

# SCIENTIFIC REPORTS

OPEN

## SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules

Received: 05 October 2016

Accepted: 13 January 2017

Published: 03 March 2017

Antoine Daina<sup>1</sup>, Olivier Michielin<sup>1,2,3</sup> & Vincent Zoete<sup>1</sup>

To be effective as a drug, a potent molecule must reach its target in the body in sufficient concentration, and stay there in a bioactive form long enough for the expected biologic events to occur. Drug development involves assessment of absorption, distribution, metabolism and excretion (ADME) increasingly earlier in the discovery process, at a stage when considered compounds are numerous but access to the physical samples is limited. In that context, computer models constitute valid alternatives to experiments. Here, we present the new SwissADME web tool that gives free access to a pool of fast yet robust predictive models for physicochemical properties, pharmacokinetics, drug-likeness and medicinal chemistry friendliness, among which in-house proficient methods such as the BOILED-Egg, iLOGP and Bioavailability Radar. Easy efficient input and interpretation are ensured thanks to a user-friendly interface through the login-free website <http://www.swissadme.ch>. Specialists, but also nonexpert in cheminformatics or computational chemistry can predict rapidly key parameters for a collection of molecules to support their drug discovery endeavours.

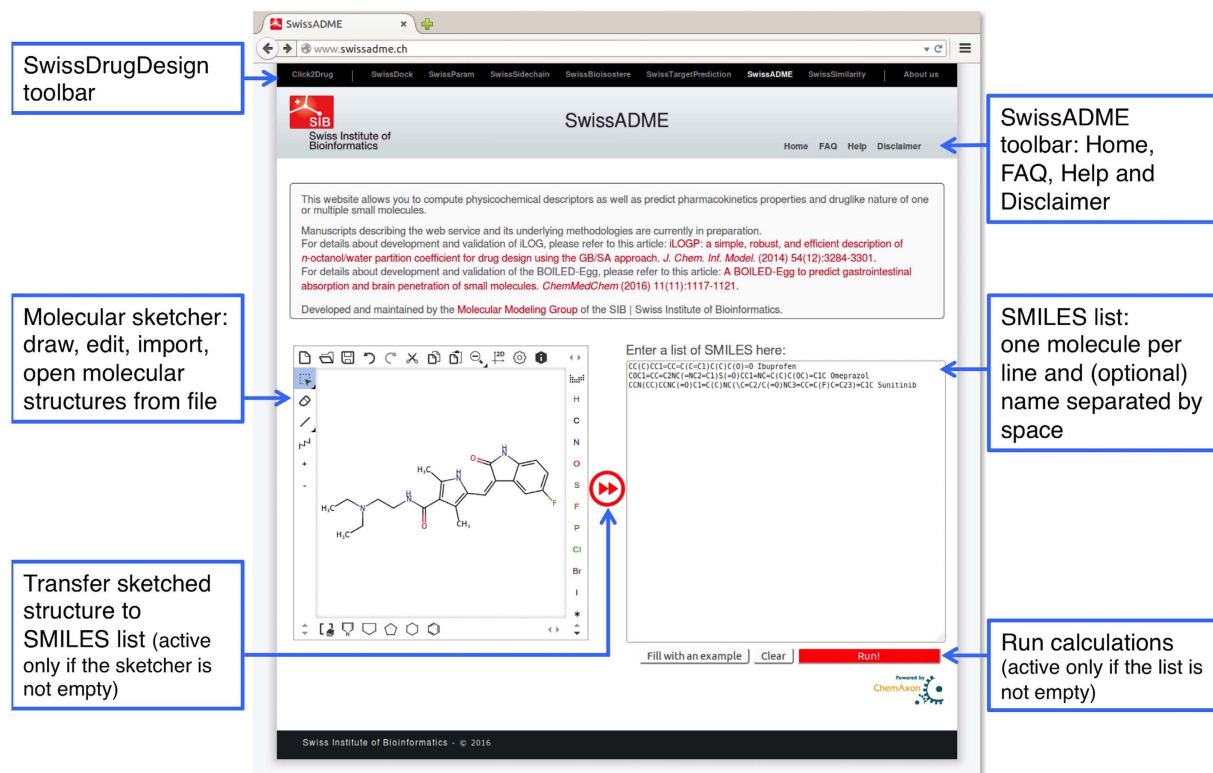
During the time- and resource-consuming processes of drug discovery and development, a large number of molecular structures are evaluated according to very diverse parameters in order to steer the selection of which chemicals to synthesize, test and promote, with the final goal to identify those with the best chance to become an effective medicine for the patients. The molecules must show high biological activity together with low toxicity. Equally important is the access to and concentration at the therapeutic target in the organism. The traditional way to consider pharmacokinetics (i.e. the fate of a therapeutic compound in the organism) is to break down the various effects that impact the access to the target into individual parameters. In turn, these ADME parameters (for Absorption, Distribution, Metabolism and Excretion) can be evaluated separately by dedicated methods. It has been demonstrated that early estimation of ADME in the discovery phase reduces drastically the fraction of pharmacokinetics-related failure in the clinical phases<sup>1</sup>. Computer models have been fostered as a valid alternative to experimental procedures for prediction of ADME, especially at initial steps, when investigated chemical structures are numerous but the availability of compounds is scarce<sup>2</sup>.

A large variety of *in silico* methods share the objective of predicting ADME parameters from molecular structure<sup>3</sup>. Noteworthy, the pioneer work of Lipinski *et al.* examined orally active compounds to define physicochemical ranges for high probability to be an oral drug (i.e. the drug-likeness)<sup>4</sup>. This so-called *Rule-of-five* delineated the relationship between pharmacokinetic and physicochemical parameters.

Whereas physicochemical parameters give a global description of the structure, molecules can be directly described by substructure searches. These techniques are at the root of Structural Alert<sup>5</sup>, the PAINS<sup>6</sup> or the Lilly MedChem<sup>7</sup> filters applied to cleanse chemical libraries from compounds most likely unstable, reactive, toxic, or prone to interfere with biological assays because unspecific frequent hitters, dyes or aggregators<sup>8</sup>.

Cheminformaticians developed different molecular descriptors mined from chemical structures. One of the most popular examples is the molecular fingerprint (FP), which consists in a sequence of bits defining

<sup>1</sup>SIB Swiss Institute of Bioinformatics, Molecular Modeling Group, Quartier Sorge, Bâtiment Génomopode, CH-1015 Lausanne, Switzerland. <sup>2</sup>Department of Oncology, Centre Hospitalier Universitaire Vaudois, CH-1015 Lausanne, Switzerland. <sup>3</sup>Ludwig Institute for Cancer Research, University of Lausanne, CH-1015 Lausanne, Switzerland. Correspondence and requests for materials should be addressed to O.M. (email: [olivier.michielin@sib.swiss](mailto:olivier.michielin@sib.swiss)) or V.Z. (email: [vincent.zoete@sib.swiss](mailto:vincent.zoete@sib.swiss))



**Figure 1. SwissADME submission page.** The actual input is a list of SMILES, which contains one molecule per line with an optional name separated by a space. Molecules can be directly pasted or typed in SMILES format, or inserted through the molecular sketcher. The latter enables importing from databases, opening a local file or drawing a 2D chemical structure to be transferred to the list by clicking on the double-arrow button. When the list of molecules is ready to be submitted, the user can start the calculations by clicking on the “Run” button.

