered significant while in case of the model for L6  
338 cytotoxicity, 2 PCs were used.

### 339 3.5. Contour mapping of steric and electrostatic fields

340 The visualization of the results of the CoMFA model as 3D contour maps was performed using  
341 the ReadMOEGRID module of MOE. The contours based on the regression coefficients exported from  
342 *Open3DQSAR* were plotted with interest on where variations of steric and electrostatic properties in  
343 the structural features of the different molecules of the training set leads to the most significant  
344 increase or decrease in antitrypanosomal and cytotoxic activities. The positive and negative  
345 influences of steric interaction on activity were represented by green and white contours,  
346 respectively, while those of electrostatic interactions were denoted by blue and red contours,  
347 respectively.

### 348 3.6. Isolation, characterization and biological testing of cyclovirobuxein B (19)

349 The extraction, fractionation and biological evaluation of aerial parts of *Buxus sempervirens* L. as  
350 well as the isolation, analytical and structural characterization and bioactivity testing of *O*-  
351 tigloylcyclovirobuxein B (**18**) has been described in detail in our previous report [16]. Fraction E9  
352 obtained by spiral coil countercurrent chromatography (spCCC) displayed promising anti-*Tbr*  
353 activity [17]. Compound **19** (cyclovirobuxein B) was detected by UHPLC/+ESI QqTOF MS as a  
354 constituent in E9 and preceding spCCC fractions (157-180 and E1-E8). It was then isolated from the  
355 pooled fractions by fast centrifugal partition chromatography (FCPC) on a Kromaton (Villejuif Cedex,  
356 France) monoaxial FCPC device with a 200 mL rotor containing 1000 metal chambers using n-hexane  
357 (upper phase): acetonitrile:dichloromethane (lower phase) 10:7:3 as solvent system in ascending  
358 mode (i.e. upper phase=mobile phase) at 1300 rotations/min and a flow rate of 2.5 mL. The separation  
359 of 55 mg of the mentioned pooled fractions led to 42 eluates (25 mL each). Compound **19** (5.3 mg)  
360 was the single constituent of eluate 23. It was characterized by UHPLC/+ESI-QqTOF-MS/MS and  
361 NMR spectroscopy ( $^1H$ ,  $^{13}C$ , COSY, HSQC, HMBC). For spectroscopic methods see [16, 17]. The  
362 compound was found identical with cyclovirobuxein B which had previously been described as  
363 constituent of *B. sempervirens*, along with the tiglitate **18** [19, 20]. The bioassays for antitrypanosomal  
364 and antiplasmodial activity were the same as described in [1, 16, 17].

365 *Cyclovirobuxein B (19)* +ESI-QqTOF-MS (m/z) 415.3760 [M+H]<sup>+</sup>, 209.1956 [M+2H]<sup>2+</sup> (calcd. for  
366  $C_{27}H_{47}N_2O^+$ : 415.3683; for  $C_{27}H_{48}N_2O_2^{2+}$ : 208.1878);  $^1H$ - and  $^{13}C$ -NMR data see Supplementary  
367 Materials, Table S4. Mass- and 1D and 2D NMR spectra are shown in Figures S1–S5, Supplementary  
368 Materials.

369 **Supplementary Materials:** Detailed Statistics for CoMFA models and numerical values for model  
370 predictions (Tables S1-S3), NMR data of compound **19** (Table S4) Mass and NMR spectra of  
371 compound **19** (Figures S1-S5). All molecular models and CoMFA contour maps are available from  
372 the corresponding author on request.

373

## 374 5. Conclusions

375 The CoMFA models obtained in this study for antitrypanosomal activity against *Tbr* and for  
376 cytotoxicity of *Holarrhena* steroid alkaloids were of good statistical quality and interpretability. They  
377 explain the main structure-activity relationships within the set of investigated natural products and  
378 also provide a rationale for the considerable selectivity of several compounds against *Tbr*.  
379 Furthermore, they may allow predictions of activity and selectivity for untested steroid alkaloids in  
380 order to select further promising natural candidates and also could provide a good starting point for  
381 the rational design of (semi)synthetic analogues.

382 **Acknowledgments:** Charles Nnadi acknowledges the Federal Government of Nigeria and the Tertiary  
383 Education Trust Fund (TETFund) for his doctoral fellowship in Münster. This work is part of the activities of  
384 ResNetNPND (<http://www.ResNetNPND.org/>).

385 **Author Contributions:** TJS and NJN conceived and initiated the study on *H. africana*. TJS designed and  
386 supervised all experiments; CON isolated and identified the *Holarrhena* alkaloids used in model development,  
387 performed the QSAR study, analyzed the data and drafted the manuscript. JBA isolated and identified the  
388 compounds **18** and **19** from *Buxus sempervirens*. TJS wrote the final version of the manuscript.

389 **Conflicts of Interest:** The authors declare no conflict of interest.

