

# AfroDb: A Select Highly Potent and Diverse Natural Product Library from African Medicinal Plants

Fidele Ntie-Kang<sup>1,2,3\*</sup>, Denis Zofou<sup>4</sup>, Smith B. Babiaka<sup>5</sup>, Rolande Meudom<sup>5</sup>, Michael Scharfe<sup>3</sup>, Lydia L. Lifongo<sup>1</sup>, James A. Mbah<sup>5</sup>, Luc Meva'a Mbaze<sup>6</sup>, Wolfgang Sippl<sup>3</sup>, Simon M. N. Efange<sup>5</sup>

**1** Chemical and Bioactivity Information Centre, Department of Chemistry, Faculty of Science, University of Buea, Buea, Cameroon, **2** Center Atomic Molecular Physics, Optics and Quantum, Faculty of Science, University of Douala, Douala, Cameroon, **3** Department of Pharmaceutical Sciences, Martin-Luther University of Halle-Wittenberg, Halle (Saale), Germany, **4** Biotechnology Unit, Department of Biochemistry and Molecular Biology, Faculty of Science, University of Buea, Buea, Cameroon, **5** Department of Chemistry, Faculty of Science, University of Buea, Buea, Cameroon, **6** Department of Chemistry, Faculty of Science, University of Douala, Douala, Cameroon

## Abstract

Computer-aided drug design (CADD) often involves virtual screening (VS) of large compound datasets and the availability of such is vital for drug discovery protocols. We assess the bioactivity and “drug-likeness” of a relatively small but structurally diverse dataset (containing >1,000 compounds) from African medicinal plants, which have been tested and proven a wide range of biological activities. The geographical regions of collection of the medicinal plants cover the entire continent of Africa, based on data from literature sources and information from traditional healers. For each isolated compound, the three dimensional (3D) structure has been used to calculate physico-chemical properties used in the prediction of oral bioavailability on the basis of Lipinski’s “Rule of Five”. A comparative analysis has been carried out with the “drug-like”, “lead-like”, and “fragment-like” subsets, as well as with the Dictionary of Natural Products. A diversity analysis has been carried out in comparison with the ChemBridge diverse database. Furthermore, descriptors related to absorption, distribution, metabolism, excretion and toxicity (ADMET) have been used to predict the pharmacokinetic profile of the compounds within the dataset. Our results prove that drug discovery, beginning with natural products from the African flora, could be highly promising. The 3D structures are available and could be useful for virtual screening and natural product lead generation programs.

**Citation:** Ntie-Kang F, Zofou D, Babiaka SB, Meudom R, Scharfe M, et al. (2013) AfroDb: A Select Highly Potent and Diverse Natural Product Library from African Medicinal Plants. PLoS ONE 8(10): e78085. doi:10.1371/journal.pone.0078085

**Editor:** Joseph J. Barchi, National Cancer Institute at Frederick, United States of America

**Received:** June 13, 2013; **Accepted:** September 6, 2013; **Published:** October 30, 2013

**Copyright:** © 2013 Ntie-Kang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Financial support is acknowledged from the Chemical and Bioactivity Information Centre, University of Buea, through a Lhasa grant and also from the German Academic Exchange Service to FNK for his stay in Halle, Germany for part of his PhD, as well as from the ICTP through the OEA-AC-071 program. DZ acknowledges an International Foundation for Science, Stockholm, Sweden/Organisation for the Prohibition of Chemical Weapons, The Hague, The Netherlands (IFS/OPCW) grant (N°F/5122-1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The academic license for this work was generously offered by Schrodinger Inc. This does not alter the authors’ adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: ntiekfidele@gmail.com

## Introduction

Drug design and discovery efforts have often resorted to natural sources for hit/lead compound identification [1–5]. This is because nature is an enormous source for structurally diverse chemical scaffolds from which drugs can be isolated and/or synthesized [6–8]. Moreover, natural products (NPs) are unique in that they are often rich in stereogenic centres and cover segments of chemical space which are typically not occupied by a majority of synthetic molecules and drugs [9–11]. In addition, it can be verified that the African flora has a huge potential and remains an interesting reservoir for new drugs targeting a variety of diseases [12].

Modern drug discovery efforts also incorporate computer-aided approaches like ligand docking, pharmacophore searching, neural networking and binding free energy calculations of potential drugs towards a target receptor. The rationale behind such *in silico* methods has been to simulate the interaction between a potential drug molecule and its receptor or binding site (often the drug target) using 3D computer models [13–14]. Hence, the use of computer modeling in drug discovery otherwise known as

computer-aided drug design (CADD) requires a compound library containing 3D structures of potential leads, which need to be screened *in silico*, with the view of identifying hit compounds. If this effort is successful, then the identified hits could be confirmed as active compounds using screening assays. Such a procedure considerably cuts down the cost of drug discovery and development [15].

At the moment, efforts towards drug discovery from the African flora have been limited to random screening of extracts and/or bioassay guided fractionation of extracts from medicinal plant materials, based on information obtained from the ethnobotanical uses of the plants [8]. However, these efforts remain below expectations and their applicability as well as their impact, are not felt at the level of local populations. Moreover, most screening efforts are limited to crude extracts and/or *in vitro* screening, with only little work on clinical development of the identified active molecules. The inclusion of *in silico* methods of drug discovery into the scene would likely foster efforts towards lead optimization and facilitates the entry of most interesting compounds into clinical trials. However, this process requires the development of databases of 3D structures of compounds, which have been isolated from

medicinal plants in Africa. It should be mentioned that some of the chemical structures are available in published articles in internationally recognized peer-reviewed journals. However, locally published data (in MSc and PhD theses as well as in local or national journals) may not be readily available to the wider scientific community. Moreover, the absence of 3D structures heavily hampers *in silico* virtual screening. In our continuous efforts to build virtual natural product libraries for compounds isolated from the African flora [16–18], it has become necessary to identify selected highly potent compounds from the entire continent, generate their 3D structures and make them available for use in virtual screening campaigns.

With the exponential increase in computer power [19], it has become possible today to carry out successful virtual screening on huge databases like the ZINC library (currently 19,607,982 purchasable compounds) [20–21] in a matter of weeks [19,22]. However, such computer power is currently absent in most research laboratories on the African continent, thoughtless of guaranteeing continuous electrical power supplies for long periods of time. This has therefore necessitated the development of relatively small compound library (~1000 compounds), known to contain NPs from across the continent, which have recorded activities against a wide range of tropical diseases as well as diseases dominant in rich countries (like cancer and hypertension). We therefore present AfroDb, which is available in several file formats, and which could be highly useful in CADD efforts. An evaluation of the potential oral bioavailability has been carried out by Lipinski criteria [23] in comparison with Dictionary of Natural Products (DNP) [24]. Several parameters related to drug metabolism and pharmacokinetics (DMPK) are also computed. A diversity analysis has also been carried out in comparison with the ChemBridge Diverset library [25].

## Results and Discussion

### Origin and Description of Secondary Metabolites

The distribution of the compound collection by geographical region of plant origin is shown in Figure 1. This shows that a majority of the compounds with remarkable biological activities were derived from the Central Africa region (35%), followed by Southern Africa (23%) and East Africa (21%). The known biological activities include very specific descriptions (like inhibition or modulation of known drug targets, e.g., prolyl endopeptidase I inhibition, 11 $\beta$ -hydroxysteroid dehydrogenase inhibition,  $\alpha$ -glucosidase inhibition, enhancement of cAMP-regulated chloride conductance of cells expressing CFTRAF508, and snake venom phosphodiesterase I inhibition), while more unspecific classifications include anti-HIV, antisalmonellal, antimalarial, antileishmanial, antitubercular, antitrypanosomal, antitumour, vasodilator, vasorelaxant and hypertensive effects, estrogenic activities, and activity against *Onchocerca gutturosa*.

### Discussion of Lipinski's Oral Availability Criteria and Property Distribution

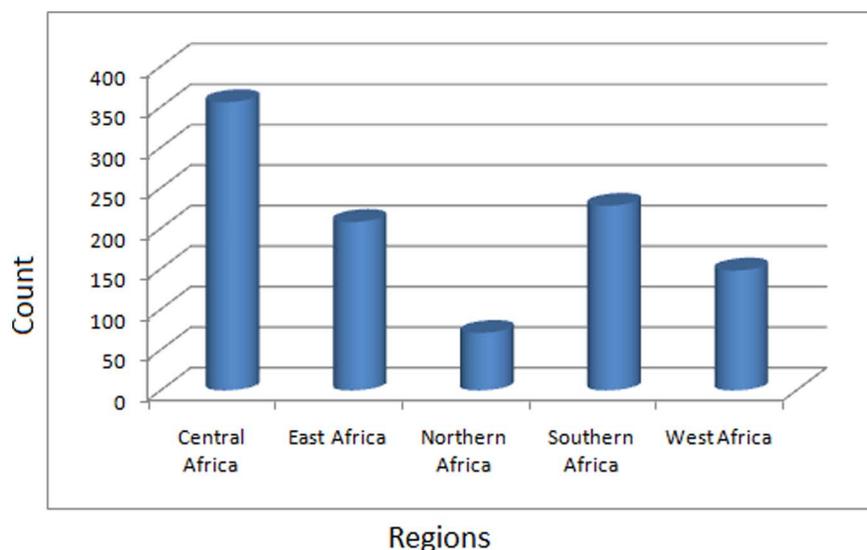
In modern drug discovery, the identification of lead compounds often involves the development of compound libraries with a high level of molecular diversity within the limits of significant “drug-like” properties [5,9]. Thus, Lipinski's criteria [23], generally referred to as the “rule of Five” (ro5), have been used in the evaluation of the likely oral availability of the compounds within the AfroDb database. In summary, Lipinski's ro5 defines a likely orally available molecule as one for which the molar weight (MW)  $\leq 500$  Daltons (Da), the logarithm of the octanol/water partition coefficient representing the lipophilicity factor ( $\log P$ )  $\leq 5$ , the

number of hydrogen bond acceptors (HBA)  $\leq 10$  and the number of hydrogen bond donors (HBD)  $\leq 5$ . A fifth rule dealing with the number of rotatable bonds (NRB) is often added to these, such that  $\text{NRB} \leq 5$ . The distributions of the compound MW,  $\log P$ , HBA and HBD were calculated and used to assess the likely oral availability of AfroDb, as shown in Figure 2. It is noteworthy that natural products exhibit a wide range of flexibility, from rigid conformationally constrained molecules to very flexible compounds. Thus, the number of rotatable bonds (NRB) within the AfroDb library was used as an additional criterion to test for the favourable drug metabolism and pharmacokinetics (DMPK) outcomes. It was observed that 57.8% of the compounds within AfroDb showed no Lipinski violations and 84.3% showed  $< 2$  violations (Figure 2A), while the peak of the distribution of the MW lies between 301 and 400 (Figure 2B). The  $\log P$  distribution showed a Gaussian shaped curve with a peak centred at 3.5  $\log P$  units (Figure 2C), while those of the HBA and HBD rose rapidly to peak values at 5 acceptors and 1 donor respectively. Moreover, both HBA and HBD graphs fell rapidly to maximum values of 27 acceptors and 14 donors (Figures 2D and E). The peak value for NRB was between 3 and 4 (Figure 2F), the graph also falling rapidly to a maximum value of 61 rotatable bonds (RBs). This last parameter indicates the high degree of flexibility of some of the NPs within the database. The MW distribution resembled those previously reported for other “drug-like” NP libraries in the literature [5,16,26], with only about 9% of MW  $> 500$  Da. In addition, only about 16% of the compounds had  $\log P$  values  $> 5$ , only about 7% having HBA  $> 10$  and only ~9% had HBD  $> 5$ . It was observed, however, that some of the compounds had  $\log P$  values as high as  $> 21$  units. The exceptionally large calculated  $\log P$  values ( $> 14$ ) observed in compounds 1 to 3 could be explained by complex partitioning in long chain aliphatic compounds, which could not be properly taken into account by the algorithm employed in the  $\log P$  prediction. The three long chain compounds with  $\log P > 14$  have been shown in Figure 3, with their plant sources and biological activities given in Table 1. The above arguments could be verified by the fact the scatter plots displaying the mutual relationship between the MW versus the other calculated parameters ( $\log P$ , HBA, HBD and NRB), Figure 4, show the highest densities of points within the Lipinski compliance regions (MW  $< 500$ ,  $-2 < \log P < 5$ , HBA  $< 10$ , HBD  $< 3$ ), and for which  $\text{NRB} < 5$ .