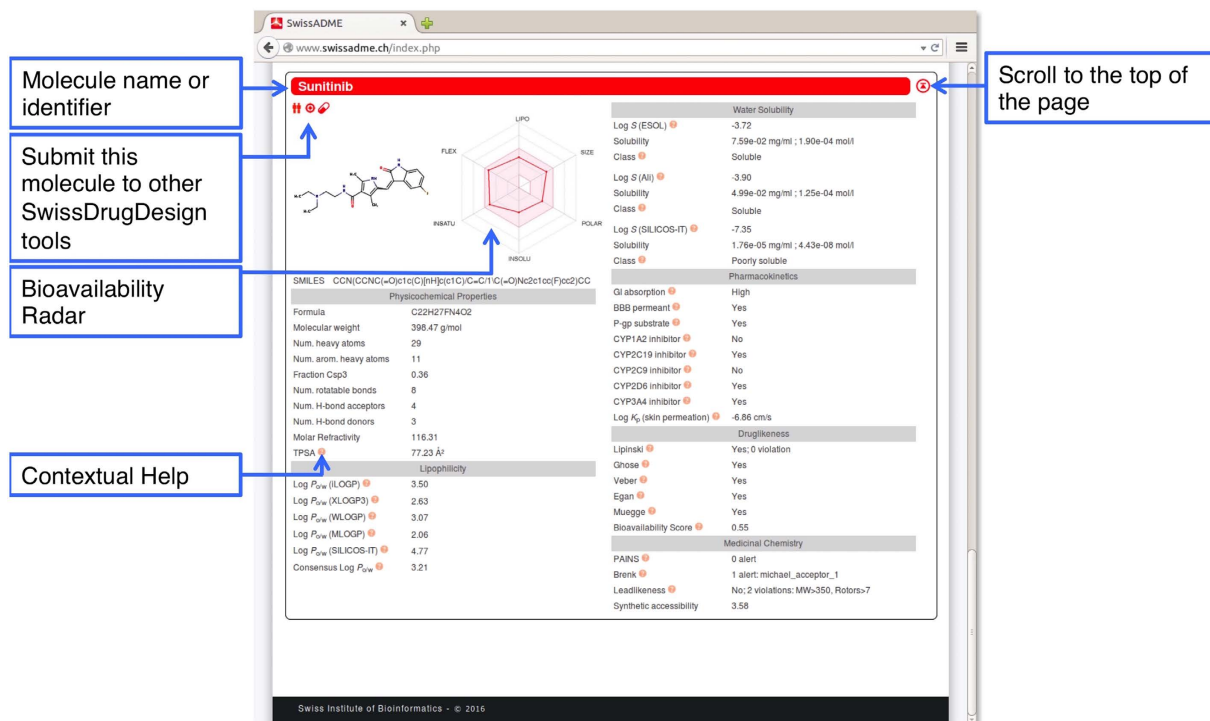
the presence or absence of chemical features in a molecule. The FP2 method<sup>9</sup> is one archetype of topological (or path-based) FP, which considers all fragments of the molecular structure following a linear path up to a given number of bonds. Every possible path is hashed to generate the bit string (i.e. the FP). A major advantage of FP is the efficiency by which computers handle such bit strings<sup>10</sup>, allowing for instance large-scale virtual screening or the rapid estimation of synthetic accessibility of molecules<sup>11</sup>. FP are also employed in classification models for ADME behaviours built by support vector machine (SVM) or Bayesian techniques<sup>12</sup>. Remarkably, computer-aided drug design (CADD) has been a pioneer field for the application of such machine learning technologies<sup>13</sup>.

Because most freely available *in silico* ADME tools focus on one specific property or model only, while generalist ADME packages are commercial software, we felt the need of gathering what we consider the most relevant computational methods to provide a global appraisal of the pharmacokinetics profile of small molecules. The methods were selected for robustness, speed, but also and importantly for ease of interpretation to enable efficient translation to medicinal chemistry through molecular design. Some of them were adapted with open-source algorithms to ensure freedom to operate for the global scientific community. Others are our own models developed and tested on purpose. When applicable, multiple predictions of the same parameter are provided to allow for a consensus view of a given property.

The SwissADME web tool presented here is freely accessible at <http://www.swissadme.ch> and meant for user-friendly submission and easy analysis of the results, also for nonexpert in CADD. Compared to the state-of-the-art of free web-based tools for ADME and pharmacokinetics (e.g. pk-CSM<sup>14</sup> and admetSAR<sup>15</sup>) and apart from unique access to proficient methods (e.g. iLOGP<sup>16</sup> or the BOILED-Egg<sup>17</sup>), SwissADME strong points are, non-exhaustively: different input methods, computation for multiple molecules, and the possibility to display, save and share results per individual molecule or through global intuitive and interactive graphs. Finally, SwissADME is integrated in the SwissDrugDesign workspace. One-click interoperability gives access to various CADD tools developed by the Molecular Modeling Group of the SIB Swiss Institute of Bioinformatics, e.g. ligand-based virtual screening (SwissSimilarity<sup>18</sup>), biotarget prediction (SwissTargetPrediction<sup>19</sup>), molecular docking (SwissDock<sup>20</sup>), bioisosteric design (SwissBioisostere<sup>21</sup>), or molecular mechanics (SwissParam<sup>22</sup>).

### Submission Web page

Accessing <http://www.swissadme.ch> in a web browser displays directly the submission page of SwissADME, where molecules to be estimated for ADME, physicochemistry, drug-likeness, pharmacokinetics and medicinal chemistry friendliness properties can be input. As shown in Fig. 1, a black toolbar at the top of the Web



**Figure 2.** Computed parameter values are grouped in the different sections of the one-panel-per-molecule output (Physicochemical Properties, Lipophilicity, Pharmacokinetics, Drug-likeness and Medicinal Chemistry). The panel is headed by the molecule name and an up-arrow button to scroll to the top of the page. The molecule is first described by its chemical structure and canonical SMILES together with the Bioavailability Radar (see Fig. 3). Contextual help can be displayed by leaving the mouse over the radar or different question mark icons next to some parameters.

page allows the user to navigate within the different SwissDrugDesign tools. A second bar gives access to different information regarding SwissADME, among which the *FAQ* and *Help* pages as well as legal disclaimer and contacts.

The input zone itself comprises a molecular sketcher based on ChemAxon's Marvin JS (<http://www.chemaxon.com>) that enables the user to import (from a file or an external database), draw, and edit a 2D chemical structure, and to transfer it to a list of molecules. This list, on the right-hand side of the submission page, is the actual input for computation. It can be edited as a standard text, allowing for typing or pasting SMILES. The list is made to contain one input molecule per line, defined by a SMILES and optionally a name separated by a space. If name is omitted, SwissADME will automatically provide an identifier.

Noteworthy, both buttons for transferring the sketch to SMILES list and for running the computation are dynamic, in the sense that they are active only if the action is possible. At the time of writing, one can expect a result in 1 to 5 seconds for a drug-like molecule.

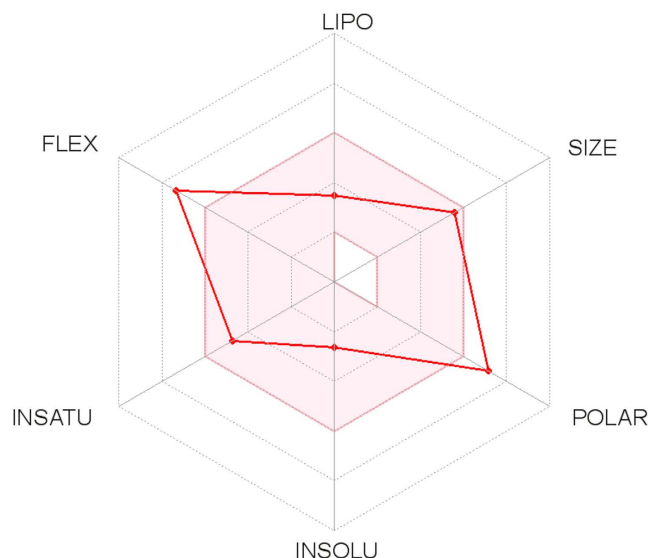
Examples can be loaded in the SMILES list by clicking on the “Fill with an example” button.

### One-panel-per-molecule Output

The output panels are loaded in the same Web page. There is one panel compiling all values for each molecule. It is filled immediately after calculation completion, one molecule after the other. This way it is possible to inspect the results for the first compounds without waiting for the whole list to be treated. This one-panel-per-molecule (Fig. 2) is headed by the molecule name and divided into different sections.

**Chemical Structure and Bioavailability Radar.** The first section, including two-dimensional chemical structure and canonical SMILES, is located below the title (Fig. 2). It shows on which chemical form the predictions were calculated (refer to *Computational Methods*). Moreover, our *Bioavailability Radar* is displayed for a rapid appraisal of drug-likeness (refer to Fig. 3). Six physicochemical properties are taken into account: lipophilicity, size, polarity, solubility, flexibility and saturation. A physicochemical range on each axis was defined by descriptors adapted from refs 23 and 24 and depicted as a pink area in which the radar plot of the molecule has to fall entirely to be considered drug-like. Leaving the mouse over the radar gives further information about the descriptors (see also *Physicochemical Properties* and *Computational Methods*).

**Physicochemical Properties.** Simple molecular and physicochemical descriptors like molecular weight (MW), molecular refractivity (MR), count of specific atom types and polar surface area (PSA) are compiled in this section. The values are computed with OpenBabel<sup>9</sup>, version 2.3.0. The PSA is calculated using the fragmental



**Figure 3. The Bioavailability Radar enables a first glance at the drug-likeness of a molecule.** The pink area represents the optimal range for each properties (lipophilicity: XLOGP3 between  $-0.7$  and  $+5.0$ , size: MW between  $150$  and  $500$  g/mol, polarity: TPSA between  $20$  and  $130$  Å<sup>2</sup>, solubility:  $\log S$  not higher than  $6$ , saturation: fraction of carbons in the  $sp^3$  hybridization not less than  $0.25$ , and flexibility: no more than  $9$  rotatable bonds). In this example, the compound is predicted not orally bioavailable, because too flexible and too polar.

technique called topological polar surface area (TPSA), considering sulfur and phosphorus as polar atoms<sup>25</sup>. This has proven a useful descriptor in many models and rules to quickly estimate some ADME properties, especially with regards to biological barrier crossing such as absorption and brain access<sup>17</sup>.

