

# Explanation Information

## ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties

Guo-Li Xiong, Zhen-Xing Wu, Jia-Cai Yi, Li Fu, Zhi-Jiang Yang, Chang-Yu Hsieh, Ming-Zhu Yin, Xiang-Xiang Zeng, Cheng-Kun Wu, Ai-Ping Lu, Xiang Chen, Ting-Jun Hou\*, Dong-Sheng Cao\*

\*Correspondence: [oriental-cds@163.com](mailto:oriental-cds@163.com), [tingjunhou@zju.edu.cn](mailto:tingjunhou@zju.edu.cn)

### 1. Physicochemical property

#### ***Molecular Weight***

Contain hydrogen atoms. Optimal:100~600, based on Drug-Like Soft rule.

#### ***Volume***

Van der Waals volume.

#### ***Density***

Density = MW / Volume

#### ***nHA***

Number of hydrogen bond acceptors. Sum of all O and N. Optimal: 0~12, based on Drug-Like Soft rule.

#### ***nHD***

Number of hydrogen bond donors. Sum of all OHs and NHs. Optimal:0~7, based on Drug-Like Soft rule.

#### ***nRot***

Number of rotatable bonds. In some situation Amide C-N bonds are not considered because of their high rotational energy barrier. Optimal:0~11, based on Drug-Like Soft rule.

#### ***nRing***

Number of rings. Smallest set of smallest rings. Optimal:0~6, based on Drug-Like Soft rule.

### **MaxRing**

Number of atoms in the biggest ring. Number of atoms involved in the biggest system ring. Optimal:0~18, based on Drug-Like Soft rule.

### **nHet**

Number of heteroatoms. Number of non-carbon atoms (hydrogens included). Optimal:1~15, based on Drug-Like Soft rule.

### **fChar**

Formal charge. Optimal:-4 ~4, based on Drug-Like Soft rule

### **nRig**

Number of rigid bonds. Number of non-flexible bonds, in opposite to rotatable bonds. Optimal:0~30, based on Drug-Like Soft rule.

### **Flexibility**

Flexibility = nRot / nRig

### **Stereo Centers**

Number of stereocenters. Optimal:  $\leq 2$ , based on Lead-Like Soft rule.

### **TPSA**

Topological polar surface area. Sum of tabulated surface contributions of polar fragments. Optimal:0~140, based on *Veber* rule.

### **logS**

- The logarithm of aqueous solubility value. The first step in the drug absorption process is the disintegration of the tablet or capsule, followed by the dissolution of the active drug. Low solubility is detrimental to good and complete oral absorption, and early measurement of this property is of great importance in drug discovery.
- Results interpretation: The predicted solubility of a compound is given as the logarithm of the molar concentration (log mol/L). Compounds in the range from -4 to 0.5 log mol/L will be considered proper.

### **logP**

- The logarithm of the n-octanol/water distribution coefficient. log P possess a leading position with considerable impact on both membrane permeability and hydrophobic binding to macromolecules, including the target receptor as well as other proteins like plasma proteins, transporters, or metabolizing enzymes.

- Results interpretation: The predicted logP of a compound is given as the logarithm of the molar concentration (log mol/L). Compounds in the range from 0 to 3 log mol/L will be considered proper.

#### **logD7.4**

- The logarithm of the n-octanol/water distribution coefficients at pH=7.4. To exert a therapeutic effect, one drug must enter the blood circulation and then reach the site of action. Thus, an eligible drug usually needs to keep a balance between lipophilicity and hydrophilicity to dissolve in the body fluid and penetrate the biomembrane effectively. Therefore, it is important to estimate the n-octanol/water distribution coefficients at physiological pH (logD7.4) values for candidate compounds in the early stage of drug discovery.
- Results interpretation: The predicted logD7.4 of a compound is given as the logarithm of the molar concentration (log mol/L). Compounds in the range from 1 to 3 log mol/L will be considered proper.