### Comparison with the Dictionary of Natural Products (DNP)

The distributions of the individual parameters for AfroDb and the DNP have been compared (Figure 5). These histograms show only data falling within the Lipinski compliance zones (MW  $< 500$ ,  $-2 \leq \log P < 5$ , HBA  $< 10$ , and HBD  $< 5$ ), the values being expressed as percentage counts of the respective databases. Our analyses showed an enhancement of the distributions of AfroDb over the DNP for Lipinski properties. As for the MW distribution histograms (Figure 5A), both curves show peaks at 301–400 Da, the AfroDb having  $> 19\%$  enhancement in MW for the region 301 to 500 Da. Below this range, the percentages were reduced for the AfroDb when compared to the DNP.

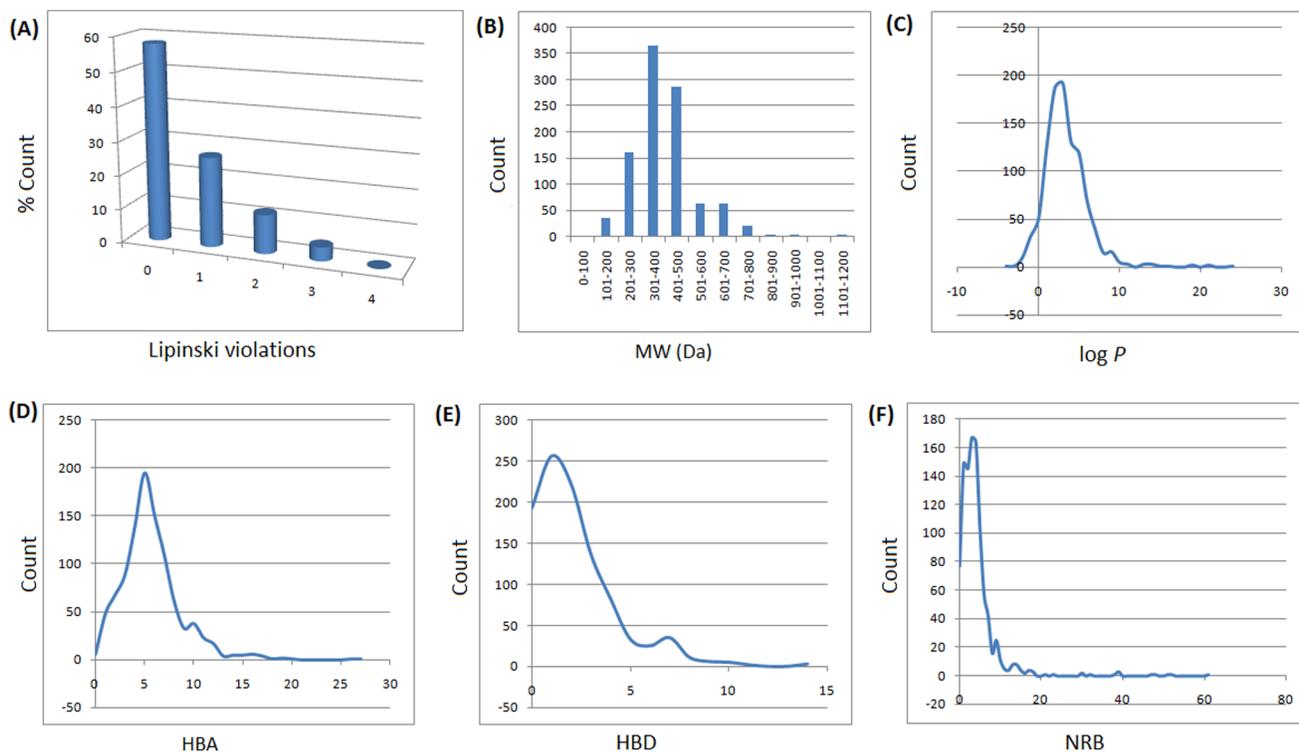
It is noteworthy that an enhancement of the MW profile is a desirable factor for a more “drug-like library”, according to Lipinski's criteria. The proportions of the two databases that satisfy Lipinski's MW property ( $< 500$  Da) were about 73% for the DNP, as compared to about 84% for AfroDb. This showed an enhancement of 11.2% for MW values between 301 and 500 Da of AfroDb over the DNP. The maximum values for the  $\log P$  distributions were respectively at 3.5 and 2.5  $\log P$  units for



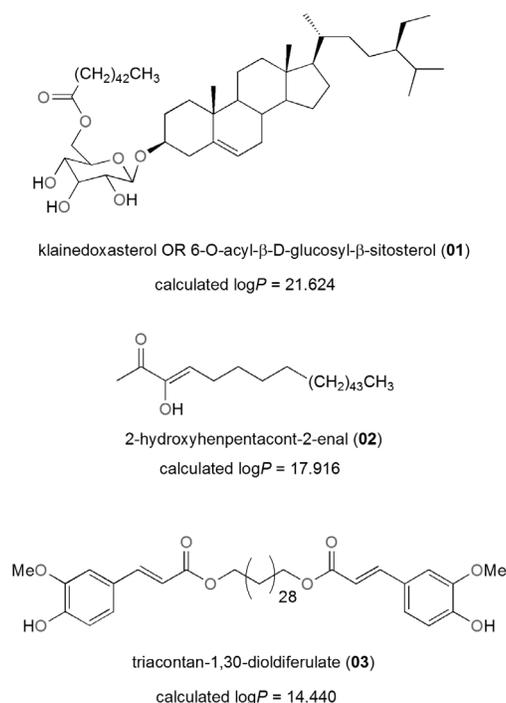
**Figure 1. Bar chart showing the distribution of the compounds within AfroDb by region of collection.**  
doi:10.1371/journal.pone.0078085.g001

AfroDb and the DNP (Figure 5B). In addition, a corresponding enhancement of 9.7% for log  $P$  values between 2 and 5 units was observed for AfroDb over the DNP. As per the HBA and HBD respectively (Figures 5C–D), AfroDb showed improvements of 17.1% for  $3 < \text{HBA} < 8$  and 10.8% for  $0 < \text{HBD} < 4$  over the DNP. The peak of the distribution for the HBA for the AfroDb is at 5 acceptors (19.3%) with a significant increase in 7 or 8

acceptors when compared to the DNP (Figure 5C). Similarly, the peak of the distribution for the HBD for the AfroDb is at 1 acceptor (25.4%) with a significant increase in 3 or 4 donors as compared to the DNP (Figure 5D). The overall summary of the four Lipinski parameters for the two datasets thus reveals that the AfroDb library is more “drug-like” than the DNP, indicating that the chances of finding “lead-like” molecules with improved



**Figure 2. Graph distribution of features that determine “drug-likeness”.** (A, B) Histogram of Lipinski violations as a percentage of the AfroDb data set and molar weight distribution, respectively. (C, D, E, F) Distribution curves of the log  $P$ , HBA, HBD and NRB, respectively, for the 1,008 compounds currently in AfroDb. For subfigure B, the x-axis label is the lower limit of binned data, e.g. 0 is equivalent to 0 to 100.  
doi:10.1371/journal.pone.0078085.g002



**Figure 3. 2D structures of the three compounds with log P values >14, included in AfroDb.**  
doi:10.1371/journal.pone.0078085.g003

DMPK properties within a library such as AfroDb are quite significant.

### Overall DMPK Compliance of the AfroDb Library

There are 24 molecular descriptors calculated by QikProp software, which are most relevant for the determination the #star parameter [30]. For a given parameter, 1 #star corresponds to the computed property of a molecule falling outside the range for 95% of known drugs. A plot of the #stars parameter (on *x*-axis) against the corresponding counts (on *y*-axis) in the AfroDb is shown within the same set of axes with those of the “drug-like”, “lead-like”, and “fragment-like” standard subsets, (Figure 6). The criteria for the respective standard subsets were defined as (MW <500; log *P* <5; HBD ≤5; HBA ≤10) [23], (150 ≤ MW ≤350; log *P* ≤4; HBD ≤3; HBA ≤6) [31–33] and (MW ≤250; −2 ≤ log *P* ≤3; HBD <3; HBA <6; NRB <3) [34]. It was observed that about 48% of the compounds within AfroDb showed #star = 0, while about 79% had #star ≤2. Among the 610 compounds of the “drug-like” subset, 73.4% had pharmacokinetic descriptors within the acceptable range for 95% of known drugs, while 98.9% showed #stars ≤2. The “lead-like” and “fragment-like” subsets were respectively 75.2% and 76.0% compliant for all of the 24 most relevant computed descriptors. The average values for 19 selected computed descriptors for all 4 compound libraries have been shown in Table 2. These values indicate a high probability of finding drug leads within the AfroDb compound library.

### Bioavailability Prediction

Two processes determine the bioavailability of a compound; absorption and liver first-pass metabolism [35]. The solubility and permeability of the compound, as well as interactions with transporters and metabolizing enzymes in the gut wall are factors responsible for absorption, while metabolism depends on the

functional group types present. The computed parameters used to assess oral absorption revolved around Jorgensen’s famous “Rule of Three” (ro3). According to the ro3, a compliance with all or some of the rules (log *S*<sub>wat</sub> > −5.7, BIP<sub>caco-2</sub> > 22 nm/s and # primary metabolites < 7) is indicative of likelihood to oral availability. Thus, the most important parameters often considered are the predicted aqueous solubility, log *S*<sub>wat</sub>, the conformation-independent predicted aqueous solubility, Clog *S*<sub>wat</sub>, the predicted qualitative human oral absorption, the predicted % human oral absorption. The solubility calculation procedure implemented depends on the similarity property space between the given molecule and its most similar analogue within the experimental training set used to develop the model implemented in QikProp, i.e., if the similarity is < 0.9, then the QikProp predicted value is taken. Otherwise, the predicted property, *P*<sub>pred</sub>, is adjusted such that:

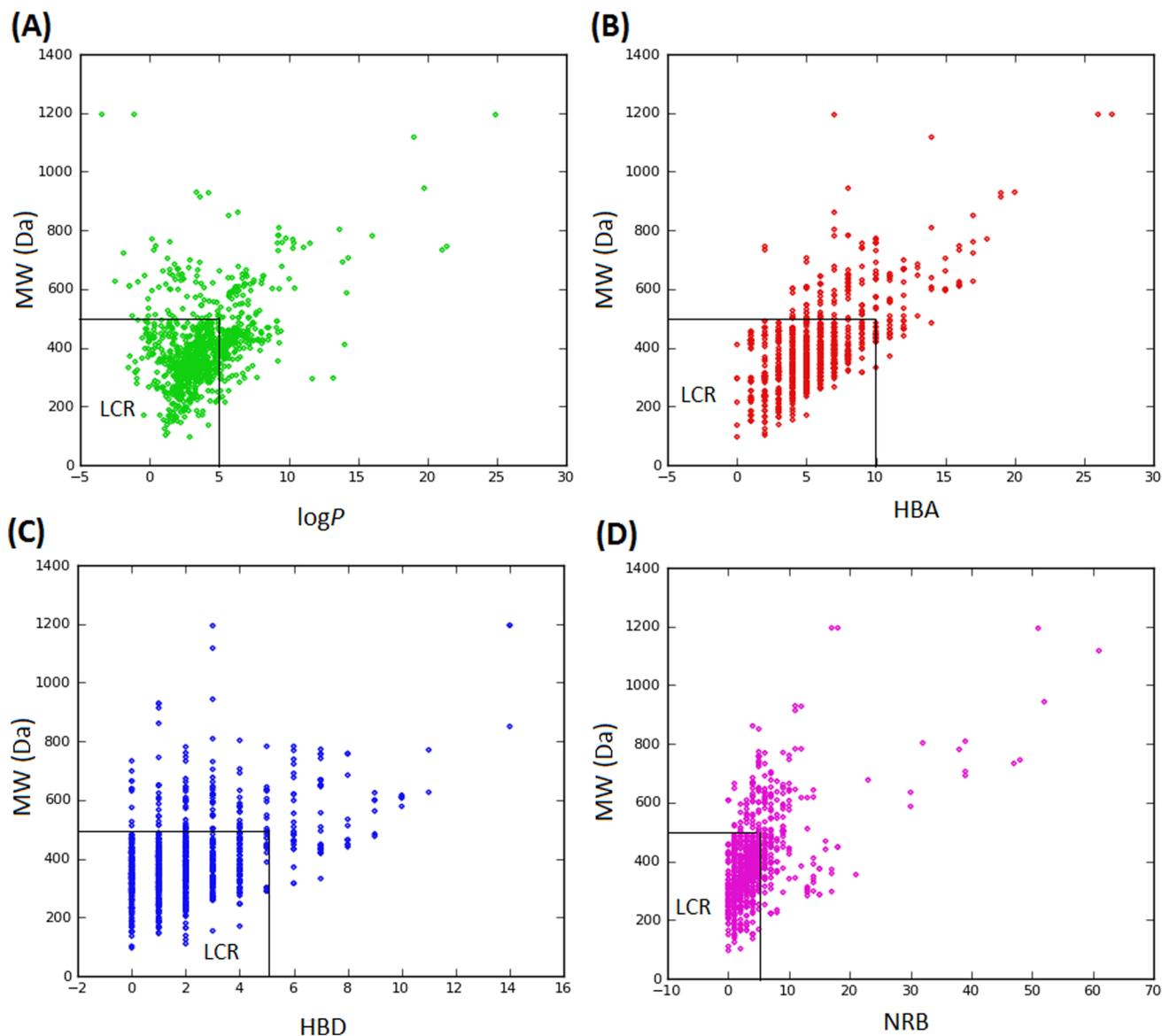
$$P_{pred} = SP_{exp} + (1 - S)P_{QP} \quad (1)$$

where *S* is the similarity, while *P*<sub>exp</sub> and *P*<sub>QP</sub> are the respective experimental and QikProp predictions for the most similar molecule within the training set. In equation (1), if *S* = 1, then the predicted property is equal to the measured experimental property of the training set compound. The distribution curves for two of the three determinants for the ro3 (log *S*<sub>wat</sub> and BIP<sub>caco-2</sub>) are shown in Figure S1. In general 43.6% of the AfroDb library was compliant with the ro3, while the respective % compliances for the various subsets were 67.6%, 92.3% and 100% for the “drug-like”, “lead-like” and “fragment-like” libraries. The most remarkable among the individual computed parameters was log *S*<sub>wat</sub>, which was met by 73.8% of the compounds within the AfroDb library. This property showed a Gaussian distribution for the “lead-like” and “fragment-like” subsets. The predicted apparent Caco-2 cell permeability, BIP<sub>caco-2</sub> (in nm s<sup>−1</sup>), model the permeability of the gut-blood barrier (for non-active transport). Even though the Caco-2 penetration parameter is not often correctly predicted computationally [36], only 38.2% of the compounds fell within the respected range for the BIP<sub>caco-2</sub> criterion. Histograms showing the predicted qualitative human oral absorption parameter (in the scale 1 = low, 2 = medium and 3 = high) are represented in Figure S2. It was observed that 55.8% of the compounds in AfroDb were predicted to have high human oral absorption. The predicted % human oral absorption (on 0 to 100% scale) shows a similar trend, 47.0% of the compounds being predicted to be absorbed at 100% and 63.1% of the compounds predicted to be absorbed at >90%.

The size of a molecule, as well as its capacity to make hydrogen bonds, its overall lipophilicity and its shape and flexibility are important properties to consider when determining permeability. Molecular flexibility has been seen as a parameter which is dependent on the NRB, a property which influences bioavailability in rats [36]. As previously mentioned, the distribution of the NRB for this dataset showed a peak value lying between 3 and 4 RBs, with an average value of 6.3 (Table 2). The gap between the average and peak values could be explained by some degree of conformational flexibility in some of the bulkiest NPs (having as many as 61 RBs).

### Prediction of Blood–brain Barrier (BBB) Penetration

The BBB partition coefficient is a good indicator of the ability of a drug to have access to the central nervous system (CNS). Drugs which are too polar do not cross the BBB. The blood/brain partition coefficients (logB/B) were computed and used as a predictor for access to the CNS. The predicted CNS activity was



**Figure 4. Pairwise comparison of mutual relationships between molecular descriptors.** (A) The distribution of the calculated log  $P$  versus MW, (B) HBA against MW, (C) HBD against MW and (D) NRB versus MW. LCR represents the Lipinski compliant regions. doi:10.1371/journal.pone.0078085.g004

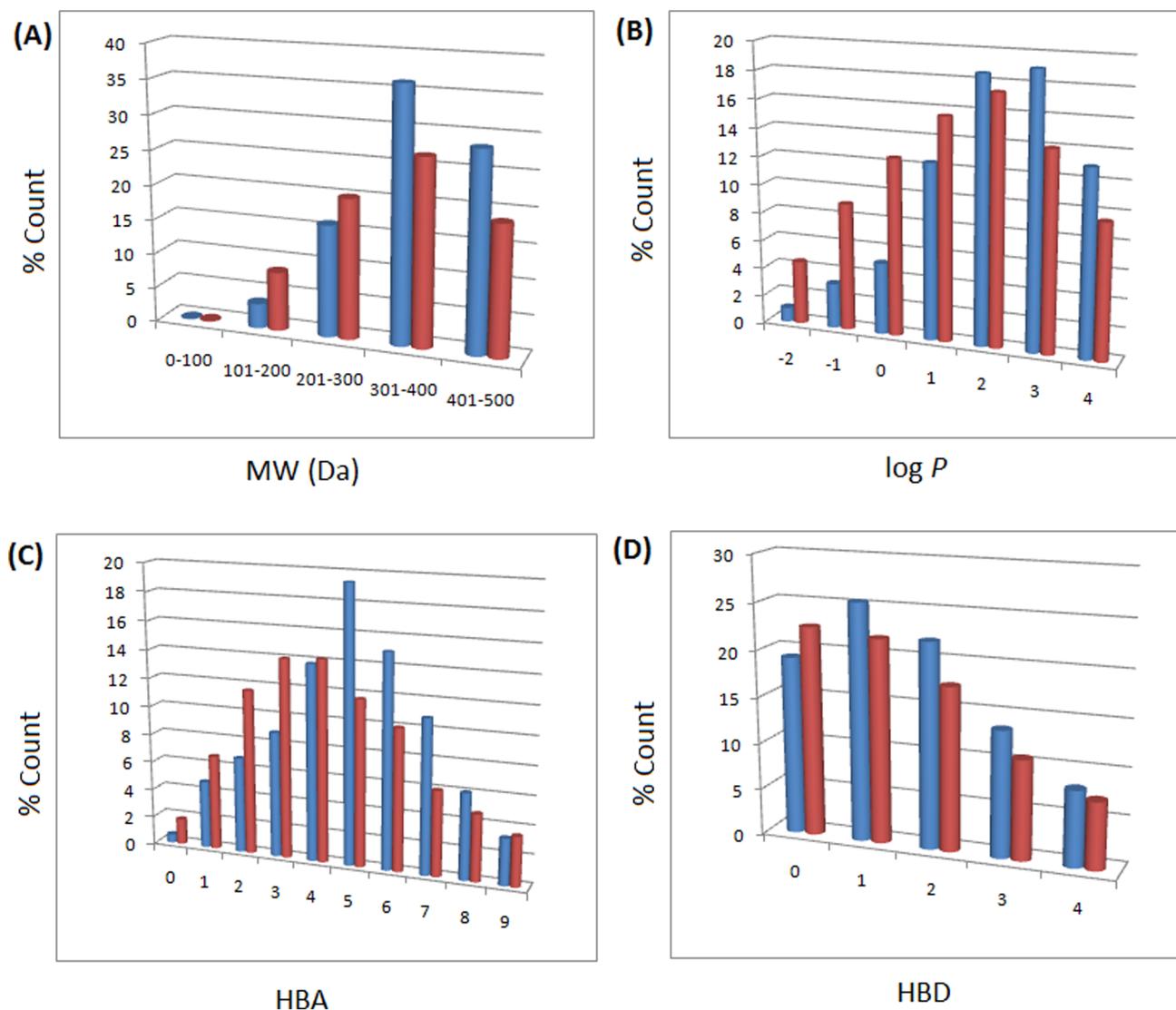
computed on a  $-2$  (inactive) to  $+2$  (active) scale and showed that only 3.8% of the compounds in AfroDb could be active in the CNS (predicted CNS activity  $>1$ ). A distribution of logB/B (Figure 7A) showed a Gaussian-shaped curve with a peak at  $-0.5$  logB/B units (the same for all the standard subsets), with  $>96\%$  of

the compounds in AfroDb falling within the recommended range for the predicted brain/blood partition coefficient ( $-3.0$  to  $1.2$ ). Madin-Darby canine kidney (MDCK) monolayers, are widely used to make oral absorption estimates, the reason being that these cells also express transporter proteins, but only express very low

**Table 1. Sources and biological activities of metabolites with calculated log  $P > 14$  found in AfroDb.**

Compound	Plant source (country)	Measured activity	Reference
01	<i>Klainedoxa gabonensis</i> (Cameroon)	xanthine oxidase inhibitory activity	[27]
02	<i>Hugonia castaneifolia</i> (Tanzania)	antifungal activity against <i>Cladosporium cucumericum</i>	[28]
03	<i>Stereospermum acuminatissimum</i> (Cameroon)	antifungal activities against <i>Candida</i> sp.	[29]

doi:10.1371/journal.pone.0078085.t001



**Figure 5. Comparison of property distribution for the two datasets by percentage distributions.** (A) MW, (B) log  $P$ , (C) HBA and (D) HBD. DNP in red and AfroDb in blue. For subfigure B, the x-axis label is the lower limit of binned data, e.g.  $-2$  is equivalent to  $-2$  to  $-1$ . doi:10.1371/journal.pone.0078085.g005

levels of metabolizing enzymes [36]. They are also used as an additional criterion to predict BBB penetration. Thus, our calculated apparent MDCK cell permeability could be considered to be a good mimic for the BBB penetration (for non-active transport). It was estimated that about 53% of the compounds had apparent MDCK cell permeability values falling within the recommended range of  $25\text{--}500\text{ nm s}^{-1}$  for 95% of known drugs. This situation was not improved in the “drug-like” and “lead-like” subsets ( $\sim 55\%$  and  $\sim 44\%$  respectively).