**Lipophilicity.** The partition coefficient between *n*-octanol and water ( $\log P_{o/w}$ ) is the classical descriptor for *Lipophilicity*. It has a dedicated section in SwissADME due to the critical importance of this physicochemical property for pharmacokinetics drug discovery<sup>26,27</sup>. Many computational methods for  $\log P_{o/w}$  estimation were developed with diverse performance on various chemical sets. Common practice is to use multiple predictors either to select the most accurate methods for a given chemical series or to generate consensus estimation. The models behind the predictors should be as diverse as possible to increase the prediction accuracy through consensus  $\log P_{o/w}$ <sup>28</sup>. In that regard, SwissADME gives access to five freely available predictive models; i.e. XLOGP3, an atomistic method including corrective factors and knowledge-based library<sup>29</sup>; WLOGP, our own implementation of a purely atomistic method based on the fragmental system of Wildman and Crippen<sup>30</sup>; MLOGP, an archetype of topological method relying on a linear relationship with 13 molecular descriptors implemented from refs 31 and 32; SILICOS-IT, an hybrid method relying on 27 fragments and 7 topological descriptors (<http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/filter-it/1.0.2/filter-it.html>, accessed June 2016); and finally iLOGP, our in-house physics-based method relying on free energies of solvation in *n*-octanol and water calculated by the Generalized-Born and solvent accessible surface area (GB/SA) model. iLOGP was benchmarked on two drug or drug-like external sets and performed equally as or better than six well-established predictors<sup>16</sup>. The consensus  $\log P_{o/w}$  is the arithmetic mean of the values predicted by the five proposed methods.

**Water Solubility.** Having a soluble molecule greatly facilitates many drug development activities, primarily the ease of handling and formulation<sup>33</sup>. Moreover, for discovery projects targeting oral administration, solubility is one major property influencing absorption<sup>34</sup>. As well, a drug meant for parenteral usage has to be highly soluble in water to deliver a sufficient quantity of active ingredient in the small volume of such pharmaceutical dosage<sup>35</sup>. Two topological methods to predict *Water Solubility* are included in SwissADME. The first one is an implementation of the ESOL model<sup>36</sup> and the second one is adapted from Ali *et al.*<sup>37</sup>. Both differ from the seminal general solubility equation<sup>38</sup> since they avoid the melting point parameter; the latter being challenging to predict. Moreover they demonstrate strong linear correlation between predicted and experimental values ( $R^2 = 0.69$  and  $0.81$ , respectively). SwissADME third predictor for solubility was developed by SILICOS-IT. The linear correlation coefficient of this fragmental method corrected by molecular weight is  $R^2 = 0.75$  (<http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/filter-it/1.0.2/filter-it.html>, accessed June 2016).

All predicted values are the decimal logarithm of the molar solubility in water ( $\log S$ ). SwissADME also provides solubility in mol/l and mg/ml along with qualitative solubility classes (please refer to *Computational Methods*).

**Pharmacokinetics.** Specialized models, whose predictions are compiled in the *Pharmacokinetics* section, evaluate individual ADME behaviours of the molecule under investigation.

Model	SwissADME					Previous models					
	TR/TS <sup>[a]</sup>	ACC <sub>CV</sub> <sup>[b]</sup>	AUC <sub>CV</sub> <sup>[c]</sup>	ACC <sub>ext</sub> <sup>[d]</sup>	AUC <sub>ext</sub> <sup>[e]</sup>	TR/TS	ACC <sub>CV</sub>	AUC <sub>CV</sub>	ACC <sub>ext</sub>	AUC <sub>ext</sub>	Reference
P-gp substrate	1033/415	0.72	0.77	0.89	0.94	544/n.c.	0.71 <sup>[f]</sup>	n.c.	n.c.	n.c.	53
						484/300	0.64	n.c.	0.59	n.c.	71
						332/n.c.	0.74 <sup>[f]</sup>	0.77 <sup>[f]</sup>	n.c.	n.c.	15
						212/120	0.74	n.c.	0.88	n.c.	57
CYP1A2 inhibitor	9145/3000	0.83	0.90	0.84	0.91	9145/3000	n.c.	n.c.	0.88	0.95	54
						12099/2804	0.82 <sup>[f]</sup>	n.c.	0.68	0.81	77
						7208/7128	0.88 <sup>[g]</sup>	n.c.	n.c.	0.93	55
CYP2C19 inhibitor	9272/3000	0.80	0.86	0.80	0.87	9272/3000	n.c.	n.c.	0.85	0.91	54
						11885/2691	0.79 <sup>[f]</sup>	n.c.	0.81	0.84	77
						6038/5923	0.81 <sup>[g]</sup>	n.c.	n.c.	0.89	55
CYP2C9 inhibitor	5940/2075	0.78	0.85	0.71	0.81	8720/3000	n.c.	n.c.	0.83	0.90	54
						12130/2579	0.78 <sup>[f]</sup>	n.c.	0.89	0.86	77
						6627/6530	0.83 <sup>[g]</sup>	n.c.	n.c.	0.89	55
CYP2D6 inhibitor	3664/1068	0.79	0.85	0.81	0.87	9726/3000	n.c.	n.c.	0.84	0.88	54
						11881/2860	0.84 <sup>[f]</sup>	n.c.	0.88	0.88	77
						7788/7761	0.90 <sup>[g]</sup>	n.c.	n.c.	0.85	55
CYP3A4 inhibitor	7518/2579	0.77	0.85	0.78	0.86	8893/5135	n.c.	n.c.	0.84	0.92	54
						11536/7025	0.78 <sup>[f]</sup>	n.c.	0.76	0.78	77
						2334/6738	0.81 <sup>[g]</sup>	n.c.	n.c.	0.87	55

**Table 1. Statistical performance of SVM classification models for substrate or inhibitor of pharmacokinetics-relevant protein, P-gp and CYP.** <sup>a</sup>Number of molecules in the training set (TR) and in the test set (TS); <sup>b</sup>10-fold cross-validation accuracy; <sup>c</sup>10-fold cross-validation area under receiver operating characteristic (ROC) curve; <sup>d</sup>external validation accuracy; <sup>e</sup>external validation area under ROC curve; <sup>f</sup>5-fold cross-validation; <sup>g</sup>7-fold cross-validation.

One model is a multiple linear regression, which aims at predicting the skin permeability coefficient ( $K_p$ ). It is adapted from Potts and Guy<sup>39</sup>, who found  $K_p$  linearly correlated with molecular size and lipophilicity ( $R^2 = 0.67$ ). The more negative the  $\log K_p$  (with  $K_p$  in cm/s), the less skin permeant is the molecule.

The predictions for passive human gastrointestinal absorption (HIA) and blood-brain barrier (BBB) permeation both consist in the readout of the BOILED-Egg model<sup>17</sup>, an intuitive graphical classification model, which can be displayed in the SwissADME result page by clicking the red button appearing below the sketcher when all input molecules have been processed (refer to *Graphical Output*). Other binary classification models are included, which focus on the propensity for a given small molecule to be substrate or inhibitor of proteins governing important pharmacokinetic behaviours.

The knowledge about compounds being substrate or non-substrate of the permeability glycoprotein (P-gp, suggested the most important member among ATP-binding cassette transporters or ABC-transporters) is key to appraise active efflux through biological membranes, for instance from the gastrointestinal wall to the lumen or from the brain<sup>40</sup>. One major role of P-gp is to protect the central nervous system (CNS) from xenobiotics<sup>41</sup>. Importantly as well, P-gp is overexpressed in some tumour cells and leads to multidrug-resistant cancers<sup>42</sup>.

Also essential is the knowledge about interaction of molecules with cytochromes P450 (CYP). This superfamily of isoenzymes is a key player in drug elimination through metabolic biotransformation<sup>43</sup>. It has been suggested that CYP and P-gp can process small molecules synergistically to improve protection of tissues and organisms<sup>44</sup>. One can estimate that 50 to 90% (depending on the authors) of therapeutic molecules are substrate of five major isoforms (CYP1A2, CYP2C19, CYP2C9, CYP2D6, CYP3A4)<sup>45,46</sup>. Inhibition of these isoenzymes is certainly one major cause of pharmacokinetics-related drug-drug interactions<sup>47,48</sup> leading to toxic or other unwanted adverse effects due to the lower clearance and accumulation of the drug or its metabolites<sup>49</sup>. Numerous inhibitors of the CYP isoforms have been identified. Some are affecting different CYP isoforms, while other compounds show selectivity for specific isoenzymes<sup>50</sup>. It is therefore of great importance for drug discovery to predict the propensity with which the molecule will cause significant drug interactions through inhibition of CYPs, and to determine which isoforms are affected.

SwissADME enables the estimation for a chemical to be substrate of P-gp or inhibitor of the most important CYP isoenzymes. We applied the support vector machine algorithm (SVM)<sup>51</sup> on meticulously cleansed large datasets of known substrates/non-substrates or inhibitors/non-inhibitors (for details, see *Computational Methods*). In similar contexts, SVM was found to perform better than other machine-learning algorithms for binary classification<sup>40,52</sup>. The models return “Yes” or “No” if the molecule under investigation has higher probability to be substrate or non-substrate of P-gp (respectively inhibitor or non-inhibitor of a given CYP). The statistical performance of the classification models is given in Table 1, in comparison with previous SVM models on the same targets. We restricted the benchmark to state-of-the-art methods, published after 2010.