## **2. Medicinal Chemistry**

#### **QED [1]**

- A measure of drug-likeness based on the concept of desirability. QED is calculated by integrating the outputs of the desirability functions based on eight drug-likeness related properties, including MW, log P, N<sub>HBA</sub>, N<sub>HBD</sub>, PSA, N<sub>rotb</sub>, the number of aromatic rings (N<sub>Ar</sub>), and the number of alerts for undesirable functional groups. Here, average descriptor weights were used in the calculation of QED. The QED score is calculated by taking the geometric mean of the individual desirability functions, given by  $QED = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln d_i\right)$ , where  $d_i$  indicates the  $d_{th}$  desirability function and  $n = 8$  is the number of drug-likeness related properties.
- Results interpretation: The mean QED is 0.67 for the attractive compounds, 0.49 for the unattractive compounds and 0.34 for the unattractive compounds considered too complex.
- Empirical decision: > 0.67: excellent (green); ≤ 0.67: poor (red)

#### **SAScore [2]**

- Synthetic accessibility score is designed to estimate ease of synthesis of drug-like molecules, based on a combination of fragment contributions and a complexity penalty. The score is between 1 (easy to make) and 10 (very difficult to make). The synthetic accessibility score (SAScore) is calculated as a combination of two components:  $SAScore = fragmentScore - complexityPenalty$ .
- Results interpretation: high SAScore: ≥ 6, difficult to synthesize; low SAScore: < 6, easy to synthesize
- Empirical decision: ≤ 6: excellent (green); > 6: poor (red)

### **Fsp<sup>3</sup>** [3]

- Fsp<sup>3</sup>, the number of sp<sup>3</sup> hybridized carbons/total carbon count, is used to determine the carbon saturation of molecules and characterize the complexity of the spatial structure of molecules. It has been demonstrated that the increased saturation measured by Fsp<sup>3</sup> and the number of chiral centers in the molecule increase the clinical success rate, which might be related to the increased solubility, or the fact that the enhanced 3D features allow small molecules to occupy more target space.
- Results interpretation: Fsp<sup>3</sup> ≥ 0.42 is considered a suitable value.
- Empirical decision: ≥ 0.42: excellent (green); < 0.42: poor (red)

### **MCE-18** [4]

- MCE-18 stands for medicinal chemistry evolution in 2018, and this measure can effectively score molecules by novelty in terms of their cumulative sp<sup>3</sup> complexity. It can effectively score structures by their novelty and current lead potential in contrast to simple and in many cases false positive sp<sup>3</sup> index, and given by the following equation:  $MCE18 = \left( AR + NAR + CHIRAL + SPIRO + \frac{sp^3 + Cyc - Acyc}{1 + sp^3} \right) \times Q^1$ , where AR is the presence of an aromatic or heteroaromatic ring (0 or 1), NAR is the presence of an aliphatic or a heteroaliphatic ring (0 or 1), CHIRAL is the presence of a chiral center (0 or 1), SPIRO is the presence of a spiro point (0 or 1), sp<sup>3</sup> is the portion of sp<sup>3</sup>-hybridized carbon atoms (from 0 to 1), Cyc is the portion of cyclic carbons that are sp<sup>3</sup> hybridized (from 0 to 1), Acyc is a portion of acyclic carbon atoms that are sp<sup>3</sup> hybridized (from 0 to 1), and Q1 is the normalized quadratic index.
- Results interpretation: < 45: uninteresting, trivial, old scaffolds, low degree of 3D complexity and novelty; 45~63: sufficient novelty, basically follow the trends of currently observed in medicinal chemistry; 63~78: high structural similarity to the compounds disclosed in patent records; >78: need to be inspected visually to assess their target profile and drug-likeness.
- Empirical decision: ≥ 45: excellent (green); < 45: poor (red)

### **NPscore** [5]

- The Natural Product-likeness score is a useful measure which can help to guide the design of new molecules toward interesting regions of chemical space which have been identified as “bioactive regions” by natural evolution. The calculation consists of molecule fragmentation, table lookup, and summation of fragment contributions.
- Results interpretation: The calculated score is typically in the range from -5 to 5. The higher the score is, the higher the probability is that the molecule is a NP.