#### Prediction of Dermal Penetration

This factor is important for drugs administered through the skin. The distribution of computed skin permeability parameter,  $\log K_p$ , showed smooth Gaussian-shaped graphs centred at  $-2.5$   $\log K_p$  units for all 4 datasets (Figure S3), with  $\sim 92\%$  of the compounds in the AfroDb database falling within the recommended range for  $>95\%$  of known drugs. The predicted maximum transdermal transport rates,  $J_m$  (in  $\mu\text{ cm}^{-2}\text{ hr}^{-1}$ ), were computed from the predicted aqueous solubility ( $S_{wat}$ ) and skin

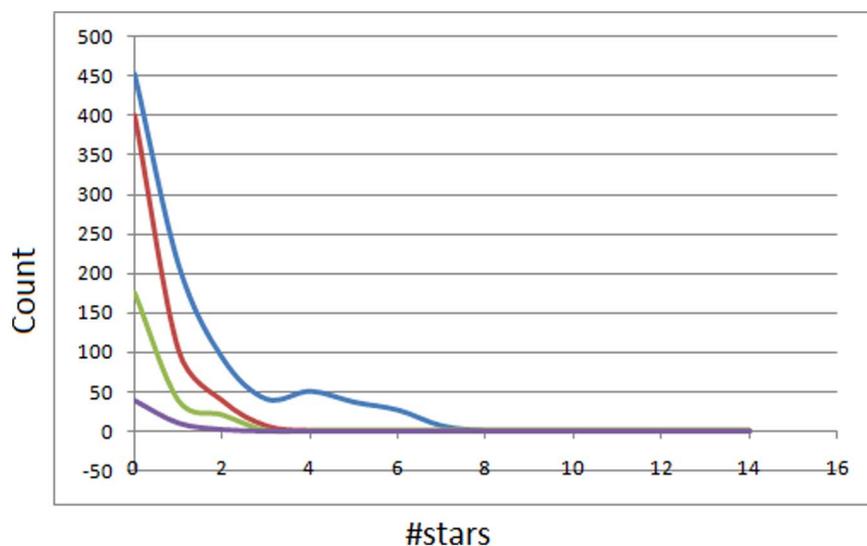
permeability ( $K_p$ ), using the relation (2):

$$J_m = K_p \times MW \times S_{wat} \quad (2)$$

This parameter showed variations from 0 to  $>1,600\ \mu\text{ cm}^{-2}\text{ hr}^{-1}$ , with only about 2.5% of the compounds in AfroDb having predicted value of  $J_m > 100\ \mu\text{ cm}^{-2}\text{ hr}^{-1}$ .

#### Prediction of Plasma-protein Binding

The degree to which a drug binds to the blood plasma proteins may seriously affect its efficacy. This is because binding of drugs to plasma proteins (like human serum albumin, lipoprotein, glycoprotein,  $\alpha$ ,  $\beta$ , and  $\gamma$  globulins) greatly reduces the quantity of the drug in general blood circulation and hence the less bound a drug is, the more efficiently it can traverse cell membranes or diffuse. The predicted plasma-protein binding has been estimated by the prediction of binding to human serum albumin; the  $\log K_{HSA}$  parameter (recommended range is  $-1.5$  to  $1.5$  for 95% of known



**Figure 6. Distribution curves for #stars within the AfroDb library, along with the standard “drug-like”, “lead-like” and “fragment-like” subsets.** Blue = AfroDb library, red = “drug-like” subset, green = “lead-like” subset and violet = “fragment-like” subset.  
doi:10.1371/journal.pone.0078085.g006

drugs). Figure 7B shows the variation of this calculated parameter within the AfroDb dataset, as well as for the standard subsets. This equally gave smooth Gaussian-shaped curves centred on 0.5  $\log K_{HSA}$  units for the total and “drug-like” libraries and  $-0.5 \log K_{HSA}$  units for the “lead-like” and “fragment-like” subsets. In addition, our calculations revealed that  $>92\%$  of the compounds within the AfroDb library complied with this parameter, which is an indication that a bulk of the compounds are likely to circulate freely within the blood stream and hence have access to their respective target sites.

### Metabolism Prediction

An estimated number of possible metabolic reactions has also been predicted by QikProp and used to determine whether the molecules can easily gain access to the target site after entering the blood stream. The average estimated number of possible metabolic reactions for the AfroDb library is between 6 and 7, while those of the standard subsets are respectively between 4 and 5, between 3 and 4 and between 2 and 3 (Table 2). Even though some of the compounds are likely to undergo as many as up to 25 metabolic reactions due to the complexity of some of the plant secondary metabolites within the database (Figure S4), up to about 80% of the compounds are predicted to undergo the recommended number of metabolic steps (1 to 8 reactions), with the situation improving to  $\sim 91\%$  and  $\sim 97\%$  in the “drug-like” and “lead-like” subsets respectively.

### Prediction of Blockage of Human Ether-a-go-go-related Gene Potassium (HERG $K^+$ ) Channel

Human ether-a-go-go related gene (HERG) encodes a potassium ion ( $K^+$ ) channel that is implicated in the fatal arrhythmia known as *torsade de pointes* or the long QT syndrome [37]. The HERG  $K^+$  channel, which is best known for its contribution to the electrical activity of the heart that coordinates the heart's beating, appears to be the molecular target responsible for the cardiac toxicity of a wide range of therapeutic drugs [38]. HERG has also been associated with modulating the functions of some cells of the nervous system and with establishing and maintaining cancer-like features in leukemic cells [39]. Thus, HERG  $K^+$  channel blockers

are potentially toxic and the predicted  $IC_{50}$  values often provide reasonable predictions for cardiac toxicity of drugs in the early stages of drug discovery [40]. In this work, the estimated or predicted  $IC_{50}$  values for blockage of this channel have been used to model the process *in silico*. The recommended range for predicted  $\log IC_{50}$  values for blockage of HERG  $K^+$  channels ( $\log HERG$ ) is  $> -5$ . A distribution curve for the variation of the predicted  $\log HERG$  is shown in Figure 7C, which is a left-slanted Gaussian-shaped curve, with a peak at  $-5.5 \log HERG$  units for the total library and  $-4.5 \log HERG$  units for all the standard subsets. It was observed that, in general, this parameter was predicted to fall within the recommended range for about 58% of the compounds within the AfroDb database,  $\sim 70\%$  for the “drug-like” subset and  $\sim 90\%$  for the “lead-like” subset.

### Diversity Analysis

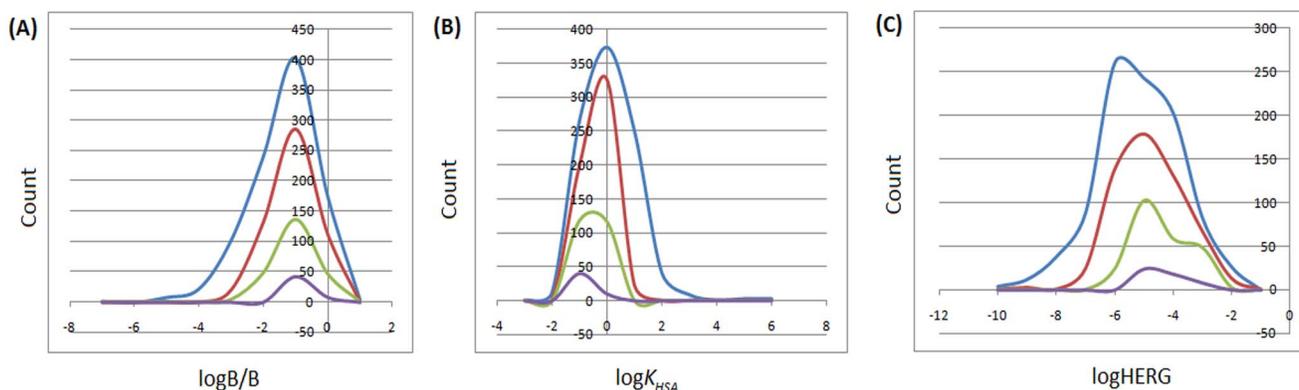
In order to reduce redundancy and enhance the coverage of biological activity and chemical space, a dataset for virtual screening must have the requirement of diversity. In this case, we carried out a simple molecular descriptor comparison with a relatively larger known diverse library (the DIVERSet<sup>TM</sup> Database, containing 48,651 compounds) from the ChemBridge Corporation [25]. Histograms showing the calculated descriptors; MW, HBA, HBD,  $\log P$ , NRB, number of rings (NR), number of oxygens (NO), and total polar surface area (TPSA) are shown in Figure 8 for AfroDb (in light green) and the ChemBridge dataset (in red). The regions shown in dark green represent regions of intersection. The MW of the AfroDb dataset stretches up to about 1000 Da, while that of the ChemBridge dataset is restricted to the range  $200 \leq MW \leq 500$  Da. This observation could be explained by the complexity and large sizes of some of the compounds within the natural product library. The large proportion of very large and complex NPs in AfroDb, could also explain the average molar weight ( $\overline{MW} = 406$ Da), when compared to those of the standard “drug-like”, “lead-like” libraries and typical drugs ( $\overline{MW} = 310$ Da for typical drugs) [41]. This same explanation holds for the trend which is observed in the distributions of  $\log P$ , HBD, NCC, NO, NRB, NR, TPSA and HBA for AfroDb, when compared with the ChemBridge dataset. It was generally observed that the AfroDb

**Table 2.** Summary of average predicted pharmacokinetic property distributions of the total AfroDb library in comparison with the various subsets.

Library name	<sup>a</sup> Lib. size	<sup>b</sup> % compl.	<sup>c</sup> MW (Da)	<sup>d</sup> LogP	<sup>e</sup> HBA	<sup>f</sup> HBD	<sup>g</sup> NRB
Total	1008	48	406	3.99	5.76	1.67	6.30
Drug-like	610	73	328	2.99	4.89	1.25	4.24
Lead-like	239	75	266	2.44	3.91	0.87	3.43
Fragment-like	51	76	219	1.89	3.39	0.60	1.40
Library name	<sup>h</sup> LogB/B	<sup>i</sup> BIP <sub>caco-2</sub> (nm s <sup>-1</sup> )	<sup>j</sup> S <sub>mol</sub> (Å <sup>2</sup> )	<sup>k</sup> S <sub>mol,hfob</sub> (Å <sup>2</sup> )	<sup>l</sup> V <sub>mol</sub> (Å <sup>3</sup> )	<sup>m</sup> LogS <sub>wat</sub> (S in mol L <sup>-1</sup> )	<sup>n</sup> LogK <sub>HSA</sub>
Total	-0.90	1516	674	415	1265	-5.11	0.59
Drug-like	-0.63	1663	568	312	1030	-3.88	0.21
Lead-like	-0.57	2032	492	235	860	-3.11	-0.02
Fragment-like	-0.29	1983	424	139	712	-2.50	-0.20
Library name	<sup>o</sup> MDCK	<sup>p</sup> Ind <sub>coh</sub>	<sup>q</sup> Glob	<sup>r</sup> QP <sub>polrz</sub> (Å <sup>3</sup> )	<sup>s</sup> LogHERG	<sup>t</sup> LogK <sub>p</sub>	<sup>u</sup> # metab
Total	859	0.009	0.84	41.78	-4.68	-2.84	6.13
Drug-like	944	0.008	0.87	33.75	-4.33	-2.73	4.85
Lead-like	1206	0.005	0.89	27.54	-3.99	-2.55	3.56
Fragment-like	1078	0.005	0.91	23.27	-3.90	-2.39	2.28