The quantification of the models performance is not straightforwardly comparable, because the training sets are different, most of the published models are less than 10-fold cross-validated and some statistical parameters

are missing. Nevertheless, the SwissADME classifiers are competitive with previous models in term of robustness, with cross-validation accuracy ( $ACC_{CV}$ ) grossly at the same level. Furthermore, cross-validated areas under receiver operating characteristic (ROC) curves ( $AUC_{CV}$ ) are equal to the corresponding values found in the literature. Likewise, external prediction power is difficult to compare, as each test set includes different molecules. However, the predictive capacity of SwissADME classifiers is grossly equivalent to the related SVM methods, both in terms of external accuracy ( $ACC_{ext}$ ) and external area under ROC curve ( $AUC_{ext}$ ). The models for which external validation was not found (P-gp of refs 15 and 53) have to be taken with extreme caution since they possibly suffer from overfitting biases. Noteworthy, some of the published models (e.g. CYP2C9 and CYP3A4 of ref. 54 or CYP2C9 of ref. 55) were built on severely unbalanced training sets and tested on clearly unbalanced external sets. As demonstrated by Carbon-Mangels *et al.*<sup>56</sup> the relevance of machine-learning classification methods, and especially SVM, are negatively impacted by datasets with one significantly more populated class. In that case, accuracy measurements are overestimated and prone to mislead the construction and the evaluation of the model. Moreover, some SVM models were published with small training and test sets (P-gp of refs 15 and 57), which imply questionable capacity of generalization and broadness of applicability domains. We emphasize that for SwissADME classifiers, both training and test sets were carefully cleansed and checked for size, diversity and balance between classes. Furthermore, our SVM models rely merely on molecular and physicochemical descriptors generated by SwissADME. We believe that this improves robustness and sustainability of the underlying methodologies. In particular, not using molecular fingerprints, molecular graphs or other structural descriptions can be an handicap to generate high statistical values but should also limit overfitting biases and yield more generalist predictive models, not necessarily influenced by specific chemical scaffolds or moieties. In our practice, these well-performing models able to estimate important ADME behaviours are of great support for pharmacokinetics optimization and evaluation of small molecules.

**Drug-likeness.** As defined earlier, “drug-likeness” assesses qualitatively the chance for a molecule to become an oral drug with respect to bioavailability. Drug-likeness was established from structural or physicochemical inspections of development compounds advanced enough to be considered oral drug-candidates. This notion is routinely employed to perform filtering of chemical libraries to exclude molecules with properties most probably incompatible with an acceptable pharmacokinetics profile. This SwissADME section gives access to five different rule-based filters, with diverse ranges of properties inside of which the molecule is defined as drug-like. These filters often originate from analyses by major pharmaceutical companies aiming to improve the quality of their proprietary chemical collections. The Lipinski (Pfizer) filter is the pioneer rule-of-five implemented from ref. 4. The Ghose (Amgen), Veber (GSK), Egan (Pharmacia) and Muegge (Bayer) methods were adapted from refs 58–61, respectively. Multiple estimations allow consensus views or selection of methods best fitting the end-user’s specific needs in terms of chemical space or project-related demands. Any violation of any rule described here appears explicitly in the output panel.

The Abbot Bioavailability Score<sup>62</sup> is similar but seeks to predict the probability of a compound to have at least 10% oral bioavailability in rat or measurable Caco-2 permeability. This semi-quantitative rule-based score relying on total charge, TPSA, and violation to the Lipinski filter defines four classes of compounds with probabilities of 11%, 17%, 56% or 85%. Like the other methods in this section, it primary focuses on the fast screening of chemical libraries, to select the best molecules to be purchased, synthesized or promoted at a further stage of a medicinal chemistry project.

**Medicinal Chemistry.** The purpose of this section is to support medicinal chemists in their daily drug discovery endeavours. Two complementary pattern recognition methods allow for identification of potentially problematic fragments. PAINS (for pan assay interference compounds, a.k.a. frequent hitters or promiscuous compounds) are molecules containing substructures showing potent response in assays irrespective of the protein target. Such fragments, yielding false positive biological output, have been identified by Baell *et al.*<sup>6</sup> in analysing six orthogonal assays and breaking down the molecules active on 2 or more assays into 481 recurrent fragments, considered as potentially leading to promiscuous compounds. SwissADME returns warnings if such moieties are found in the molecule under evaluation.

Besides, we implemented Structural Alert, which consists in a list of 105 fragments identified by Brenk *et al.*<sup>5</sup> to be putatively toxic, chemically reactive, metabolically unstable or to bear properties responsible for poor pharmacokinetics. In SwissADME, it is possible to have a chemical description of the problematic fragments found in a given molecule by flying over the “question mark” icon appearing after the fragment list. This is implemented for both PAINS and Brenk filters. By applying these and other physicochemical filters to design screening libraries, Brenk *et al.*<sup>5</sup> observed that most of the remaining compounds satisfy criteria for “leadlikeness”. This concept is similar to drug-likeness, yet focusing on physicochemical boundaries defining a good lead, i.e. a molecular entity suitable for optimization. By definition, leads are subjected to chemical modifications that will most likely increase size and lipophilicity<sup>63</sup>. As a consequence, leads are required to be smaller and less hydrophobic than drug-like molecules. Since it is crucial for a chemist to judge whether a given molecule is suitable to initiate lead optimization, in addition to structural filters, we implemented a rule-based method for leadlikeness, which was adapted from ref. 64.

One of the key aspects of CADD activities is to help the selection of the most promising virtual molecules that will be synthesized and submitted to biological assays or other experiments. Synthetic accessibility (SA) is a major factor to consider in this selection process. Obviously, for a reasonable number of molecules, medicinal chemists are the best able to determine SA. However, when too many molecular structures prevent an expert evaluation, *in silico* estimation can be used for pre-filtering. Ertl & Schuffenhauer<sup>11</sup> proposed a fingerprint-based approach for SA estimation but including closed-source information about fingerprint definition that prevents a straightforward implementation in our tool open to the scientific community. As a consequence, we have built

	Test set		SwissADME			Original models published with the test set			
	#molecule <sup>[a]</sup>	#chemist <sup>[b]</sup>	MAE <sup>[c]</sup>	RMSE <sup>[d]</sup>	r <sup>[e]</sup>	MAE	RMSE	r	Reference
1	40	9	0.63	0.81	0.96	0.91	1.13	0.94	11
2	114	4	1.02	1.33	0.62	1.07	1.41	0.62	65

**Table 2. Statistical performance of synthetic accessibility (SA) scores on two external test sets.** <sup>a</sup>Number of external molecules; <sup>b</sup>number of chemist marks; <sup>c</sup>mean average error; <sup>d</sup>root mean square error; <sup>e</sup>linear correlation coefficient.

our own fragmental method by analysing more than 13 millions compounds immediately deliverable by vendors. We assumed that the most frequent molecular fragments (technically, FP2 bits, refer to *Computational Methods*) in this large collection indicates a probably high SA, while rare fragments imply a difficult synthesis. For a given molecule, the fragmental contributions to SA are summed and corrected by the terms describing size and complexity, such as macrocycles, chiral centres, or spiro functions as defined by Ertl & Schuffenhauer<sup>11</sup>. After normalization, the SA Score ranges from 1 (very easy) to 10 (very difficult). To assess the performance of the method developed for SwissADME, we retrieved two test sets of SA previously published. Both sets involved external molecules, whose difficulty of synthesis was marked from 1 to 10 by nine<sup>11</sup> and four<sup>65</sup> medicinal chemists, respectively. The averaged expert score can then be compared to an *in silico* SA Score. As seen in Table 2, the predictive capacity of all three methods appears very dependent of the test set. Indeed the SAs of set 1, smaller and evaluated by more chemists, turned out to be much more robustly predictable than set 2. Human evaluation of synthetic complexity is undeniably subjective and relies on individual chemist's training and experience. However, significant linear correlation and small errors — especially with SwissADME SA Score that outperformed the reference methods on both sets with smaller errors, and equal or higher linear correlation coefficients — demonstrate how this simple and fast methodology can help prioritizing molecules to synthesize.