### **Lipinski Rule** [6]

- Content: MW ≤ 500; logP ≤ 5; Hacc ≤ 10; Hdon ≤ 5
- Results interpretation: If two properties are out of range, a poor absorption or permeability is possible, one is acceptable.

- Empirical decision: < 2 violations: excellent (green); ≥2 violations: poor (red)

#### **Pfizer Rule** <sup>[7]</sup>

- Content: logP > 3; TPSA < 75
- Results interpretation: Compounds with a high log P (>3) and low TPSA (<75) are likely to be toxic.
- Empirical decision: two conditions satisfied: poor (red); otherwise: excellent (green)

#### **GSK Rule** <sup>[8]</sup>

- Content: MW ≤ 400; logP ≤ 4
- Results interpretation: Compounds satisfying the GSK rule may have a more favorable ADMET profile.
- Empirical decision: 0 violations: excellent (green); otherwise: poor (red)

#### **Golden Triangle** <sup>[9]</sup>

- Content: 200 ≤ MW ≤ 500; -2 ≤ logD ≤ 5
- Results interpretation: Compounds satisfying the GoldenTriangle rule may have a more favourable ADMET profile.
- Empirical decision: 0 violations: excellent (green); otherwise: poor (red)

#### **PAINS** <sup>[10]</sup>

- Pan Assay Interference Compounds (PAINS) is one of the most famous frequent hitters filters, which comprises 480 substructures derived from the analysis of FHs determined by six target-based HTS assay. By application of these filters, it is easier to screen false positive hits and to flag suspicious compounds in screening databases. One of the most authoritative medicine magazines *Journal of Medicinal Chemistry* even requires authors to provide the screening results with the PAINS alerts of active compounds when submitting manuscripts.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

#### **ALARM NMR Rule** <sup>[11]</sup>

- Thiol reactive compounds. There are 75 substructures in this endpoint.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

#### **BMS Rule** <sup>[12]</sup>

- Undesirable, reactive compounds. There are 176 substructures in this endpoint.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

### **Chelator Rule<sup>[13]</sup>**

- Chelating compounds. There are 55 substructures in this endpoint.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

## **3. Absorption**

### **Caco-2 Permeability**

- Before an oral drug reaches the systemic circulation, it must pass through intestinal cell membranes via passive diffusion, carrier-mediated uptake or active transport processes. The human colon adenocarcinoma cell lines (Caco-2), as an alternative approach for the human intestinal epithelium, has been commonly used to estimate in vivo drug permeability due to their morphological and functional similarities. Thus, Caco-2 cell permeability has also been an important index for an eligible candidate drug compound.
- Results interpretation: The predicted Caco-2 permeability of a given compound is given as the log cm/s. A compound is considered to have a proper Caco-2 permeability if it has predicted value  $> -5.15 \log \text{ cm/s}$ .
- Empirical decision:  $> -5.15$ : excellent (green); otherwise: poor (red)

### **MDCK Permeability**

- Madin–Darby Canine Kidney cells (MDCK) have been developed as an in vitro model for permeability screening. Its apparent permeability coefficient,  $P_{app}$ , is widely considered to be the in vitro gold standard for assessing the uptake efficiency of chemicals into the body.  $P_{app}$  values of MDCK cell lines are also used to estimate the effect of the blood-brain barrier (BBB).
- Results interpretation: The unit of predicted MDCK permeability is cm/s. A compound is considered to have a high passive MDCK permeability for a  $P_{app} > 20 \times 10^{-6} \text{ cm/s}$ , medium permeability for  $2-20 \times 10^{-6} \text{ cm/s}$ , low permeability for  $< 2 \times 10^{-6} \text{ cm/s}$ .
- Empirical decision:  $> 2 \times 10^{-6} \text{ cm/s}$ : excellent (green), otherwise: poor (red)