<sup>a</sup>Size or number of compounds in library;<sup>b</sup>Percentage of compounds with #star=0;<sup>c</sup>Molar weight (range for 95% of drugs: 130–725 Da);<sup>d</sup>Logarithm of partitioning coefficient between *n*-octanol and water phases (range for 95% of drugs: -2 to 6);<sup>e</sup>Number of hydrogen bonds accepted by the molecule (range for 95% of drugs: 2–20);<sup>f</sup>Number of hydrogen bonds donated by the molecule (range for 95% of drugs: 0–6);<sup>g</sup>Number of rotatable bonds (range for 95% of drugs: 0–15);<sup>h</sup>Logarithm of predicted blood/brain barrier partition coefficient (range for 95% of drugs: -3.0 to 1.0);<sup>i</sup>Predicted apparent Caco-2 cell membrane permeability in Boehringer-Ingelheim scale, in nm/s (range for 95% of drugs: <5 low, >100 high);<sup>j</sup>Total solvent-accessible molecular surface, in Å<sup>2</sup> (probe radius 1.4 Å) (range for 95% of drugs: 300–1000 Å<sup>2</sup>);<sup>k</sup>Hydrophobic portion of the solvent-accessible molecular surface, in Å<sup>2</sup> (probe radius 1.4 Å) (range for 95% of drugs: 0–750 Å<sup>2</sup>);<sup>l</sup>Total volume of molecule enclosed by solvent-accessible molecular surface, in Å<sup>3</sup> (probe radius 1.4 Å) (range for 95% of drugs: 500–2000 Å<sup>3</sup>);<sup>m</sup>Logarithm of aqueous solubility (range for 95% of drugs: -6.0 to 0.5);<sup>n</sup>Logarithm of predicted binding constant to human serum albumin (range for 95% of drugs: -1.5 to 1.2);<sup>o</sup>Predicted apparent MDCK cell permeability in nm/sec (<25 poor, >500 great);<sup>p</sup>Index of cohesion interaction in solids (0.0 to 0.05 for 95% of drugs);<sup>q</sup>Globularity descriptor (0.75 to 0.95 for 95% of drugs);<sup>r</sup>Predicted polarizability (13.0 to 70.0 for 95% of drugs);<sup>s</sup>Predicted IC<sub>50</sub> value for blockage of HERG K<sup>+</sup> channels (concern<-5);<sup>t</sup>Predicted skin permeability (-8.0 to -1.0 for 95% of drugs);<sup>u</sup>Number of likely metabolic reactions (range for 95% of drugs: 1–8).

doi:10.1371/journal.pone.0078085.t002

**Figure 7.** Distribution curves for some computed ADME parameters. (A) logB/B, (B) logK<sub>HSA</sub>, (C) logHERG. For subfigure B, the x-axis label is the lower limit of binned data, e.g. -2 is equivalent to -2 to -1. The colour codes are according to Figure 5.

doi:10.1371/journal.pone.0078085.g007

dataset covers a different physico-chemical space from the ChemBridge Diversity dataset. Principal component analysis (PCA) was as well used as a means of comparing the extent of diversity of the two datasets. This consists in reducing the dimensionality of the calculated descriptors by linearly transforming the data, by calculating a new and smaller set of descriptors, which are uncorrelated and normalised (mean = 0, variance = 1). The PCA scatter plot of the previously calculated physico-chemical properties of the AfroDb (light green) and ChemBridge Diverset database (red), shown in Figure 9, is a visual representation of the molecules in the respective datasets, as described by the 3 selected principal components (PC1, PC2 and PC3). Each point shown corresponds to a molecule, the spread of the points representing the diversity of the respective datasets. The first three principal components (PCs) explain 80.1% (AfroDb) and 63.7% (ChemBridge) of the variance of the individual datasets. The larger number of outliers in the case of the AfroDb dataset (away from the centre and towards the sides of the cube) indicates a wider sampling of the chemical space compared to the ChemBridge Diverset collection.

### Searching for Most Common Substructures

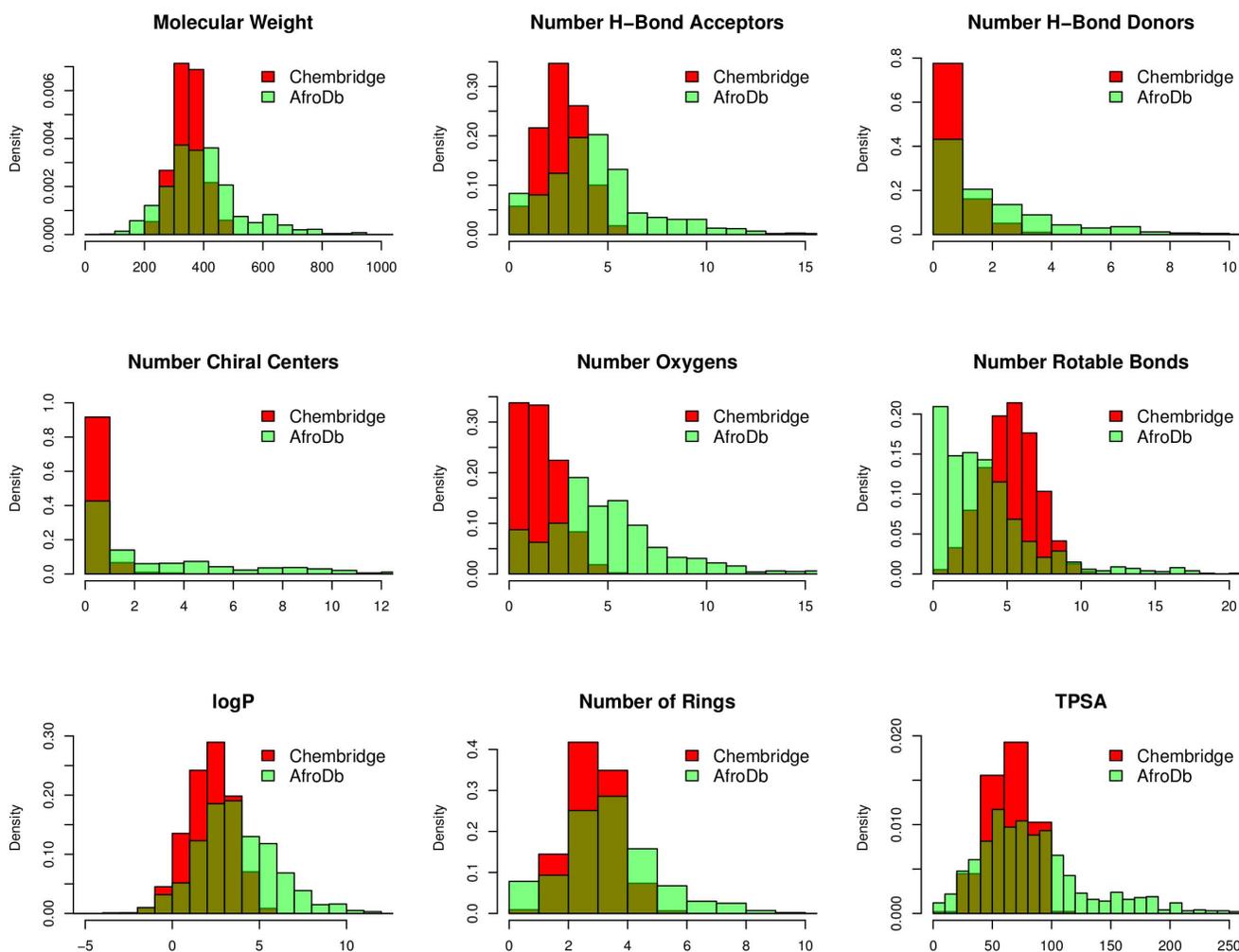
The most common substructure selection (MCSS) panel for compound selection (Figure 10) is based on substructures that can be synthetically combined and are common in “drug-like” molecules and allows a direct selection and identification of compounds containing such substructures. The panel highlights the large diversity of the rings present in the NPs of AfroDb.

### Tautomer Generation and Synthetic Accessibility Evaluation

Of the 1008 compounds within AfroDb, we were able to identify 2308 possible tautomers, and respectively 1463, 524 and 103 possible tautomers for the standard subsets (“drug-like”, “lead-like” and “fragment-like”). The number of possible analogues were estimated at 9098, even the synthetic accessibility scores were quite weak for most of the proposed, as was expected for NPs.

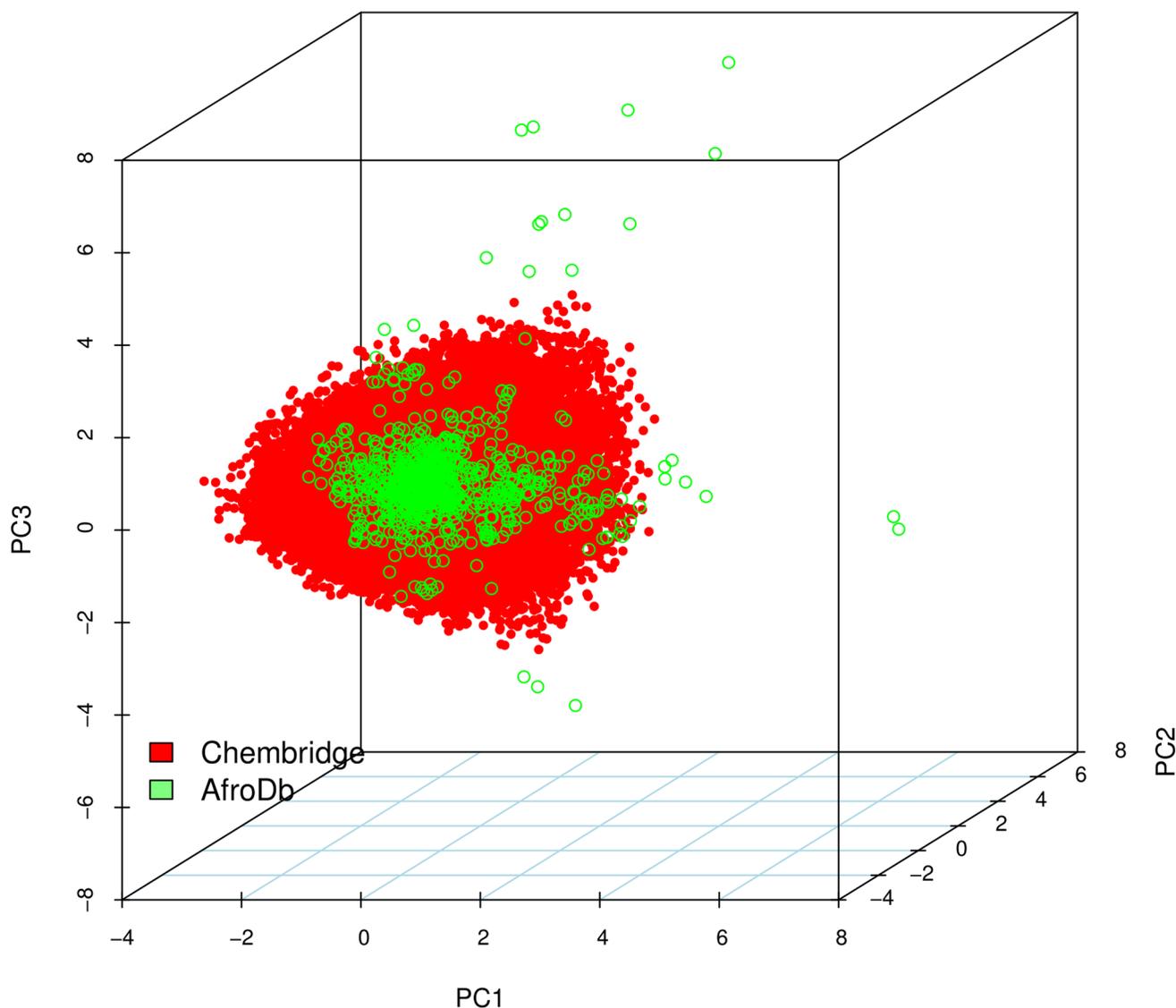
### Some Selected Highly Potent Compounds