### Graphical output

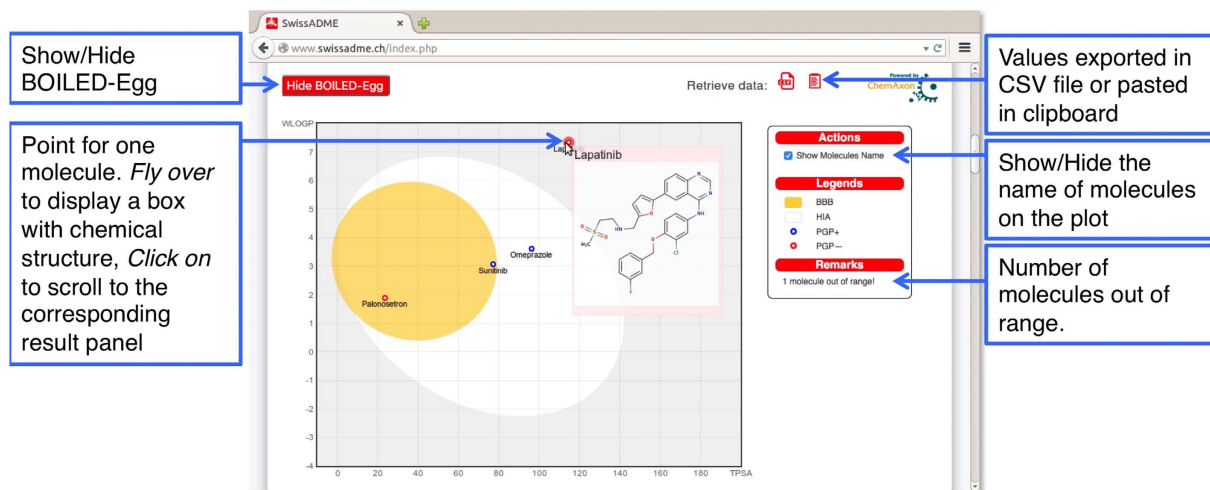
After all calculations completed, the “Show BOILED-Egg” red button appears below the sketcher to display the graphical output on the same page (as exemplify in Fig. 4). This consists primarily in the BOILED-Egg, an intuitive method to predict simultaneously two key ADME parameters, i.e. the passive gastrointestinal absorption (HIA) and brain access (BBB). Although conceptually very simple as it relies on two physicochemical descriptors only (WLOGP and TPSA, for lipophilicity and apparent polarity), this classification model was built with extreme care regarding statistical significance and robustness<sup>17</sup>. As shown in Fig. 4, the egg-shaped classification plot includes the yolk (i.e. the physicochemical space for highly probable BBB permeation) and the white (i.e. the physicochemical space for highly probable HIA absorption). Both compartments are not mutually exclusive and the outside grey region stands for molecules with properties implying predicted low absorption and limited brain penetration. In practice, the BOILED-Egg has proven straightforward interpretation and efficient translation to molecular design in a variety of drug discovery settings. Whereas the predictive power of the BOILED-Egg is broad in term of chemical space, it is restricted to passive penetration through gastro-intestinal wall and BBB. We took benefit of its implementation within SwissADME to enrich the graphical output with the prediction of P-gp substrate, which is the most important active efflux mechanism involved in those biological barriers<sup>40</sup> (refer to the SVM model described in *Pharmacokinetics*). As a result, the user conveniently obtains on the same graph a global evaluation about passive absorption (inside/outside the white), passive brain access (inside/outside the yolk) and active efflux from the CNS or to the gastrointestinal lumen by colour-coding: blue dots for P-gp substrates (PGP+) and red dots for P-gp non-substrate (PGP-).

Contrary to the one-panel-per-molecule concept for other parameters, the graphical output includes prediction for all molecules submitted to SwissADME, thus additional capacities were implemented to enable interactive navigation and easy evaluation. Flying over a specific point makes a semi-transparent box appear including the name and structure of the molecule. Clicking on a specific point makes the page scroll to the corresponding output panel including all predictions for the molecule. Getting back to the graphical output is achieved by clicking on the red up-arrow on the top-right corner of the panel. Moreover, on the right of the plot are displayed possible actions (at the moment, to show the name of the molecules on the graph, only), legends and possible remarks (the number of molecules outside the range of the plot). The user may want to hide the plot by clicking the “Hide BOILED-Egg” red button.

### Help, saving and interoperability

SwissADME is user-friendly and has been conceived for a variety of users, including nonexpert in cheminformatics or computational chemistry. Efforts were put to insure ease of input and interpretability of output. These aspects are supported by the *Help* page, which provides technical guidance on the usage of the website and by the *FAQ*, where we collected not only technical but also scientific and general questions. Contextual help about specific items can be displayed by leaving the mouse over question marks or some boxes on the pages.

Moreover, SwissADME is an integral part of the SwissDrugDesign program, an ambitious initiative driven by the Molecular Modeling Group of the SIB Swiss Institute of Bioinformatics that aims at providing a collection of free Web-based tools covering many aspects of CADD. As such, SwissADME users can access the different Web sites, through the black toolbar at the top of the page. By clicking on the name of the tool, a new tab opens with the corresponding submission page. We have undertaken the implementation of more advanced interoperability capabilities. At the time of writing this manuscript, three icons were placed top left of the output panel (below



**Figure 4.** The BOILED-Egg<sup>17</sup> allows for intuitive evaluation of passive gastrointestinal absorption (HIA) and brain penetration (BBB) in function of the position of the molecules in the WLOGP-versus-TPSA referential. The white region is for high probability of passive absorption by the gastrointestinal tract, and the yellow region (yolk) is for high probability of brain penetration. Yolk and white areas are not mutually exclusive. In addition the points are coloured in blue if predicted as actively effluxed by P-gp (PGP+) and in red if predicted as non-substrate of P-gp (PGP-). For an interactive analysis, the user can leave the mouse over a dot to show the structure of the molecule and click on the dot to scroll to the corresponding output panel. In this example, Lapatinib is predicted as not absorbed and not brain penetrant (outside the Egg), Omeprazole is predicted well-absorbed but not accessing the brain (in the white) and PGP+ (blue dot), Sunitinib is predicted as passively crossing the BBB (in the yolk), but pumped-out from the brain (blue dot), and Palonosetron is predicted as brain-penetrant (in the yolk) and not subject to active efflux (red dot). One molecule is predicted not absorbed and not BBB permeant because outside of the range of the plot (streptomycin with a TPSA of 331.43 Å<sup>2</sup> and a WLOGP of -7.74).

the name of the molecule, see Fig. 2). They allow one-click submission of the molecule to SwissTargetPrediction (for reverse screening to predict protein target, by clicking on the red target), SwissSimilarity (for ligand-based virtual screening, by clicking on the red twins) and to re-submit the molecule to SwissADME (for instance to get a single compound graphical output, by clicking the red pill). Reversely, other SwissDrugDesign websites include the SwissADME pill icon for estimation of ADME, pharmacokinetics, drug-likeness and medicinal chemistry friendliness regarding a small molecule output of any CADD process.

Importantly, the user has the opportunity to save or share SwissADME results in files (CSV) or by copying the values in the clipboard of the local computer and pasting in text or spreadsheet applications for further processing.

## Conclusions

The SwissADME Web tool enables the computation of key physicochemical, pharmacokinetic, drug-like and related parameters for one or multiple molecules. In one hand, efforts were put in the backend to embed free open-access and fast predictive models showing statistical significance, predictive power, intuitive interpretation, and straightforward translation to molecular design. These models are adapted from published renowned approaches or in-house original methods, specially developed and thoroughly benchmarked. On the other hand, we focused on an ergonomic and user-friendly graphical interface for the cost- and login-free Web site <http://www.swissadme.ch>. The latter enables easy input and efficient analysis of the output through interactive capabilities, and does not require any prior knowledge in CADD. Moreover, interoperability allows for direct access to other SwissDrugDesign web tools, including SwissSimilarity<sup>66</sup> (virtual screening), SwissBioisostere<sup>21</sup> (ligand-based design), SwissTargetPrediction<sup>19</sup> (prediction of biotargets), SwissDock<sup>20</sup> (molecular docking), SwissSideChain<sup>67</sup> (protein modeling) and SwissParam<sup>22</sup> (molecular mechanics).

As a result, SwissADME has been designed to support the entire community (specialists and nonexperts) in their drug discovery endeavours.

## Computational methods

**Programming and scripting.** The SwissADME website was written in HTML, PHP5, and JavaScript, whereas the backend of computation was mainly coded in Python 2.7. The use of additional libraries or software for specific tasks is mentioned in the corresponding paragraph.

**Submission Page.** The molecule inputted through the sketcher Marvin JS (version 16.4.18, 2016, [www.chemaxon.com](http://www.chemaxon.com)) are converted into SMILES by JChem Web Services (version 14.9.29, 2013, [www.chemaxon.com](http://www.chemaxon.com)) installed on one of our servers. This on-the-fly conversion allows seamless paste of SMILES in the input list. The user has the possibility to edit this list as a standard text, e.g. to modify SMILES or add a name to the molecule.



Upon calculation submission by clicking the “Run” button, the SMILES of each molecule is canonicalised by OpenBabel (version 2.3.0, 2012, <http://openbabel.org>)<sup>9</sup> and processed individually. Several actions are performed through JChem Web Services APIs. First hydrogen atoms are added to the molecular structure, which is dearomatized (i.e. kekulised), neutralised and checked by the Standardizer API. Then a tridimensional conformation is generated through the StringMolExport function with the Clean3D option and stored in MOL2 format. Besides, a two-dimensional image created through the MolConverter API is displayed on demand when scrolling the output web page.