### **Pgp-inhibitor**

- The inhibitor of P-glycoprotein. The P-glycoprotein, also known as MDR1 or 2 ABCB1, is a membrane protein member of the ATP-binding cassette (ABC) transporters superfamily. It is probably the most promiscuous efflux transporter, since it recognizes a number of structurally different and apparently unrelated xenobiotics; notably, many of them are also CYP3A4 substrates.
- Results interpretation: Category 0: Non-inhibitor; Category 1: Inhibitor. The output value is the probability of being Pgp-inhibitor, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### ***Pgp-substrate***

- As described in the Pgp-inhibitor section, modulation of P-glycoprotein mediated transport has significant pharmacokinetic implications for Pgp substrates, which may either be exploited for specific therapeutic advantages or result in contraindications.
- Results interpretation: Category 0: Non-substrate; Category 1: substrate. The output value is the probability of being Pgp-substrate, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### ***HIA***

- Human intestinal absorption. As described above, the human intestinal absorption of an oral drug is the essential prerequisite for its apparent efficacy. What's more, the close relationship between oral bioavailability and intestinal absorption has also been proven and HIA can be seen an alternative indicator for oral bioavailability to some extent.
- Result interpretation: A molecule with an absorbance of less than 30% is considered to be poorly absorbed. Accordingly, molecules with a HIA >30% were classified as HIA- (Category 0), while molecules with a HIA < 30% were classified as HIA+(Category 1). The output value is the probability of being HIA+, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### ***F<sub>20%</sub>***

- The human oral bioavailability 20%. For any drug administrated by the oral route, oral bioavailability is undoubtedly one of the most important pharmacokinetic parameters because it is the indicator of the efficiency of the drug delivery to the systemic circulation.
- Result interpretation: Molecules with a bioavailability  $\geq 20\%$  were classified as F<sub>20%-</sub> (Category 0), while molecules with a bioavailability < 20% were classified as F<sub>20%+</sub> (Category 1). The output value is the probability of being F<sub>20%+</sub>, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### ***F<sub>30%</sub>***

- The human oral bioavailability 30%. For any drug administrated by the oral route, oral bioavailability is undoubtedly one of the most important pharmacokinetic parameters because it is the indicator of the efficiency of the drug delivery to the systemic circulation.
- Result interpretation: Molecules with a bioavailability  $\geq 30\%$  were classified as F<sub>30%-</sub> (Category 0), while molecules with a bioavailability < 30% were classified as F<sub>30%+</sub> (Category 1). The output value is the probability of being F<sub>30%+</sub>, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

## 4. Distribution

### PPB

- Plasma protein binding. One of the major mechanisms of drug uptake and distribution is through PPB, thus the binding of a drug to proteins in plasma has a strong influence on its pharmacodynamic behavior. PPB can directly influence the oral bioavailability because the free concentration of the drug is at stake when a drug binds to serum proteins in this process.
- Result interpretation: A compound is considered to have a proper PPB if it has predicted value  $< 90\%$ , and drugs with high protein-bound may have a low therapeutic index.
- Empirical decision:  $\leq 90\%$ : excellent (green); otherwise: poor (red).

### VD

- Volume Distribution. The VD is a theoretical concept that connects the administered dose with the actual initial concentration present in the circulation and it is an important parameter to describe the in vivo distribution for drugs. In practical, we can speculate the distribution characters for an unknown compound according to its VD value, such as its condition binding to plasma protein, its distribution amount in body fluid and its uptake amount in tissues.
- Result interpretation: The unit of predicted VD is L/kg. A compound is considered to have a proper VD if it has predicted VD in the range of 0.04-20L/kg.
- Empirical decision: 0.04-20: excellent (green); otherwise: poor (red)