Our discussion also includes selected compounds within the AfroDb database with promising biological activities, which have been isolated from African medicinal plants from the very first



**Figure 8. A simple descriptor-based comparison of the AfroDb database and the ChemBridge Diversity database.** Comparison of typical physico-chemical property distributions (MW, HBA, HBD, NCC, NO, NR, log *P*, NR and TPSA) in the AfroDb (green) and ChemBridge Diverset (red) database. All histograms and scatterplots were generated with the R software [85]. doi:10.1371/journal.pone.0078085.g008

## PCA Descriptors

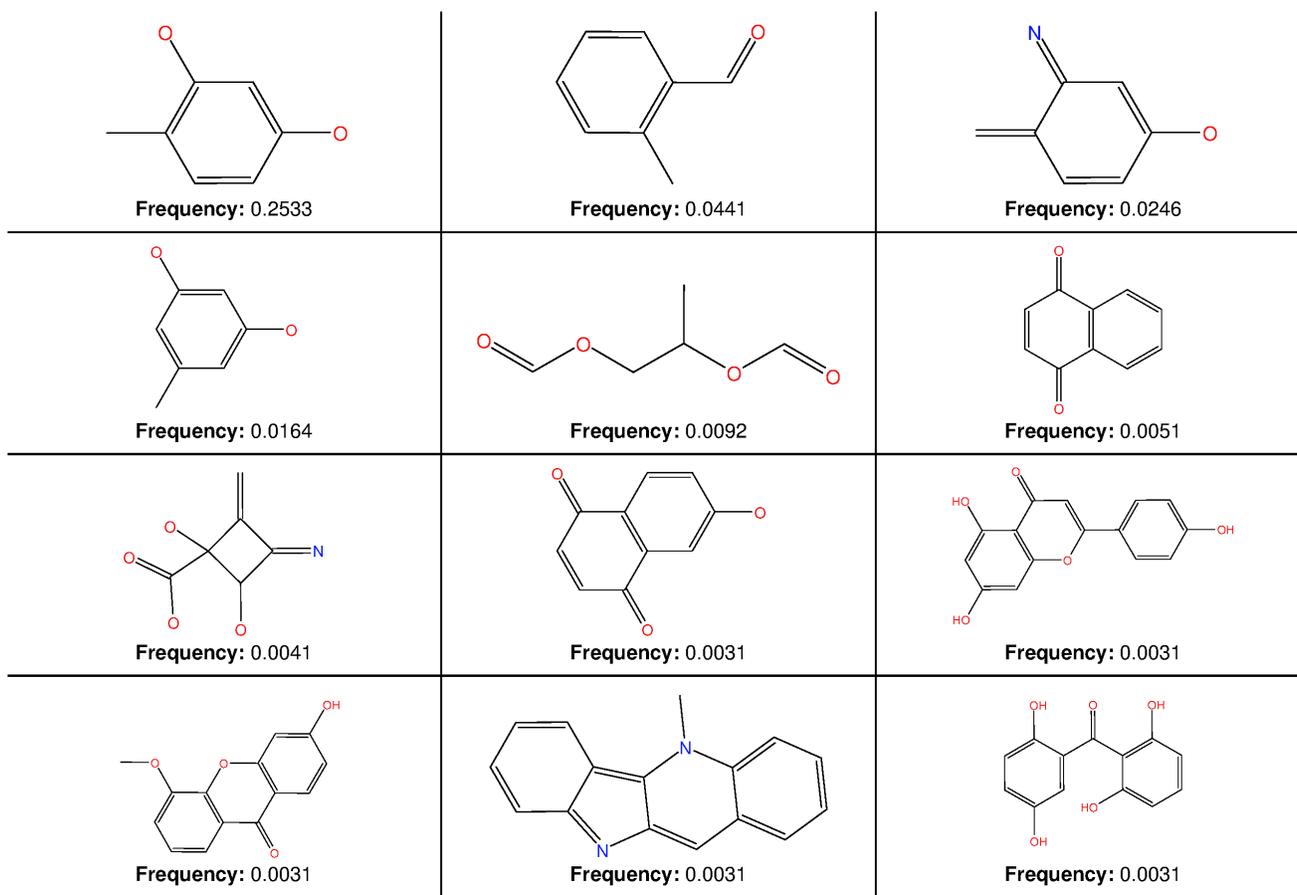


**Figure 9. A principal component analysis (PCA) plot, showing the comparison of the chemical space defined by the NPs in AfroDb (green) and the chemical space represented by NPs in the ChemBridge Diversity (red) databases.**  
doi:10.1371/journal.pone.0078085.g009

time. These have been summarised in Table 3, while their 2D structures are shown in Figure 11.

Among the promising compounds with anticancer activities, the trachylobane diterpene ent-trachyloban-3 $\beta$ -ol (**04**), derived from *Croton zambesicus* (Euphorbiaceae) in Benin, has been shown to exert a dose dependent cytotoxic effect, which varies between cell lines. Induction of apoptosis in HL-60 cells could be detected at a concentration of 50  $\mu$ M after 24-h treatment. This compound was also able to induce apoptosis in human promyelocytic leukemia cells via caspase-3 activation in a concentration-dependent manner [42]. Other promising cytotoxic agents included are six aporphines (**05–10**) from *Cassytha filiformis*, among which actinodaphnine (**05**), cassythine (**06**), and dicentrine (**07**) were also shown to possess antitrypanosomal properties *in vitro* on *Trypano-*

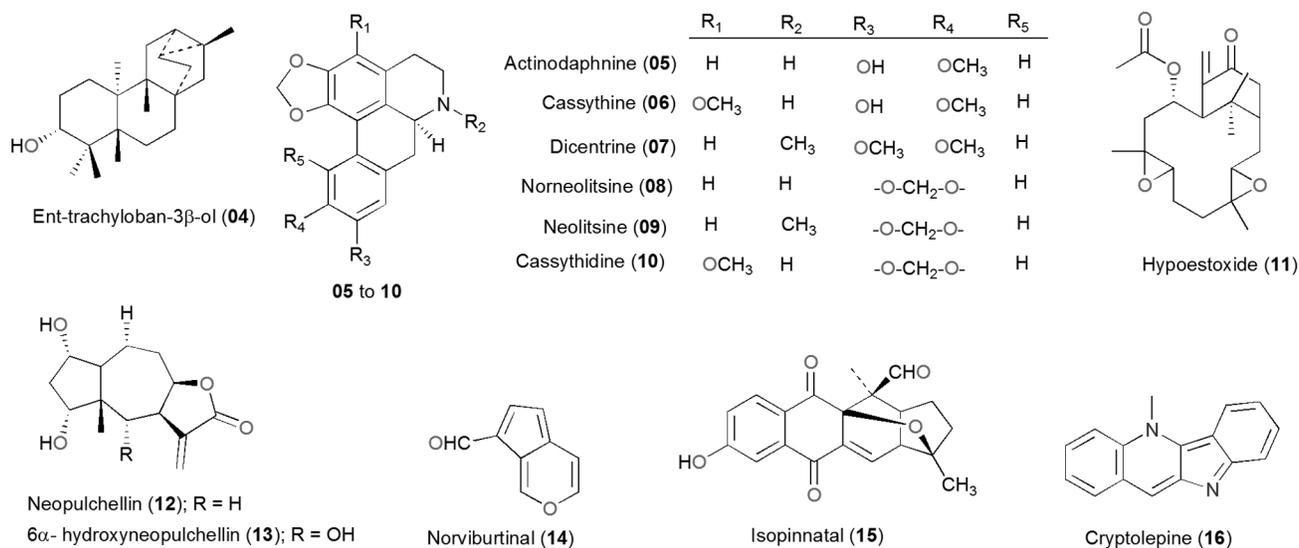
*soma brucei brucei* [43]. Hypoestoxide (HE, **11**), a natural diterpenoid [a bicyclo (9, 3, 1) pentadecane], derived from *Hypoestes rosea* growing in Nigeria [44] has been reported to be a potent nonsteroidal anti-inflammatory drug. This compound is also reported to be non mutagenic and possesses antiangiogenic and antitumor activities, also inhibiting the activity of I $\kappa$ B kinase [45,46]. Other potent anticancer compounds have been isolated from *Gaillardia aristata* growing in Egypt [47]. These include neopulchellin (**12**) and 6 $\alpha$ -hydroxyneopulchellin (**13**), with respective IC<sub>50</sub> values of 0.43 and 0.32 mgmL<sup>-1</sup> against human cancer cell lines (breast (MCF7)) and 0.46, 0.34 mgmL<sup>-1</sup> against human colon (HCT116), respectively. Examples of NPs with anticancer properties derived from plants growing in the Southern African region include norviburtinal (**14**) and isopinnatal (**15**).



**Figure 10. MCSS panel in AfroDb, featuring the most common cyclic structures included in the database.**  
doi:10.1371/journal.pone.0078085.g010

These compounds were derived from *Kigelia pinnata* harvested in Zimbabwe and have shown cytotoxic activity against various cancer cell lines after 144 h exposure [48].

The anti-malarial properties of some of the compounds included in the AfroDb dataset have been discussed in several review papers [49–53], as well as compounds with antiparasitic properties [54] and antimicrobial activities [55]. One of the



**Figure 11. 2D structures of selected promising compounds derived from the African flora and included in AfroDb.**  
doi:10.1371/journal.pone.0078085.g011

**Table 3.** Summary of selected promising potent compounds derived from African medicinal plants and currently included in AfroDb.

Compound	Plant source (country)	Measured activity(ies)	Reference
04	<i>Croton zambesicus</i> (Benin)	Induction of apoptosis in Human promyelocytic leukemia cells	[42]
05–10	<i>Cassipoupa filiformis</i> (Benin)	Cytotoxic, antitrypanosomal	[43]
11	<i>Hypoestes rosea</i> (Nigeria)	Anti-inflammatory, antiangiogenic and antitumor activities, inhibiting the activity of I $\kappa$ B kinase	[44–46]
12–13	<i>Gaillardia aristata</i> (Egypt)	Anticancer	[47]
14–15	<i>Kigelia pinnata</i> (Zimbabwe)	Anticancer	[48]
16	<i>Sida acuta</i> (Cote d'Ivoire)	Anti-malarial	[56]

doi:10.1371/journal.pone.0078085.t003

promising anti-malarial compounds include cryptolepine (**16**) derived from *Sida acuta* (Malvaceae), a plant growing in Côte d'Ivoire, which traditional healers commonly use for the treatment of malaria [56]. This compound has shown antiplasmodial activity against the FcM29-Cameroon (chloroquine-resistant strain) and a Nigerian chloroquine-sensitive strain, with IC<sub>50</sub> values of 0.13 and 0.17  $\mu$ g/mL respectively.

### Usefulness of the AfroDb Library

It is important to mention that virtual screening results sometimes provide insight and direct natural product chemists to search for theoretically active principles with attractive ADMET profiles, which have been previously isolated, but not tested for activity against specified drug targets (if samples are absent). This “resurrection” process could prove to be a better procedure for lead search than the random screening, which is common practice in most laboratories in Africa. AfroDb is constantly being updated; meanwhile a MySQL platform to facilitate the searching of this database and ordering of compound samples is under development within our group and will also be published subsequently. However, 3D structures of the compounds, as well as their physico-chemical properties that were used to evaluate “drug-likeness” and the DMPK profile, can be freely downloaded in the (Dataset S1). In addition, information about compound sample availability can be obtained on request from the authors of this paper or from the pan-African Natural Products Library (p-ANAPL) project [57–58].