## 390 References

- 391 1. Nnadi, C. O.; Nwodo, N. J.; Kaiser, M.; Brun, R.; Schmidt, T. J. Steroid Alkaloids from *Holarrhena africana*  
392 with Strong Activity against *Trypanosoma brucei rhodesiense*. *Molecules*, **2017**, *22*(7), 1129.
- 393 2. Babu, S.; Sohn, H.; Madhavan, T. Computational analysis of CRTh2 receptor antagonist: A ligand-based  
394 CoMFA and CoMSIA approach. *Comput. Biol. Chem.*, **2015**, *56*, 109-121.
- 395 3. Lill, M. A. Multi-dimensional QSAR in drug discovery. *Drug Discovery Today*, **2007**, *12*(23-24), 1013-1017.
- 396 4. Wellsow, J.; Machulla, H. J.; Kovar, K. A. 3D QSAR of serotonin transporter ligands: CoMFA and CoMSIA  
397 studies. *Quant. Struct. Act. Relat.*, **2002**, *21*, 577-589.
- 398 5. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effects of  
399 shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, **1988**, *110*(18), 5959-5967.
- 400 6. Cook, C. E.; Wani, M. C.; Lee, Y. W.; Fail, P. A.; Petrow, V. Reversal of activity profile in analogs of the  
401 antiprogesterin RU 486: effect of a 16 $\alpha$ -substituent on progestational (agonist) activity. *Life Sci.*, **1993**, *52*(2),  
402 155-162.
- 403 7. Polanski, J.; Gieleciak, R.; Bak, A. Probability issues in molecular design: Predictive and modeling ability  
404 in 3D-QSAR schemes. *Comb. Chem. High Throughput Screening*, **2004**, *7*(8), 793-807.
- 405 8. Chen, H.; Li, Q.; Yao, X.; Fan, B.; Yuan, S.; Panaye, A.; Doucet, J. P. 3D-QSAR and Docking Study of the  
406 Binding Mode of Steroids to Progesterone Receptor in Active Site. *Mol. Inf.*, **2003**, *22*(6), 604-613.
- 407 9. Kapou, A.; Benetis, N. P.; Durdagi, S.; Nikolaropoulos, S.; Mavromoustakos, T. 3D QSAR/CoMFA and  
408 CoMSIA studies on antileukemic steroidal esters coupled with conformationally flexible nitrogen  
409 mustards. *J. Chem. Inf. Model.*, **2008**, *48*(11), 2254-2264.
- 410 10. Loughney, D. A.; Schwender, C. F. A comparison of progestin and androgen receptor binding using the  
411 CoMFA technique. *Journal of Computer-Aided Molecular Design*, **1992**, *6*(6), 569-581.
- 412 11. Norinder, U. Experimental design based 3D QSAR analysis of steroid-protein interactions: application to  
413 human CBG complexes. *Journal of Computer-Aided Molecular Design*, **1990**, *4*(4), 381-389.
- 414 12. Norinder, U. 3-D QSAR analysis of steroid/protein interactions: The use of difference maps. *Journal of*  
415 *Computer-Aided Molecular Design*, **1991**, *5*(5), 419-426.
- 416 13. Norinder, U. 3D-QSAR investigation of the tripos benchmark steroids and some protein-tyrosine kinase  
417 inhibitors of styrene type using the TDQ approach. *J. Chemom.*, **1996**, *10*(5-6), 533-545.
- 418 14. Chemical Computing Group, Molecular Operating Environment (MOE) rel. 2011.10. Chemical Computing  
419 Group Inc, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, **2014**.
- 420 15. Tosco, P.; Balle, T. Open3DQSAR: a new open-source software aimed at high throughput chemometric  
421 analysis of molecular interaction fields. *J. Mol. Model.*, **2011**, *17*(1), 201-208.
- 422 16. Althaus, J. B.; Jerz, G.; Winterhalter, P.; Kaiser, M.; Brun, R.; Schmidt, T. J. Antiprotozoal activity of *Buxus*  
423 *sempervirens* and activity-guided isolation of *O*-tigloylcyclovirobuxeine-B as the main constituent active  
424 against *Plasmodium falciparum*. *Molecules*, **2014**, *19*(5), 6184-6201.

- 425 17. Althaus, J. B. Natural products with antiprotozoal activity: Alkamides isolated from Asteraceae and  
426 triterpene alkaloids from *Buxus sempervirens* L. Doctoral Thesis, IPBP, University of Münster, Germany,  
427 **2015**.
- 428 18. Martin, Y. C.; Willett, P.; Heller, S. R. *Designing bioactive molecules: Three-dimensional techniques and*  
429 *applications (Computer applications in Chemistry collection)*. 1<sup>st</sup> ed.; American Chemical Society. Washington  
430 D.C, USA, **1998**, pp. 181-194; ISBN: 9780841234901.
- 431 19. Kupchan, S.M.; Ohta, G. *Buxus* alkaloids. X. The Isolation and Constitution of Cyclovirobuxine-B. *J. Org.*  
432 *Chem.* **1966**, *31*, 608–610.
- 433 20. Kupchan, S. M.; Kennedy, R. M.; Schleigh, W. R.; Ohta, G. *Buxus* alkaloids- XII. Benzamide alkaloids from  
434 *Buxus sempervirens* L. *Tetrahedron* **1967**, *23*, 4563–4586.  
435