### BBB Penetration

- Drugs that act in the CNS need to cross the blood–brain barrier (BBB) to reach their molecular target. By contrast, for drugs with a peripheral target, little or no BBB penetration might be required in order to avoid CNS side effects.
- Result interpretation: The unit of BBB penetration is cm/s. Molecules with  $\log_{BB} > -1$  were classified as BBB+ (Category 1), while molecules with  $\log_{BB} \leq -1$  were classified as BBB- (Category 0). The output value is the probability of being BBB+, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### Fu

- The fraction unbound in plasms. Most drugs in plasma will exist in equilibrium between either an unbound state or bound to serum proteins. Efficacy of a given drug may be affect by the degree to which it binds proteins within blood, as the more that is bound the less efficiently it can traverse cellular membranes or diffuse.
- Result interpretation:  $>20\%$ : High Fu; 5-20%: medium Fu;  $<5\%$  low Fu.
- Empirical decision:  $\geq 5\%$ : excellent (green);  $< 5\%$ : poor (red).



## 5. Metabolism

### **CYP 1A2 / 2C19 / 2C9 / 2D6 / 3A4 inhibitor**

### **CYP 1A2 / 2C19 / 2C9 / 2D6 / 3A4 substrate**

- Based on the chemical nature of biotransformation, the process of drug metabolism reactions can be divided into two broad categories: phase I (oxidative reactions) and phase II (conjugative reactions). The human cytochrome P450 family (phase I enzymes) contains 57 isozymes and these isozymes metabolize approximately two-thirds of known drugs in human with 80% of this attribute to five isozymes—1A2, 3A4, 2C9, 2C19 and 2D6. Most of these CYPs responsible for phase I reactions are concentrated in the liver.
- Result interpretation: Category 0: Non-substrate / Non-inhibitor; Category 1: substrate / inhibitor. The output value is the probability of being substrate / inhibitor, within the range of 0 to 1.

## 6. Excretion

### **CL**

- The clearance of a drug. Clearance is an important pharmacokinetic parameter that defines, together with the volume of distribution, the half-life, and thus the frequency of dosing of a drug.
- Result interpretation: The unit of predicted CL penetration is ml/min/kg. >15 ml/min/kg: high clearance; 5-15 ml/min/kg: moderate clearance; <5 ml/min/kg: low clearance.
- Empirical decision:  $\geq 5$ : excellent (green); < 5: poor (red).

### **$T_{1/2}$**

- The half-life of a drug is a hybrid concept that involves clearance and volume of distribution, and it is arguably more appropriate to have reliable estimates of these two properties instead.
- Result interpretation: Molecules with  $T_{1/2} > 3$  were classified as  $T_{1/2}^-$  (Category 0), while molecules with  $T_{1/2} \leq 3$  were classified as  $T_{1/2}^+$  (Category 1). The output value is the probability of being  $T_{1/2}^+$ , within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

## **7. Toxicology**

### ***hERG Blockers***

- The human ether-a-go-go related gene. The During cardiac depolarization and repolarization, a voltage-gated potassium channel encoded by hERG plays a major role in the regulation of the exchange of cardiac action potential and resting potential. The hERG blockade may cause long QT syndrome (LQTS), arrhythmia, and Torsade de Pointes (TdP), which lead to palpitations, fainting, or even sudden death.
- Result interpretation: Molecules with  $IC_{50}$  more than 10  $\mu$ M or less than 50% inhibition at 10  $\mu$ M were classified as hERG - (Category 0), while molecules with  $IC_{50}$  less than 10  $\mu$ M or more than 50% inhibition at 10  $\mu$ M were classified as hERG+ (Category 1). The output value is the probability of being hERG+, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### ***H-HT***

- The human hepatotoxicity. Drug induced liver injury is of great concern for patient safety and a major cause for drug withdrawal from the market. Adverse hepatic effects in clinical trials often lead to a late and costly termination of drug development programs.
- Result interpretation: Category 0: H-HT negative(-); Category 1: H-HT positive(+). The output value is the probability of being toxic, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### ***DILI***

