# Supplementary information

## Identification of steroid-like natural products as potent antiplasmodial agents by 2D and 3D similarity-based virtual screening

Elumalai Pavadai[a#*], Gurminder Kaur[a#], Sergio Wittlin[b,c] and Kelly Chibale[a, d, e*]

[a]Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa.
[b]Swiss Tropical and Public Health Institute, Socinstrasse 57, 4051 Basel, Switzerland
[c]University of Basel, 4002 Basel, Switzerland
[d]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Rondebosch 7701, South Africa.
[e]South African Medical Research Council Drug Discovery and Development Research Unit, University of Cape Town, Rondebosch 7701, South Africa.

[#]Equally contributed

*Corresponding authors:
Phone: +91 44 2246 0042; E-mail: elumalai.pavadai.a@gmail.com
Phone: +27 21 650 2553; Fax: +27 21 650 5195; E-mail: Kelly.Chibale@uct.ac.za

### Supplementary information S1. 2D fingerprint-based screening

***Database preparation***: The in-house database containing 708 compounds was prepared and duplicates were removed using the *Prepare Ligands* protocol of Discovery Studio (DS4) [Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, San Diego, USA], returning 602 compounds that were used for the fingerprint-based screening.

***2D Fingerprints***: Fingerprints were generated employing DS4 which contains 13 fingerprint methods: one MDL PublicKeys and 12 extended connectivity fingerprints. Extended-connectivity fingerprints are circular topological fingerprints and their generation process systematically records the neighbourhood of each non-hydrogen atom into multiple circular layers up to a given bond diameter.[1] The extended connectivity fingerprints are further sub grouped into three ECFPs (ECFP_2, ECFP_4 and ECFP_6), three FCFPs

(FCFP_2, FCFP_4, and FCFP_6), three ECFCs (ECFC_2, ECFC_4, ECFC_6) and three FCFCs (FCFC_2, FCFC_4, FCFC_6). ECFPs (atom-types) and FCFPs (functional class) consider only the presence or absence of a feature while ECFCs and FCFCs also consider as many times as a feature has appeared, the numeric 2,4 and 6 represent bond diameter length or maximum number of bonds considered. Of thirteen fingerprints methods available in DS4, ECFC_4, FCFP_2 and FCFC_4 were used for virtual screening with the prepared database. The similarity between the fusidic acid and the database compounds was assessed using the Tanimoto coefficient (Tc) [Equation (1)].

$$\mathrm{Tc(A,\ B)} = \frac{c}{a + b - c} \qquad (1)$$

Where *a* and *b* are the number of features present in compounds A and B respectively, c is the number of features shared by A and B. Tc quantifies the fraction of features common to A and B to the total number of features of A and B, where the c term in the denominator corrects for double counting of the features. The value of Tc lies between 0 and 1 where 0 stands for 'not similar' and 1 stands for 'highly similar'.

**Supplementary information S2. 3D shape-based screening**

***Database preparation*:** The database compounds and query were optimized using the *LigPrep* module of the Schrödinger suite with OPLS_2005 force fields. This module generates possible 3D conformations for each ligand with various ionization states at pH 7.0 ± 2.0, tautomers, stereochemistries and ring conformations. All the generated conformations were saved as maestro files and used for screening. Subsequently, a conformational search was then performed using *ConfGen* in the fast search mode and different energy conformations of the query and database molecules were generated. Only the lowest energy conformation of the query was allowed to screen the database.

***3D shape-based screening***: Shape-based screening was performed by Phase-shape module of Maestro 8.0, Schrodinger (Small-Molecule Drug Discovery Suite Phase, Schrödinger, LLC, New York, USA). Phase-shape considers volume overlapping of all the conformations of the molecule with the shape query and returns an aligned structure which shows the best overlap with the shape query. Four different phase-shape methods are available in this software: Shape_none, Shape_MMod, Shape_element, Shape_Pharm. Shape_none calculates

volume overlaps only and does not discriminate between the different atoms. Shape_MMod calculates volume overlaps only between atoms that have the same MacroModel atom type. Shape_element and Shape_pharm also consider atom-related information in their calculations. Shape_element calculates volume overlaps only between atoms of the same element. Shape_pharm calculates volume overlaps between atoms that have the same pharmacophore type (acceptor, donor etc.).

Of four shape methods available in *Shape screening* module of the Schrödinger suite, Shape_element and Shape_pharm were employed for virtual screening with the prepared database.The shape similarity coefficient was used to assess the similarity between the query molecule and database compounds.Like Tanimoto, shape similarity coefficient lies between 0 and 1 where 0 stands for 'not similar' and 1 stands for 'highly similar'.

**Supplementary information S3. *In vitro Pf*NF54 Assay**

Compounds were screened against the sensitive strain of P. falciparum (NF54) in vitro using the modified [$^3$H]hypoxanthine incorporation assay (please refer to: In vitro and in vivo interaction of synthetic peroxide RBx11160 (OZ277) with piperaquine in Plasmodium models. Snyder C, Chollet J, Santo-Tomas J, Scheurer C, Wittlin S., Exp Parasitol. 2007 Mar;115(3):296-300).