**One-panel-per-molecule Output.** The *Bioavailability Radar* in the first section of the One-panel-per-molecule output complements the two-dimensional image from the JChem webserver and the canonical SMILES calculated by OpenBabel. We use the JpGraph PHP library (version 3.5.0b1, 2016, <http://jgraph.net>) to produce the radar plot, which bears six axes for six important properties for oral bioavailability. Each property is defined by a descriptor of SwissADME and a range of optimal values is depicted as a pink area. The latter is inspired from commonly accepted bioavailability and drug-likeness guidelines<sup>23,24</sup>. For saturation, the ratio of sp<sup>3</sup> hybridized carbons over the total carbon count of the molecule (Fraction Csp3) should be at least 0.25. For size, the molecular weight (MW calculated by OpenBabel) should be between 150 and 500 g/mol. For polarity, the TPSA<sup>25</sup> should be between 20 and 130 Å<sup>2</sup>. For solubility, log S (calculated with the ESOL model<sup>36</sup>) should not exceed 6. For lipophilicity, XLOGP3<sup>29</sup> should be in the range from -0.7 to +6.0. For flexibility, the molecule should not have more than 9 rotatable bonds. To be estimated as drug-like, the red line of the compound under study must be fully included in the pink area. Any deviation represents a suboptimal physicochemical property for oral bioavailability.

All descriptors and molecular parameters of the *Physicochemical Properties* section are computed through the OpenBabel API (version 2.3.0, 2012, <http://openbabel.org>)<sup>9</sup>. Noteworthy, the topological polar surface area (TPSA) is strictly based on the fragmental system provided by Ertl *et al.*<sup>25</sup> including polar sulfur and phosphorus atoms.

Multiple freely available computational methods to predict *n*-octanol/water partition coefficient (log  $P_{o/w}$ ) values are made available in the *Lipophilicity* section. iLOGP (for implicit log  $P$ ) is an in-house physics-based methods relying on Gibbs free energy of solvation calculated by GB/SA in water and *n*-octanol. Generalized-born (GB) parameters are computed through the GBMV2 method<sup>68</sup> and solvent-accessible surface area (SA) is the analytical approximation generated by CHARMM (version c36b1, 2011, <https://www.charmm.org>)<sup>69</sup>. The iLOGP implemented in SwissADME corresponds to Model9 of the seminal publication<sup>16</sup>, which was trained on 11,993 molecules ( $r = 0.72$ ,  $MAE = 0.89$ , and  $RMSE = 1.14$  against experimental log  $P$ ). 5-fold crossvalidation ensured robustness ( $q^2_{CV} = 0.52$ ,  $MAE_{CV} = 0.89$ , and  $RMSE_{CV} = 1.14$ ) and external test benchmarks showed the excellent predictive power and extended applicability domain compared to well-established methods. XLOGP3 values are obtained through the command-line Linux program (version 3.2.2, courtesy of CCBG, Shanghai Institute of Organic Chemistry) including the knowledge-based corrections<sup>29</sup>. WLOGP is our own implementation of the atomistic method developed by Wildman and Crippen<sup>30</sup>. MLOGP values are computed through an in-house implementation of Moriguchi's topological method<sup>31,32</sup>. SILICOS-IT is the log  $P_{o/w}$  estimation returned by executing the FILTER-IT program (version 1.0.2, 2013, <http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/filter-it/1.0.2/filter-it.html>). Finally, SwissADME gives a consensus log  $P_{o/w}$  value, which is the arithmetic mean of the five predictive values mentioned above.

Similarly to lipophilicity, the *Water Solubility* section includes multiple predictive methods for the user to choose between the most accurate model for a given chemical series and an averaged consensus value. The ESOL model<sup>36</sup> is a QSPR model establishing the linear relationship between log S and five molecular parameters, i.e. MW, the number of rotatable bonds, the fraction of aromatic heavy atoms and Daylight's CLOGP. Because the lipophilicity descriptor is not freely available, the implementation of ESOL in SwissADME replaces CLOGP by XLOGP3 as parameter in the linear equation to predict log S. XLOGP3 is known to perform well on external datasets and to return similar predictions as CLOGP<sup>28</sup>. The other three parameters were computed with OpenBabel. Likewise, Ali *et al.*<sup>37</sup> linked log S with log  $P_{o/w}$  and TPSA. The model implemented in SwissADME corresponds to the model 3 of the original publication, with XLOGP3 as lipophilicity descriptor. The third solubility method available in SwissADME is the log S estimated by the FILTER-IT program (version 1.0.2, 2013, <http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/filter-it/1.0.2/filter-it.html>). This prediction is based on a system of 16 fragmental contributions modulated by the squared root of MW. All three models are predicting log S values, which are also translated within SwissADME into solubility in mol/l and mg/ml. Finally a qualitative estimation of the solubility class is given according to the following log S scale: insoluble < -10 < poorly < -6 < moderately < -4 < soluble < -2 < very < 0 < highly.

The *Pharmacokinetics* section proposes one linear method for skin permeation, which relies on the simple QSPR model by Potts and Guy<sup>39</sup> linking the decimal logarithm of the skin permeability coefficient (log  $K_p$  in cm/s) with MW and log  $P_{o/w}$ . The model implemented in SwissADME uses XLOGP3 as lipophilicity descriptor. Besides, most of the models in this section are machine-learning binary classifiers for important ADME behaviours. Passive gastro-intestinal (HIA) absorption and blood-brain barrier (BBB) permeation are predicted with the BOILED-Egg model, which defines favourable and unfavourable zones in the log  $P_{o/w}$  versus PSA physicochemical space for passive diffusion through both physiological barriers<sup>17</sup>. The classification showed 10-fold cross-validation accuracy of 92% and 88% for BBB and HIA, respectively (refer to *Graphical Output*).

Six other classification models are part of the *Pharmacokinetics* section to predict the propensity of the molecule under investigation to be substrate or inhibitor of important pharmacokinetics-related proteins, for which large diverse and balanced datasets were retrieved and meticulously cleansed. For P-glycoprotein1 (P-gp), the training set consists of 521 substrates and 512 non-substrates extracted from the Metrabase database<sup>70</sup> (<http://www-metrabase.ch.cam.ac.uk>, accessed January 2016), whereas the test set was obtained from ref. 71.

To ensure truly external validation, molecules overlapping with the training set were removed from the test set, which finally includes 215 substrates and 200 non-substrates. For CYP major isoforms, all datasets were those of Veith *et al.*<sup>50</sup> and downloaded from the PubChem database<sup>72</sup> (<http://pubchem.ncbi.nlm.nih.gov>, accessed February 2016). In case of unbalanced dataset (all except CYP1A2 and CYP2C19), sufficient chemical diversity was guaranteed by clustering with the Ward method and a reciprocal nearest neighbour (RNN) algorithm<sup>73</sup>, the more populated class to lessen. The number of molecules (described by circular fingerprints) of the large class is reduced by defining clusters with the JKlustor program (version 14.9.29, 2014, <http://www.chemaxon.com>). Only the centre of each cluster (i.e. the molecule that has the smallest sum of dissimilarities to the other molecules in the cluster) is included in the training or test set to balance. As a result, the training sets involved respectively 4301 CYP1A2 inhibitors and 4844 CYP1A2 non-inhibitors; 4284 CYP2C19 inhibitors and 4988 CYP2C19 non-inhibitors; 2940 CYP2C9 inhibitors and 3000 CYP2C9 non-inhibitors; 1814 CYP2D6 inhibitors and 1850 CYP2D6 non-inhibitors; and 3758 CYP3A4 inhibitors and 3760 CYP3A4 non-inhibitors. The test sets involved respectively 1412 CYP1A2 inhibitors and 1588 CYP1A2 non-inhibitors; 1386 CYP2C19 inhibitors and 1614 CYP2C19 non-inhibitors; 1020 CYP2C9 inhibitors and 1055 CYP2C9 non-inhibitors; 528 CYP2D6 inhibitors and 540 CYP2D6 non-inhibitors; and 1289 CYP3A4 inhibitors and 1290 CYP3A4 non-inhibitors.