- Drug-induced liver injury (DILI) has become the most common safety problem of drug withdrawal from the market over the past 50 years.
- Result interpretation: Category 0: DILI negative(-); Category 1: DILI positive(+). The output value is the probability of being toxic, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### ***AMES Toxicity***

- The Ames test for mutagenicity. The mutagenic effect has a close relationship with the carcinogenicity, and it is the most widely used assay for testing the mutagenicity of compounds.
- Result interpretation: Category 0: AMES negative(-); Category 1: AMES positive(+). The output value is the probability of being toxic, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### ***Rat Oral Acute Toxicity***

- Determination of acute toxicity in mammals (e.g. rats or mice) is one of the most important tasks for the safety evaluation of drug candidates.

- Result interpretation: Category 0: low-toxicity, > 500 mg/kg; Category 1: high-toxicity; < 500 mg/kg. The output value is the probability of being toxic, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **FDAMDD**

- The maximum recommended daily dose provides an estimate of the toxic dose threshold of chemicals in humans.
- Result interpretation: Category 1: FDAMDD positive(+),  $\leq 0.011$  mmol/kg -bw/day; Category 0: FDAMDD negative(-), > 0.011 mmol/kg-bw/day. The output value is the probability of being toxic, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **Skin Sensitization**

- Skin sensitization is a potential adverse effect for dermally applied products. The evaluation of whether a compound, that may encounter the skin, can induce allergic contact dermatitis is an important safety concern.
- Result interpretation: Category 1: Sensitizer; Category 0: Non-sensitizer. The output value is the probability of being toxic, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **Carcinogenicity**

- Among various toxicological endpoints of chemical substances, carcinogenicity is of great concern because of its serious effects on human health. The carcinogenic mechanism of chemicals may be due to their ability to damage the genome or disrupt cellular metabolic processes. Many approved drugs have been identified as carcinogens in humans or animals and have been withdrawn from the market.
- Result interpretation: Category 1: carcinogens; Category 0: non-carcinogens. Chemicals are labelled as active (carcinogens) or inactive (non-carcinogens) according to their TD<sub>50</sub> values. The output value is the probability of being toxic, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **Eye Corrosion / Irritation**

- Assessing the eye irritation/corrosion (EI/EC) potential of a chemical is a necessary component of risk assessment. Cornea and conjunctiva tissues comprise the anterior surface of the eye, and hence cornea and conjunctiva tissues are directly exposed to the air and easily suffer injury by chemicals. There are several substances, such as chemicals used in manufacturing, agriculture and warfare, ocular pharmaceuticals, cosmetic products, and household products, that can cause EI or EC.
- Result interpretation: Category 1: corrosives / irritants chemicals; Category 0: non-corrosives / non-irritants chemicals. The output value is the probability of being toxic, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **Respiratory Toxicity**

- Among these safety issues, respiratory toxicity has become the main cause of drug withdrawal. Drug-induced respiratory toxicity is usually underdiagnosed because it may not have distinct early signs or symptoms in common medications and can occur with significant morbidity and mortality. Therefore, careful surveillance and treatment of respiratory toxicity is of great importance.
- Result interpretation: Category 1: respiratory toxicants; Category 0: non-respiratory toxicants. The output value is the probability of being toxic, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **Bioconcentration Factor**

- The bioconcentration factor BCF is defined as the ratio of the chemical concentration in biota as a result of absorption via the respiratory surface to that in water at steady state. It is used for considering secondary poisoning potential and assessing risks to human health via the food chain. The unit of BCF is log<sub>10</sub>(L/kg).

### **IGC<sub>50</sub>**

- 48 hour Tetrahymena pyriformis IGC<sub>50</sub> (concentration of the test chemical in water in mg/L that causes 50% growth inhibition to Tetrahymena pyriformis after 48 hours). The unit of IGC<sub>50</sub> is  $-\log_{10}[(\text{mg/L})/(1000 \cdot \text{MW})]$ .