**Supplementary information S4. Cytotoxicity Assay**

Compounds were screened for *in vitro* cytotoxicity against a mammalian cell-line, Chinese Hamster Ovarian (CHO) using the 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazoliumbromide (MTT) assay. The MTT assay is used as a colorimetric assay for cellular growth and survival, and compares well with other available assays.[4, 5] The tetrazolium salt MTT was used to measure all growth and chemo sensitivity. Compounds were tested in triplicate on one occasion.

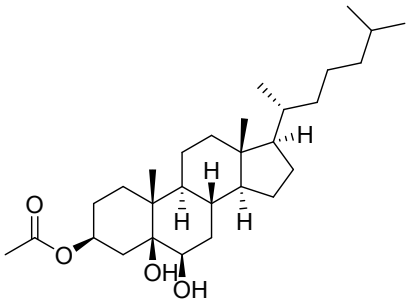
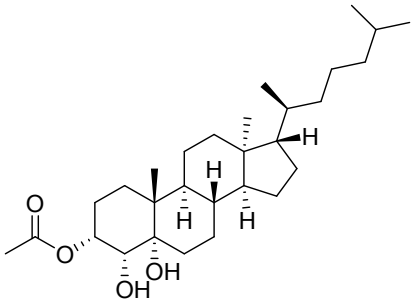
The test compounds were prepared to a 20 mg/ml stock solution in 100% DMSO and stored at -20ºC until use. Dilutions were prepared on the day of the experiment. Emetine was used as the reference drug in all experiments. The initial concentration of emetine was 100 µg/ml, which was serially diluted in complete medium with 10-fold dilutions to give 6 concentrations, the lowest being 0.001 µg/ml. The same dilution technique was applied to the all test compounds. The 50% inhibitory concentration ($IC_{50}$) values were obtained from full

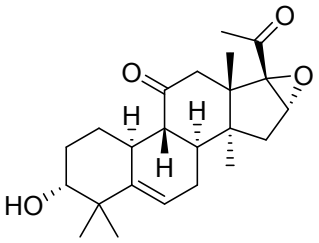
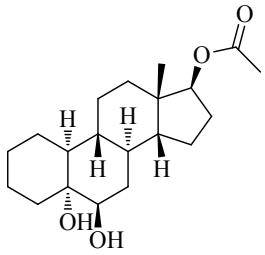
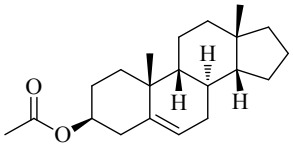
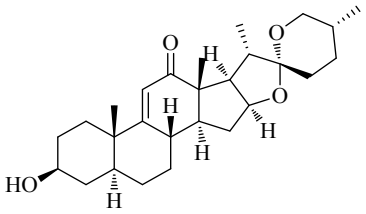
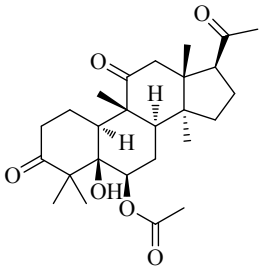
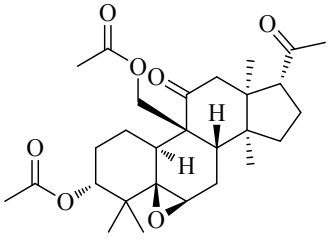
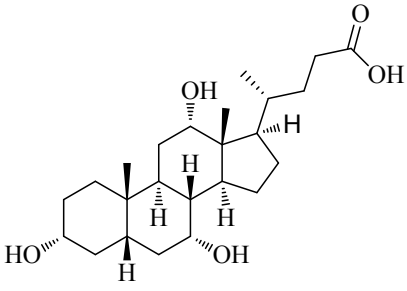
dose-response curves, using a non-lineardose-response curve fitting analysis via GraphPad Prism v.4 software.