## Materials and Methods

### Data Sources and Cutoff Points for Biological Activities

The plant sources, geographical collection sites, chemical structures of pure compounds as well as their biological activities, were retrieved from literature sources comprising mainly published articles from across the major journals of natural product chemistry, as well as MSc and PhD theses, textbook chapters, as well as unpublished conference presentations (from personal communication with the authors). References were recorded from 1971 to 2013. Following criteria used by Mahmoudi et al. [59] and Wilcox et al. [60], a pure compound was considered highly active if IC<sub>50</sub> < 0.06  $\mu$ M, being active with 0.06  $\mu$ M  $\leq$  IC<sub>50</sub>  $\leq$  5  $\mu$ M, weakly active when 5  $\mu$ M  $\leq$  IC<sub>50</sub>  $\leq$  10  $\mu$ M and compounds with IC<sub>50</sub> > 10  $\mu$ M were considered inactive. Up to weakly active compounds were selected.

### Generation of 3D Models, Optimization and Calculation of Molecular Descriptors

Based on the known chemical structures of the NPs, all 3D molecular structures were generated using the graphical user interface (GUI) of the MOE software [61] running on a Linux workstation with a 3.5 GHz Intel Core2 Duo processor. The 3D structures were generated using the builder module of MOE and energy minimization was subsequently carried out using the MMFF94 force field [62] until a gradient of 0.01 kcal/mol was reached. The 3D structures of the compounds were then saved as mol2 files subsequently included into a MOE database (.mdb) file and converted to other file formats (.sdf, .mol, .mol2 and .ldb), which are suitable for use in several virtual screening workflow protocols. Up to 10 possible tautomers were generated per compound in the dataset. The MW, NRB, log P, HBA, HBD and Lipinski violations were calculated using the molecular descriptor calculator included in the QuSAR module of the MOE package [61].

### Determination of ADMET Profiles

The previously prepared 1,008 low energy 3D chemical structures in the AfroDb library were saved in mol2 format and initially treated with LigPrep [63], distributed by Schrodinger Inc. This implementation was carried out with the graphical user interface (GUI) of the Maestro software package [64], using the OPLS forcefield [65–67]. Protonation states at biologically relevant pH were correctly assigned (group I metals in simple salts were disconnected, strong acids were deprotonated, strong bases protonated, while topological duplicates and explicit hydrogens were added). A set of ADMET-related properties (a total of 46 molecular descriptors) were calculated by using the QikProp program [68] running in normal mode. QikProp generates physically relevant descriptors, and uses them to perform ADMET predictions. An overall ADME-compliance score – drug-likeness parameter (indicated by #stars), was used to assess the pharmacokinetic profiles of the compounds within the AfroDb library. The #stars parameter indicates the number of property descriptors computed by QikProp that fall outside the optimum range of values for 95% of known drugs. The methods implemented were developed by Jorgensen et al. [69–70] and among the calculated descriptors are: the total solvent-accessible molecular surface,  $S_{mol}$  in  $\text{\AA}^2$  (probe radius 1.4  $\text{\AA}$ ) (range for 95% of drugs: 300–1000  $\text{\AA}^2$ ); the hydrophobic portion of the solvent-accessible molecular surface,  $S_{mol,hfob}$  in  $\text{\AA}^2$  (probe radius 1.4  $\text{\AA}$ ) (range for 95% of drugs: 0–750  $\text{\AA}^2$ ); the total volume of molecule enclosed by solvent-accessible molecular surface,  $V_{mol}$  in  $\text{\AA}^3$  (probe radius 1.4  $\text{\AA}$ ) (range for 95% of drugs: 500–2000  $\text{\AA}^3$ ); the logarithm

of aqueous solubility,  $\log S_{wat}$  (range for 95% of drugs:  $-6.0$  to  $0.5$ ) [69,71]; the logarithm of predicted binding constant to human serum albumin,  $\log K_{HSA}$  (range for 95% of drugs:  $-1.5$  to  $1.2$ ) [72]; the logarithm of predicted blood/brain barrier partition coefficient,  $\log B/B$  (range for 95% of drugs:  $-3.0$  to  $1.0$ ) [73–75]; the predicted apparent Caco-2 cell membrane permeability ( $BIP_{caco-2}$ ) in Boehringer–Ingelheim scale, in nm/s (range for 95% of drugs:  $<5$  low,  $>100$  high) [76–78]; the predicted apparent Madin-Darby canine kidney (MDCK) cell permeability in  $\text{nm s}^{-1}$  ( $<25$  poor,  $>500$  great) [77]; the index of cohesion interaction in solids,  $\text{Ind}_{\text{coh}}$ , calculated from the HBA, HBD and the surface area accessible to the solvent, SASA ( $S_{mol}$ ) by the relation  $\text{Ind}_{\text{coh}} = \text{HBA} \times \text{HBD}^{1/2} / S_{mol}$  (0.0 to 0.05 for 95% of drugs) [71]; the globularity descriptor,  $\text{Glob} = (4\pi r^2) / S_{mol}$ , where  $r$  is the radius of the sphere whose volume is equal to the molecular volume (0.75 to 0.95 for 95% of drugs); the predicted polarizability,  $QP_{\text{polz}}$  (13.0 to 70.0 for 95% of drugs); the predicted logarithm of  $\text{IC}_{50}$  value for blockage of HERG  $\text{K}^+$  channels,  $\log \text{HERG}$  (concern  $< -5$ ) [79–80]; the predicted skin permeability,  $\log K_p$  ( $-8.0$  to  $-1.0$  for 95% of drugs) [81–82]; and the number of likely metabolic reactions,  $\#\text{metab}$  (range for 95% of drugs: 1–8).

### Diversity Analysis and Searching with Most Common Substructures

The ChemBridge Diverset dataset (48,651 compounds) was downloaded from the official ChemBridge webpage [25]. The MW, NRB,  $\log P$ ,  $\log S$ , HBA, HBD, THSA, TPSA, NO, NCC, NR and number of Lipinski violations were calculated using the molecular descriptor calculator included in the QuSAR module of the MOE package [61]. The LibMCS program of JKlustor [83] was used for maximum common substructure clustering of the AfroDb database. In the MCSS search, only structures with  $\text{MW} \leq 700$  were included, since MCSS clustering is only feasible on small molecules. This means, only 975 of the 1008 compounds of the AfroDb were analyzed for MCSS. The compounds were fragmented using the RECAP algorithm [84] and the frequency of the generated fragments were analysed, leading to the identification of the most frequent substructures.

### Conclusions

To the best of our knowledge, AfroDb represents the largest “drug-like” and diverse collection of 3D structures of NPs covering the geographical region of the entire African continent, which is readily available for download and use in virtual screening campaigns. Virtual screening workflows usually involve docking a compound library into the binding site of a target receptor and using scoring functions and binding free energy calculations to identify putative binders. The availability of 3D structures of the compounds to be used for docking is of utmost importance.

### References

- Akerele O (1993) In Summary of WHO guidelines for the assessment of herbal medicine. Herbalgram 28: 13–19.
- Potterat O, Hamburger M (2008) Drug discovery and development with plant-derived compounds. In: Petersen F, Amstutz R, editors. Progress in drug research: natural compounds as drugs. Birkhäuser: Verlag AG. Pp. 45–118.
- Chin YW, Balunas MJ, Chai HB, Kinghorn AD (2006) Drug discovery from natural sources. The AAPS Journal 8(2): E239–E253.
- Li JWH, Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? Science 325: 161–165.
- Quinn RJ, Carroll AR, Pham MB, Baron P, Palframan ME, et al. (2008) Developing a drug-like natural product library. J Nat Prod 71: 464–468.
- Harvey AL (2008) Natural products in drug discovery. Drug Discov Today 13: 894–901.
- Newman DJ (2008) Natural products as leads to potential drugs: an old process or the new hope for drug discovery? J Med Chem 51: 2589–2599.

Therefore the availability of such structures within AfroDb, as well as their calculated physico-chemical properties and indicators of “drug-likeness” within this newly developed database will facilitate the drug discovery process from leads that have been identified from African medicinal plants.

### Supporting Information

**Figure S1 Distribution curves for compliance to Jorgensen’s “Rule of Three”.** (A) calculated  $\log S_{wat}$  against count, (B) predicted  $BIP_{caco-2}$  against count. Colour codes are as defined in Figure 6.

(TIF)

**Figure S2 Histograms showing the distribution of human oral absorption predictions.**

(TIF)

**Figure S3 Distribution curves for the predicted skin penetration parameter.** Colour codes are as defined in Figure 6.

(TIF)

**Figure S4 Graphs showing the distribution of the predicted number of metabolic reactions for compounds in AfroDb.** Colour codes are as defined in Figure 6.

(TIF)

**Dataset S1 SDF 3D structures of compounds currently included in AfroDb, along with computed physico-chemical descriptors used to predict drug-likeness and DMPK profile.** This file is saved in.sdf format (which can be viewed using many drug discovery software including Maestro, MOE, Discovery studio, LigandScout, etc.) or into.ldb format using the software LigandScout. Requests for other file formats could be addressed to the authors of this article.

(SDF)

### Acknowledgments

Mr. Bakoh Ndingkokhar, Chemistry Department, University of Buea, Cameroon assisted in the generation of the 3D models while Ms Inna Slynko, Department of Pharmaceutical Sciences, Martin-Luther University of Halle-Wittenberg, Germany, is acknowledged for assisting in carrying out the predicted MedChem Transformations and evaluation of the synthetic accessibilities.

### Author Contributions

Conceived and designed the experiments: WS LMM DZ FNK SMNE. Performed the experiments: FNK DZ MS SBB. Analyzed the data: MS FNK DZ SBB RM JAM. Contributed reagents/materials/analysis tools: WS DZ FNK LMM MS RM SMNE JAM LLL. Wrote the paper: FNK DZ SBB MS.

- Efange SMN (2002) Natural products: a continuing source of inspiration for the medicinal chemist. In: Iwu MM, Wootton JC, editors. Advances in Phytomedicine. Amsterdam: Elsevier Science. 61–69.
- Grabowski K, Baringhaus KH, Schneider G (2008) Scaffold diversity of natural products: inspiration for combinatorial library design. Nat Prod Rep 25:892–904.
- Grabowski K, Schneider G (2007) Properties and architecture of drugs and natural products revisited. Curr Chem Biol 1: 115–127.
- Wetzel S, Schuffenhauer A, Roggo S, Ertl P, Waldmann H (2007) Cheminformatic analysis of natural products and their chemical space. Chimia Int J Chem 61:355–360.
- Hostettmann K, Marston A, Ndjoko K, Wolfender JL (2000) The potential of African plants as a source of drugs. Curr Org Chem 4: 973–1010.