### **LC<sub>50</sub>FM**

- 96 hour fathead minnow LC<sub>50</sub> (concentration of the test chemical in water in mg/L that causes 50% of fathead minnow to die after 96 hours). The unit of LC<sub>50</sub>FM is  $-\log_{10}[(\text{mg/L})/(1000 \cdot \text{MW})]$ .

### **LC<sub>50</sub>DM**

- 48 hour Daphnia magna LC<sub>50</sub> (concentration of the test chemical in water in mg/L that causes 50% of Daphnia magna to die after 48 hours). The unit of LC<sub>50</sub>DM is  $-\log_{10}[(\text{mg/L})/(1000 \cdot \text{MW})]$ .

### **NR-AR**

- Androgen receptor (AR), a nuclear hormone receptor, plays a critical role in AR-dependent prostate cancer and other androgen related diseases. Endocrine disrupting chemicals (EDCs) and their interactions with steroid hormone receptors like AR may cause disruption of normal endocrine function as well as interfere with metabolic homeostasis, reproduction, developmental and behavioral functions.
- Result interpretation: Category 1: actives ; Category 0: inactives. The output value is the probability of being AR agonists, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **NR-AR-LBD**

- Androgen receptor (AR), a nuclear hormone receptor, plays a critical role in AR-dependent prostate cancer and other androgen related diseases. Endocrine disrupting chemicals (EDCs) and their interactions with

steroid hormone receptors like AR may cause disruption of normal endocrine function as well as interfere with metabolic homeostasis, reproduction, developmental and behavioral functions.

- Result interpretation: Category 1: actives ; Category 0: inactives. Molecules that labeled 1 in this bioassay may bind to the LBD of androgen receptor. The output value is the probability of being actives, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

#### **NR-AhR**

- The Aryl hydrocarbon Receptor (AhR), a member of the family of basic helix-loop-helix transcription factors, is crucial to adaptive responses to environmental changes. AhR mediates cellular responses to environmental pollutants such as aromatic hydrocarbons through induction of phase I and II enzymes but also interacts with other nuclear receptor signaling pathways.
- Result interpretation: Category 1: actives ; Category 0: inactives. Molecules that labeled 1 may activate the aryl hydrocarbon receptor signaling pathway. The output value is the probability of being actives, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

#### **NR-Aromatase**

- Endocrine disrupting chemicals (EDCs) interfere with the biosynthesis and normal functions of steroid hormones including estrogen and androgen in the body. Aromatase catalyzes the conversion of androgen to estrogen and plays a key role in maintaining the androgen and estrogen balance in many of the EDC-sensitive organs.
- Result interpretation: Category 1: actives ; Category 0: inactives. Molecules that labeled 1 are regarded as aromatase inhibitors that could affect the balance between androgen and estrogen. The output value is the probability of being actives, within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

#### **NR-ER**

- Estrogen receptor (ER), a nuclear hormone receptor, plays an important role in development, metabolic homeostasis and reproduction. Endocrine disrupting chemicals (EDCs) and their interactions with steroid hormone receptors like ER causes disruption of normal endocrine function. Therefore, it is important to understand the effect of environmental chemicals on the ER signaling pathway.
- Result interpretation: Category 1: actives ; Category 0: inactives. The output value is the probability of being actives within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

#### **NR-ER-LBD**

- Estrogen receptor (ER), a nuclear hormone receptor, plays an important role in development, metabolic homeostasis and reproduction. Two subtypes of ER, ER-alpha and ER-beta have similar expression

patterns with some uniqueness in both types. Endocrine disrupting chemicals (EDCs) and their interactions with steroid hormone receptors like ER causes disruption of normal endocrine function.