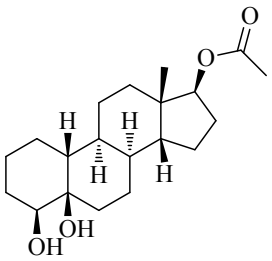
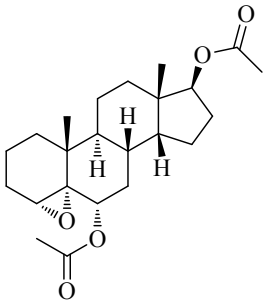
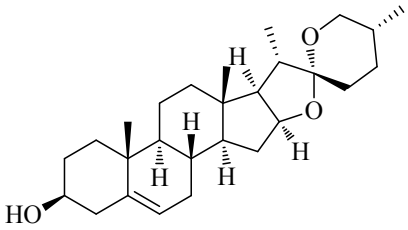
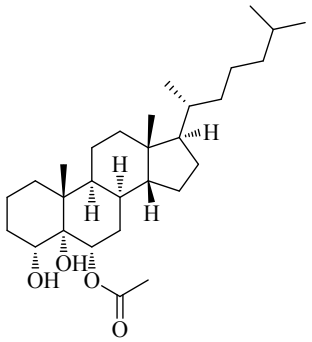
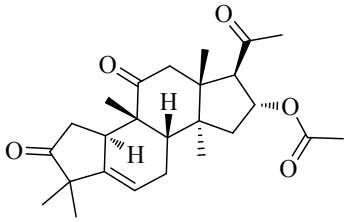
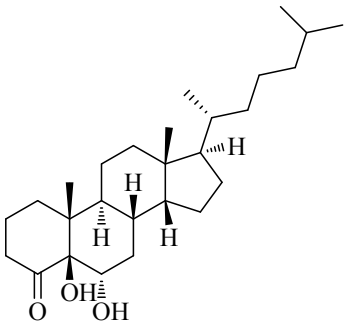
## 5. References

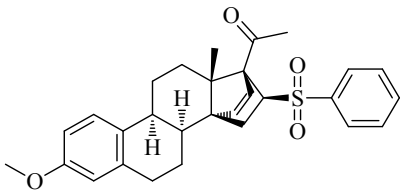
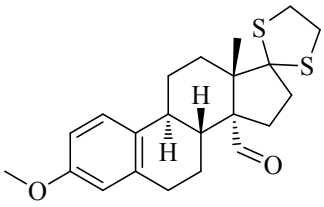
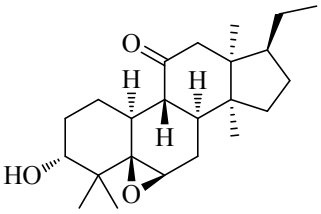

1. D. Rogers,M. Hahn,*J. Chem. Inf. Model.*,2010, **50**, 742-754.
2. W. Trager, J. B. Jensen, *Science*,1976, **193**, 673-675.
3. M.T.Makler, J. M. Ries, J. A. Williams,J. E. Bancroft, R. C. Piper,B. L. Gibbins, D.J.Hinrichs,*Am. J. Trop. Med.Hyg*.,1993, **48**, 739-741.
4. L. V. Rubinstein,R. H. Shoemaker, K. D. Paull, R. M.Simon, S. Tosini, P. Skehan, D. aScudiero, A. Monks, M. R. Boyd, *J. Natl. cancer Inst.*,1990, **82**, 1113-1118.
5. T. Mosmann, *J. Immunol. Methods*,1983, **65**, 55-63.


**Supplementary Table S1**.Chemical structures of inactive compounds along with their log P values and similarity score.

| Compd. ID | Structure | *Pf*NF54 $IC_{50}$ (µM)* | CHO $IC_{50}$ (µM) | log P | Similarity score |
|---|---|---|---|---|---|
| **1** | Fusidic Acid | 59 | ND | 4.45 | ND |
| 11[a] |  | >18.64 | ND | 5.37 | 0.68 |
| 12[a] |  | >21.42 | ND | 5.38 | 0.68 |

| | | | | | |
|---|---|---|---|---|---|
| 13[a] |  | >24 | ND | 2.82 | 0.7 |
| **14**[a] |  | >25 | ND | 3.298 | 0.684211 |
| **15**[a] |  | >25 | ND | 5.446 | 0.666667 |
| **16**[a] |  | >25 | ND | 4.012 | 0.666667 |
| **17**[b] |  | >25 | ND | 2.976 | 0.607029 |
| **18**[b] |  | >25 | ND | 2.992 | 0.636646 |
| **19**[a] |  | >25 | ND | 3.308 | 0.722222 |

| | | | | | |
|---|---|---|---|---|---|
| **20**[a] |  | >25 | ND | 3.313 | 0.684211 |
| **21**[b] |  | >25 | ND | 3.547 | 0.60596 |
| **22**[e] |  | >25 | ND | 4.572 | 0.394735 |
| **23**[a] |  | >25 | ND | 5.372 | 0.684211 |
| **24**[e] |  | >25 | ND | 2.981 | 0.412334 |
| **25**[e] |  | >25 | ND | 5.207 | 0.408731 |

| | | | | | |
|---|---|---|---|---|---|
| **26**[d] |  | >25 | ND | 4.241 | 0.689822 |
| **27**[e] |  | >25 | ND | 3.489 | 0.570233 |
| **28**[e] |  | >25 | ND | 3.575 | 0.427041 |

ND: Not Determined. . [*]Mean from at least 2 independent experiments with chloroquine-sensitive strain (NF54) of *P. falciparum*. [a]Compound screened by FCFP_2. [b]Compound screened by ECFC_4. [c]Compound screened by FCFC_4. [d]Compound screened by Shape_ele. [e]Compound screened by Shape_pharm.