SwissADME's backend calculations were ran to generate 50 molecular and physicochemical descriptors per molecule (described in the Supplementary Table S1). For a given model, a descriptor was rejected if non-zero values for all molecules in the training set are less than 20% or if the coefficient of variation is less than 3%. In case of correlation higher than 0.9 between remaining descriptors, a selection is made based on F-score. The selected descriptors for each model are shown in Supplementary Tables S2–S7. These tables also include the minimum and maximum values for each descriptor among all molecules used in the training. This enables beholding the broadness of physicochemical space involved and the applicability domain of the SVM models. The predictive capability of each model can be further appraised on Supplementary Table S8, where external accuracy was split in sensitivity and specificity to ensure that positive and negative molecules are predicted with the same level of robustness. The final training and test sets with selected descriptors were normalized and the respective model ready to be built. First, the libSVM support vector machine python library (version 3.20, 2015, <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>)<sup>74</sup> was used for multi-step grid-based optimization of the best coefficients for the above-selected descriptors as well as for the soft-margin permissivity (C) and the hyper-parameter ( $\gamma$ ) of the RBF Gaussian kernel function. The 10-fold crossvalidated accuracy ( $ACC_{CV}$ ) for each model was so maximized and  $AUC_{CV}$  was calculated. In a second step, the so-built models were used on the external test sets (normalized according to the training set) in order to evaluate predictive power in terms of external accuracy ( $ACC_{ext}$ ) and  $AUC_{ext}$ . All final SVM models were stored in separate files, which are read through the libSVM API upon SwissADME job submission.

*The Drug-likeness section includes six rule-based methods.* The Lipinski rule-of-five is exactly as described in ref. 4 including  $MLOGP < 4.15$  as lipophilicity threshold. The Ghose filter is the method detailed in the original publication<sup>58</sup>, where the atomic log  $P$  is calculated with WLOGP. The very simple yet efficient Veber filter is implemented directly from the seminal paper<sup>59</sup>. The Egan filter is yielded from the Egan Egg<sup>60,75</sup>, but the closed-source ALOGP98 was replaced by WLOGP<sup>17</sup>. The Muegge filter<sup>61</sup> was adapted to fit SwissADME implementation and usage by calculating XLOGP3 as lipophilicity descriptor. The Bioavailability Score was implemented without changes from Martin *et al.*<sup>62</sup>. Similarly, the leadlikeness filter included in the *Medicinal Chemistry* section was adapted from the original rule<sup>64</sup> by using XLOGP3 as lipophilicity descriptor.

Both methods for identification of problematic fragments within the *Medicinal Chemistry* section, i.e. PAINS<sup>6</sup> and Brenk's Structural Alert<sup>5</sup>, were implemented using the SMARTS recognition capability of OpenBabel API. The SMARTS definitions for PAINS were retrieved from the Filter-it distribution (version 1.0.2, 2013, <http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/filter-it/1.0.2/filter-it.html>). Little formatting and cleansing were needed to obtain a screenable collection of 481 fragments. Brenk provided directly the SMARTS descriptions of 105 unwanted chemical groups in the supplementary material of the seminal paper<sup>5</sup>.

SwissADME Synthetic Accessibility (SA) Score is based primarily on the assumption that the frequency of molecular fragments in 'really' obtainable molecules correlates with the ease of synthesis. The fragmental contribution to SA should be favourable for frequent chemical moieties and unfavourable for rare moieties. We examined the 'All Now' subset of ZINC database<sup>76</sup> (version 12, 2014, <http://zinc.docking.org/subsets/all-now> accessed April 2015) including 12'782'590 compounds immediately deliverable by vendors. This collection was not submitted to any other filter. FP2 fingerprints of all molecules were computed by OpenBabel<sup>9</sup> thus generating 12'782'590 bit strings (i.e. the fingerprints). The frequency of occupancy of each of the 1024 bits for the entire library was calculated and its contribution to SA obtained by applying natural logarithm to the normalized bit count. As a result, this fragmental system of bits allows extremely fast evaluation of any input molecule by reading its FP2 string. Fragmental contribution of bit occupancy is summed and linearly modulated by corrective factors, which are penalties for size (MW) and complexity. This latter is based on SMARTS recognition of chiral centres, spiro functions, bridged rings and macrocycles (more than 8-membered rings). The coefficients of the corrective terms are those defined by Ertl & Schuffenhauer<sup>11</sup>. Finally the score is normalized to range from 1 (very easy) to 10 (very difficult to synthesize). As crude as it may seem, SwissADME SA Score is extremely fast and performing slightly better than two similar methods previously published<sup>11,65</sup> (refer to Table 2).

**Graphical output.** The graphical output of SwissADME consists of the BOILED-Egg directly implemented from ref. 17 to predict passive diffusion through HIA and BBB by position in a WLOGP-*versus*-TPSA physicochemical space. The HIA model was trained on 660 well-absorbed and non-absorbed molecules and a 10-fold crossvalidation returned an accuracy ( $ACC_{CV}$ ) of 92% and a Matthews correlation coefficient ( $MCC_{CV}$ ) of 0.65. The BBB model was built on a training set of 260 permeant and non-permeant molecules and displayed excellent

internal statistics as well:  $ACC_{CV} = 88\%$  and  $MCC_{CV} = 0.75$ . The plot is enriched by color-coding referring to P-gp active efflux predicted by a SVM classifier (refer to *Pharmacokinetics*). We aimed at providing a global and efficient analysis of absorption and brain access for multiple molecules at once. For more advanced interactive features, the graph is generated in JavaScript using the Flot plotting library (version 0.8, 2015, <http://www.flotcharts.org/>). The user has the possibility to visualize the position of the molecules between the different BOILED-Egg compartments and their propensity of being substrate of P-gp by coloured points: blue for substrate (PGP+) and red for non-substrate (PGP-). The dots are also the mean to visualize name and chemical structure of the molecules (generated on-the-fly by the dedicated Chemaxon web server as described above) by leaving the mouse over the dot. As well, it is possible to access the corresponding output panel summarizing all calculated values for the molecule by clicking on the corresponding dot. Getting back to the BOILED-Egg is as simple as clicking on the up arrow at the top-right corner of the panel. All these functionalities, as well as interoperability with other SwissDrugDesign web tools are coded in PHP and Javascript.

## References

- Hay, M., Thomas, D. W., Craighead, J. L., Economides, C. & Rosenthal, J. Clinical development success rates for investigational drugs. *Nature Biotechnol.* **32**, 40–51 (2014).
- Dahlin, J. L., Ingles, J. & Walters, M. A. Mitigating risk in academic preclinical drug discovery. *Nature Rev. Drug Discov.* **14**, 279–294 (2015).
- Tian, S. *et al.* The application of in silico drug-likeness predictions in pharmaceutical research. *Adv Drug Deliv Rev* **86**, 2–10 (2015).
- Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Deliv. Rev.* **46**, 3–26 (2001).
- Brenk, R. *et al.* Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem* **3**, 435–444 (2008).
- Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740 (2010).
- Bruns, R. F. & Watson, I. A. Rules for Identifying Potentially Reactive or Promiscuous Compounds. *J. Med. Chem.* **55**, 9763–9772 (2012).
- Irwin, J. J. *et al.* An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **58**, 7076–7087 (2015).
- O’Boyle, N. M. *et al.* OpenBabel: An open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
- Cereto-Massaguà, A. *et al.* Molecular fingerprint similarity search in virtual screening. *Methods* **71**, 58–63 (2015).
- Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **1**, 8 (2009).
- Clark, A. M. *et al.* Open Source Bayesian Models. 1. Application to ADME/Tox and Drug Discovery Datasets. *J. Chem. Inf. Model.* **55**, 1231–1245 (2015).
- Mitchell, J. B. O. Machine learning methods in chemoinformatics. *WIREs Comput. Mol. Sci.* **4**, 468–481 (2014).
- Pires, D. E. V., Blundell, T. L. & Ascher, D. B. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem.* **58**, 4066–4072 (2015).
- Cheng, F. *et al.* admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. *J. Chem. Inf. Model.* **52**, 3099–3105 (2012).
- Daina, A., Michielin, O. & Zoete, V. iLOGP: A Simple, Robust, and Efficient Description of n-Octanol/Water Partition Coefficient for Drug Design Using the GB/SA Approach. *J. Chem. Inf. Model.* **54**, 3284–3301 (2014).
- Daina, A. & Zoete, V. A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem* **11**, 1117–1121 (2016).
- Zoete, V., Daina, A., Bovigny, C. & Michielin, O. SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening. *J. Chem. Inf. Model.* **56**, 1399–1404 (2016).
- Gfeller, D. *et al.* SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res.* **42**, W32–W38 (2014).
- Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein–small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **39**, W270–7 (2011).
- Wirth, M., Zoete, V., Michielin, O. & Sauer, W. H. B. SwissBioisostere: a database of molecular replacements for ligand design. *Nucleic Acids Res.* **41**, D1137–43 (2013).
- Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32**, 2359–2368 (2011).
- Ritchie, T. J., Ertl, P. & Lewis, R. The graphical representation of ADME-related molecule properties for medicinal chemists. *Drug Discov. Today* **16**, 65–72 (2011).
- Lovering, F., Bikker, J. & Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **52**, 6752–6756 (2009).
- Ertl, P., Rohde, B. & Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **43**, 3714–3717 (2000).
- Pliska, V., Testa, B. & van de Waterbeemd, H. In *Lipophilicity in Drug Action and Toxicology* 1–6 (Wiley-VCH Verlag GmbH, 1996).
- Arnott, J. A. & Planey, S. L. The influence of lipophilicity in drug discovery and design. *Expert Opin. Drug Discov.* **7**, 863–875 (2012).
- Mannhold, R., Poda, G. I. & Ostermann, C. Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds. *J. Pharm. Sci.* **98**, 861–893 (2009).
- Cheng, T. *et al.* Computation of Octanol–Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J. Chem. Inf. Model.* **47**, 2140–2148 (2007).
- Wildman, S. A. & Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Model.* **39**, 868–873 (1999).
- Moriguchi, I., Shuichi, H., Liu, Q., Nakagome, I. & Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull.* **40**, 127–130 (1992).
- Moriguchi, I., Shuichi, H., Nakagome, I. & Hirano, H. Comparison of reliability of log P values for Drugs calculated by several methods. *Chem. Pharm. Bull.* **42**, 976–978 (1994).
- Ritchie, T. J., Macdonald, S. J. F., Peace, S., Pickett, S. D. & Luscombe, C. N. Increasing small molecule drug developability in sub-optimal chemical space. *Med. Chem. Commun.* **4**, 673 (2013).
- Ottaviani, G. *et al.* What is modulating solubility in simulated intestinal fluids? *Eur J Pharm Sci* **41**, 452–457 (2010).
- Savjani, K. T., Gajjar, A. K. & Savjani, J. K. Drug solubility: importance and enhancement techniques. *ISRN Pharm* **2012**, 195727 (2012).
- Delaney, J. S. ESOL: Estimating Aqueous Solubility Directly from Molecular Structure. *J. Chem. Inf. Model.* **44**, 1000–1005 (2004).