13. Høltje HD, Sippl W, Rognan D, Folkers G (2008) Molecular modeling: basic principles and applications. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA. 181–258 p.
14. Kubinyi H (1998) Combinatorial and computational approaches in structure-based drug design. *Curr. Opin. Drug Discov Devel* 1, 16–27.
15. DiMasi JA, Hansen RW, Grabowski HG (2003) The price of innovation: new estimates of drug development costs. *J Health Econ* 22:151–185.
16. Ntie-Kang F, Mbah JA, Mbaze LM, Lifongo LL, Scharfe M, et al. (2013) CamMedNP: Building the Cameroonian 3D structural natural products database for virtual screening. *BMC Complement Altern Med* 13:88.
17. Zofou D, Ntie-Kang F, Sippl W, Efang SMN (2013) Bioactive natural products derived from the Central African flora against Neglected Tropical Diseases and HIV. *Nat Prod Rep* 30:1098–1120.
18. Ntie-Kang F, Onguéné PA, Scharfe M, Owono LCO, Megnassan E, et al. (2013) ConMedNP: An *in silico* assessment of the “druglikeness” and medicinal potential of a natural product library from medicinal plants in Central Africa. *RSC Advances*, submitted.
19. Couzin J (1998) Supercomputing-computer experts urge new federal initiative. *Science* 281: 762.
20. Irwin JJ, Shoichet BK (2005) ZINC – A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45(1): 177–182.
21. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) ZINC: A free tool to discover chemistry for biology. *J Chem Inf Model* 52: 1757–1768.
22. Buzbee B (1993) Workstation clusters rise and shine. *Science* 261: 852–853.
23. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Delivery Rev* 23: 3–25.
24. Chapman and Hall/CRC Press: Dictionary of Natural Products on CD-Rom. London; 2005.
25. ChemBridge Corporation website. Available: <http://chembridge.com/> Accessed 2013 June 13.
26. Feher M, Schmidt JM (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci* 43: 218–227.
27. Nkanwen ERS, Gokayev AS, Wabo HK, Bankeu JJK, Iqbal MC, et al. (2013) Lanostane-type triterpenoid and steroid from the stem bark of *Klainedoxa gabonensis*. *Fitoterapia* 86:108–114.
28. Mdee LK, Waibel R, Nkunya MHH, Jonker SA, Achenbach H (1998) Rosane diterpenes and bis-dinorditerpenes from *Hugonia castaneifolia*. *Phytochemistry* 49:1107–1113.
29. Sob SVT, Wabo HK, Tang C-P, Tane P, Ngadjui BT, et al. (2011) Phenol esters and other constituents from the stem barks of *Stereospermum acuminatissimum*. *J Asian Nat Prod Res* 13:1128–1134.
30. Schrödinger Press: QikProp 3.4 User Manual, LLC, New York, NY, 2011.
31. Teague SJ, Davis AM, Lecson PD, Opeca TI (1999) The design of leadlike combinatorial libraries. *Angew Chem, Int Ed* 38: 3743–3748.
32. Oprea TI (2002) Current trends in lead discovery: are we looking for the appropriate properties? *J Comput-Aided Mol Des* 16: 325–334.
33. Schneider G (2002) Trends in virtual computational library design. *Curr Med Chem* 9: 2095–2102.
34. Verdok ML, Cole JC, Hartshorn ML, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. *Proteins* 52: 609–623.
35. Van de Waterbeemd H, Gifford E (2003) ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* 2: 192–204.
36. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, et al. (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45: 2615–2623.
37. Hedley PL, Jørgensen P, Schlamowitz S, Wangari R, Moolman-Smook J, et al. (2009) The genetic basis of long QT and short QT syndromes: A mutation update. *Human Mutation* 30: 1486–1511.
38. Vandenberg JJ, Walker BD, Campbell TJ (2001) HERG K<sup>+</sup> channels: Friend or foe. *Trends Pharmacol Sci* 22: 240–246.
39. Chiesa N, Rosati B, Arcangeli A, Olivetto M, Wanke E (1997) A novel role for HERG K<sup>+</sup> channels: spike-frequency adaptation. *J Physiol* 501: 313–318.
40. Aronov AM (2005) Predictive in silico modeling for hERG channel blockers. *Drug Discov Today* 10: 149–155.
41. Khanna V, Ranganathan S (2011) Structural diversity of biologically interesting datasets: a scaffold analysis approach. *J Cheminform* 3: 30.
42. Block S, Gerkens P, Peulen O, Jolois O, Mingéot-Leclercq MP, et al. (2005) Induction of apoptosis in Human promyelocytic leukemia cells by a natural trichyalbane diterpene. *Anticancer Res* 25: 363–368.
43. Hoet S, Stéveny C, Block S, Opperdoes F, Colson P, et al. (2004) Alkaloids from *Cassytha filiformis* and related aporphines: antitrypanosomal activity, cytotoxicity, and interaction with DNA and topoisomerases. *Planta Med* 70: 407–413.
44. Okogun JI, Adesomoku AA, Adesida GA, Lindner HJ, Habermehl G (1982) Roseanolone: A new diterpene from *Hypoestes rosea*. *Z Naturforsch* 37c: 558–561.
45. Ojo-Amaize EA, Nchekwube EJ, Cottam HB, Bai R, Verdier-Pinard P, et al. (2002) Hypoestoxide, a natural nonmutagenic diterpenoid with antiangiogenic and antitumor activity: possible mechanisms of action. *Cancer Res* 62: 4007–4014.
46. Ojo-Amaize EA, Kapahi P, Kakkanaiah VN, Takahashi T, Shalom-Barak T, et al. (2001) Hypoestoxide, a novel anti-inflammatory natural diterpene, inhibits the activity of IκB kinase. *Cellular Immunology* 209:149–157.
47. Salama MM, Kandil ZA, Islam WT (2012) Cytotoxic compounds from the leaves of *Gaillardia aristata* Pursh. growing in Egypt. *Nat Prod Res* 26 (22): 2057–2062.
48. Jackson SJ, Houghton PJ, Retas S, Photiou A (2000) In vitro cytotoxicity of norvirburtal and isopinnatal from *Kigelia pinnata* against cancer cell lines. *Planta Med* 66: 758–761.
49. Schwikkard S, van Heerden RF (2002) Anti-malarial activity of plant metabolites. *Nat Prod Rep* 19(6):675–692.
50. Magadula JJ, Erasto P (2009) Bioactive natural products derived from the East African flora. *Nat Prod Rep* 26:1535–1554.
51. Titanji VPK, Zofou D, Ngenyen MN (2008) The anti-malarial potential of medicinal plants used for the treatment of malaria in Cameroonian Folk Medicine. *Afr J Trad CAM* 5(3):302–321.
52. Kuete V, Efferth T (2010) Cameroonian medicinal plants: pharmacology and derived natural products. *Frontiers in Pharmacology* 1:123.
53. Ntie-Kang F, Lifongo LL, Mbaze LM, Ekwelle N, Owono Owono LC, et al. (2013) Cameroonian medicinal plants: a bioactivity versus ethnobotanical survey and chemotaxonomic classification. *BMC Complement Altern Med* 13(1): 147.
54. Hoet S, Opperdoes F, Brun R, Quetin-Leclercq J (2004) Natural products active against African trypanosomes: a step towards new drugs. *Nat Prod Rep* 21: 253–264.
55. Kuete V (2010) Potential of Cameroonian plants and derived products against microbial infections: a review. *Planta Med* 76:1479–1491.
56. Banzouzi JT, Prado R, Menan H, Valentin A, Roumestan C, et al. (2004) Studies on medicinal plants of Ivory Coast: Investigation of *Sida acuta* for in vitro antiplasmodial activities and identification of an active constituent. *Phytomedicine* 11: 338–341.
57. Chibale K, Davies-Coleman M, Masimirembwa C (2012) Drug discovery in Africa: impacts of genomics, natural products, traditional medicines, insights into medicinal chemistry, and technology platforms in pursuit of new drugs. Springer.
58. pan-ANAPL: pan-African Natural Products Library. Available: <http://www.linkedin.com/groups/pANAPL-4098579/about/> Accessed 2013 June 13.
59. Mahmoudi N, De Julian-Ortiz JV, Ciceron L, Galvez J, Mazier D, et al. (2006) Identification of new antimalarial drugs by linear discriminant analysis and topological virtual screening. *J Antimicrob Chemother* 57: 489–497.
60. Bickii J, Njifutie N, Foyere JA, Basco LK, Ringwald P (2000) In vitro antimalarial activity of limonoids from *Khaya grandifoliola* C.D.C. (Meliaceae). *J Ethnopharmacol* 69: 27–33.
61. Chemical Computing Group Inc.: *Molecular Operating Environment Software*. Montreal; 2010.
62. Halgren TA (1996) Merck molecular forcefield. *J Comput Chem* 17: 490–641.
63. Schrödinger: *LigPrep software*, version 2.5, LLC, New York, NY; 2011.
64. Schrödinger: *Maestro*, version 9.2, LLC, New York, NY; 2011.
65. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, et al. (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J Chem Theory Comput* 6: 1509–1519.
66. Jørgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118(45): 11225–11236.
67. Jørgensen WL, Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* 110(6): 1657–1666.
68. Schrödinger: *QikProp*, version 3.4, LLC, New York, NY, 2011.
69. Jørgensen WL, Duffy EM (2002) Prediction of drug solubility from structure. *Adv Drug Deliv Rev* 54: 355–366.
70. Duffy EM, Jørgensen WL (2000) Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. *J Am Chem Soc* 122: 2878–2888.
71. Jørgensen WL, Duffy EM (2000) Prediction of drug solubility from Monte Carlo simulations. *Bioorg Med Chem Lett* 10: 1155–1158.
72. Colmenarejo G, Alvarez-Pedraglio A, Lavandera JL (2001) Cheminformatic models to predict binding affinities to human serum albumin. *J Med Chem* 44: 4370–4378.
73. Luco JM (1999) Prediction of brain-blood distribution of a large set of drugs from structurally derived descriptors using partial least squares (PLS) modelling. *J Chem Inf Comput Sci* 39: 396–404.
74. Kelder J, Grootenhuis PD, Bayada DM, Delbresine LP, Ploemen JP (1999) Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm Res* 16: 1514–1519.
75. Ajay, Bermis GW, Murkco MA (1999) Designing libraries with CNS activity. *J Med Chem* 42: 4942–4951.
76. Yazdaniyan M, Glynn SL, Wright JL, Hawi A (1998) Correlating partitioning and caco-2 cell permeability of structurally diverse small molecular weight compounds. *Pharm Res* 15: 1490–1494.
77. Irvine JD, Takahashi L, Lockhart K, Cheong J, Tolan JW, et al. (1999) MDCK (Madin-Darby canine kidney) cells: a tool for membrane permeability screening. *J Pharm Sci* 88: 28–33.
78. Stenberg P, Norinder U, Luthman K, Artursson P (2001) Experimental and computational screening models for the prediction of intestinal drug absorption. *J Med Chem* 44: 1927–1937.
79. Cavalli A, Poluzzi E, De Ponti F, Recanatini M (2002) Toward a pharmacophore for drugs inducing the long QT syndrome: Insights from a CoMFA Study of HERG K<sup>+</sup> channel blockers. *J Med Chem* 45: 3844–3853.

80. De Ponti F, Poluzzi E, Montanaro N (2001) Organising evidence on QT prolongation and occurrence of Torsades de Pointes with non-antiarrhythmic drugs: a call for consensus. *Eur J Clin Pharmacol* 57: 185–209.
81. Potts RO, Guy RH (1992) Predicting skin permeability. *Pharm Res* 9: 663–669.
82. Potts RO, Guy RH (1995) A predictive algorithm for skin permeability: The effects of molecular size and hydrogen bond activity. *Pharm Res* 12: 1628–1633.
83. ChemAxon: JChem software, version 5.11.3; 2012. Website. Available: <https://www.chemaxon.com/jchem/doc/user/LibMCS.html>/Accessed 2013 June 13.
84. Lewell XQ, Judd DB, Watson SP, Hann MM (1998) RECAP-retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci* 38: 511–522.
85. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna; 2012. Website <http://www.R-project.org/> Accessed 2013 June 13.