37. Ali, J., Camilleri, P., Brown, M. B., Hutt, A. J. & Kirton, S. B. Revisiting the general solubility equation: in silico prediction of aqueous solubility incorporating the effect of topographical polar surface area. *J. Chem. Inf. Model.* **52**, 420–428 (2012).
38. Yalkowsky, S. H. & Valvani, S. C. Solubility and partitioning I: Solubility of nonelectrolytes in water. *J. Pharm. Sci.* **69**, 912–922 (1980).
39. Potts, R. O. & Guy, R. H. Predicting Skin Permeability. *Pharm. Res.* **09**, 663–669 (1992).
40. Montanari, F. & Ecker, G. F. Prediction of drug-ABC-transporter interaction—Recent advances and future challenges. *Adv. Drug Deliv. Rev.* **86**, 17–26 (2015).
41. Szakács, G., Váradi, A., Ozvegy-Laczka, C. & Sarkadi, B. The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). *Drug Discov. Today* **13**, 379–393 (2008).
42. Sharom, F. J. ABC multidrug transporters: structure, function and role in chemoresistance. *Pharmacogenomics* **9**, 105–127 (2008).
43. Testa, B. & Kraemer, S. D. The Biochemistry of Drug Metabolism – An Introduction - Testa - 2007 - Chemistry & Biodiversity - Wiley Online Library. *Chem. Biodivers* (2007).
44. van Waterschoot, R. A. B. & Schinkel, A. H. A critical analysis of the interplay between cytochrome P450 3A and P-glycoprotein: recent insights from knockout and transgenic mice. *Pharmacological Reviews* **63**, 390–410 (2011).
45. Wolf, C. R., Smith, G. & Smith, R. L. Pharmacogenetics. *British Medical Journal* (2000).
46. Di, L. The role of drug metabolizing enzymes in clearance. *Expert Opin. Drug Metab. Toxicol.* **10**, 379–393 (2014).
47. Hollenberg, P. F. Characteristics and common properties of inhibitors, inducers, and activators of CYP enzymes. *Drug Metab. Rev.* **34**, 17–35 (2002).
48. Huang, S.-M. *et al.* New era in drug interaction evaluation: US Food and Drug Administration update on CYP enzymes, transporters, and the guidance process. *J. Clin. Pharmacol.* **48**, 662–670 (2008).
49. Kirchmair, J. *et al.* Predicting drug metabolism: experiment and/or computation? *Nature Rev. Drug Discov.* **14**, 387–404 (2015).
50. Veith, H. *et al.* Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nature Biotechnol.* **27**, 1050–1055 (2009).
51. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
52. Mishra, N. K., Agarwal, S. & Raghava, G. P. Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule. *BMC Pharmacol.* **10**, 8 (2010).
53. Sedykh, A. *et al.* Human intestinal transporter database: QSAR modeling and virtual profiling of drug uptake, efflux and interactions. *Pharm. Res.* **30**, 996–1007 (2015).
54. Rostkowski, M., Spjuth, O. & Rydberg, P. WhichCyp: prediction of cytochromes P450 inhibition. *Bioinformatics* **29**, 2051–2052 (2013).
55. Sun, H., Veith, H., Xia, M., Austin, C. P. & Huang, R. Predictive models for cytochrome p450 isozymes based on quantitative high throughput screening data. *J. Chem. Inf. Model.* **51**, 2474–2481 (2011).
56. Carbon-Mangels, M. & Hutter, M. C. Selecting Relevant Descriptors for Classification by Bayesian Estimates: A Comparison with Decision Trees and Support Vector Machines Approaches for Disparate Data Sets. *Mol. Inf.* **30**, 885–895 (2011).
57. Wang, Z. *et al.* P-glycoprotein substrate models using support vector machines based on a comprehensive data set. *J. Chem. Inf. Model.* **51**, 1447–1456 (2011).
58. Ghose, A. K., Viswanadhan, V. N. & Wendoloski, J. J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1**, 55–68 (1999).
59. Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
60. Egan, W. J., Merz, K. M. & Baldwin, J. J. Prediction of Drug Absorption Using Multivariate Statistics. *J. Med. Chem.* **43**, 3867–3877 (2000).
61. Muegge, I., Heald, S. L. & Brittelli, D. Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* **44**, 1841–1846 (2001).
62. Martin, Y. C. A Bioavailability Score. *J. Med. Chem.* **48**, 3164–3170 (2005).
63. Hann, M. M. & Keserü, G. M. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. *Nature Rev. Drug Discov.* **11**, 355–365 (2012).
64. Teague, S., Davis, A., Leeson, P. & Oprea, T. The Design of Leadlike Combinatorial Libraries. *Angew. Chem. Int. Ed. Engl.* **38**, 3743–3748 (1999).
65. Fukunishi, Y., Kurosawa, T., Mikami, Y. & Nakamura, H. Prediction of synthetic accessibility based on commercially available compound databases. *J. Chem. Inf. Model.* **54**, 3259–3267 (2014).
66. Zoete, V., Daina, A., Bovigny, C. & Michielin, O. SwissSimilarity. A web tool for low to ultra high-throughput ligand-based virtual screening. *J. Chem. Inf. Model.* **58**, 1399–1404 (2016).
67. Gfeller, D., Michielin, O. & Zoete, V. SwissSidechain: a molecular and structural database of non-natural sidechains. *Nucleic Acids Res.* **41**, D327–32 (2013).
68. Lee, M. S., Feig, M., Salsbury, F. R. & Brooks, C. L. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **24**, 1348–1356 (2003).
69. Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
70. Mak, L. *et al.* Metrabase: a cheminformatics and bioinformatics database for small molecule transporter data analysis and (Q)SAR modeling. *J. Cheminform.* **7**, 31 (2015).
71. Poongavanam, V. V., Haider, N. N. & Ecker, G. F. G. Fingerprint-based in silico models for the prediction of P-glycoprotein substrates and inhibitors. *Bioorg. Med. Chem.* **20**, 5388–5395 (2012).
72. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res.* **44**, D1202–13 (2016).
73. Ward, J. H. Jr. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* (1963).
74. Chang, C.-C. & Lin, C.-J. LIBSVM. *ACM Trans. Intell. Syst. Technol.* **2**, 1–27 (2011).
75. Egan, W. J. & Lauri, G. Prediction of intestinal permeability. *Adv. Drug Deliv. Rev.* **54**, 273–289 (2002).
76. Irwin, J. J. & Shoichet, B. K. ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177–182 (2005).
77. Cheng, F. *et al.* Classification of Cytochrome P450 Inhibitors and Noninhibitors Using Combined Classifiers. *J. Chem. Inf. Model.* **51**, 996–1011 (2011).

## Acknowledgements

The authors would like to thank the SIB Swiss Institute of Bioinformatics ([www.sib.swiss](http://www.sib.swiss)) for funding and its high-performance computing centre (Vital-IT, [www.vital-it.ch](http://www.vital-it.ch)) for computational resources. This work was also partly supported by the MelanomX grant from SystemsX, the Swiss Initiative in Systems Biology, and by the Solidar-Immun Foundation to V.Z. We acknowledge ChemAxon Ltd. ([www.chemaxon.com](http://www.chemaxon.com)) for the academic licensing agreement as well as the Shanghai Institute of Organic Chemistry (SIOC), China for the XLOGP license.

## Author Contributions

A.D. and V.Z. wrote the backend code, the website and the manuscript. A.D. built the models, performed testing and benchmarks. O.M. and V.Z. supervised the project. All authors approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Daina, A. *et al.* SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 42717; doi: 10.1038/srep42717 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017