- Result interpretation: Category 1: actives ; Category 0: inactives. The output value is the probability of being actives within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

#### **NR-PPAR-gamma**

- The peroxisome proliferator-activated receptors (PPARs) are lipid-activated transcription factors of the nuclear receptor superfamily with three distinct subtypes namely PPAR alpha, PPAR delta (also called PPAR beta) and PPAR gamma (PPAR $\gamma$ ). All these subtypes heterodimerize with Retinoid X receptor (RXR) and these heterodimers regulate transcription of various genes. PPAR-gamma receptor (glitazone receptor) is involved in the regulation of glucose and lipid metabolism.
- Result interpretation: Category 1: actives ; Category 0: inactives. The output value is the probability of being actives within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

#### **SR-ARE**

- Oxidative stress has been implicated in the pathogenesis of a variety of diseases ranging from cancer to neurodegeneration. The antioxidant response element (ARE) signaling pathway plays an important role in the amelioration of oxidative stress. The CellSensor ARE-bla HepG2 cell line (Invitrogen) can be used for analyzing the Nrf2/antioxidant response signaling pathway. Nrf2 (NF-E2-related factor 2) and Nrf1 are transcription factors that bind to AREs and activate these genes.
- Result interpretation: Category 1: actives ; Category 0: inactives. The output value is the probability of being actives within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

#### **SR-ATAD5**

- ATPase family AAA domain-containing protein 5. As cancer cells divide rapidly and during every cell division they need to duplicate their genome by DNA replication. The failure to do so results in the cancer cell death. Based on this concept, many chemotherapeutic agents were developed but have limitations such as low efficacy and severe side effects etc. Enhanced Level of Genome Instability Gene 1 (ELG1; human ATAD5) protein levels increase in response to various types of DNA damage.
- Result interpretation: Category 1: actives ; Category 0: inactives. The output value is the probability of being actives within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **SR-HSE**

- Heat shock factor response element. Various chemicals, environmental and physiological stress conditions may lead to the activation of heat shock response/ unfolded protein response (HSR/UPR). There are three heat shock transcription factors (HSFs) (HSF-1, -2, and -4) mediating transcriptional regulation of the human HSR.
- Result interpretation: Category 1: actives ; Category 0: inactives. The output value is the probability of being actives within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **SR-MMP**

- Mitochondrial membrane potential (MMP), one of the parameters for mitochondrial function, is generated by mitochondrial electron transport chain that creates an electrochemical gradient by a series of redox reactions. This gradient drives the synthesis of ATP, a crucial molecule for various cellular processes. Measuring MMP in living cells is commonly used to assess the effect of chemicals on mitochondrial function; decreases in MMP can be detected using lipophilic cationic fluorescent dyes.
- Result interpretation: Category 1: actives ; Category 0: inactives. The output value is the probability of being actives within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **SR-p53**

- p53, a tumor suppressor protein, is activated following cellular insult, including DNA damage and other cellular stresses. The activation of p53 regulates cell fate by inducing DNA repair, cell cycle arrest, apoptosis, or cellular senescence. The activation of p53, therefore, is a good indicator of DNA damage and other cellular stresses.
- Result interpretation: Category 1: actives ; Category 0: inactives. The output value is the probability of being actives within the range of 0 to 1.
- Empirical decision: 0-0.3: excellent (green); 0.3-0.7: medium (yellow); 0.7-1.0(++): poor (red)

### **Acute Toxicity Rule**

- Molecules containing these substructures may cause acute toxicity during oral administration. There are 20 substructures in this endpoint.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

### **Genotoxic Carcinogenicity Rule**

- Molecules containing these substructures may cause carcinogenicity or mutagenicity through genotoxic mechanisms. There are 117 substructures in this endpoint.

- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

#### ***NonGenotoxic Carcinogenicity Rule***

- Molecules containing these substructures may cause carcinogenicity through nongenotoxic mechanisms. There are 23 substructures in this endpoint.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

#### ***Skin Sensitization Rule***

- Molecules containing these substructures may cause skin irritation. There are 155 substructures in this endpoint. Molecules containing these substructures may cause skin irritation.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

#### ***Aquatic Toxicity Rule***

- Molecules containing these substructures may cause toxicity to liquid(water). There are 99 substructures in this endpoint.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

#### ***NonBiodegradable Rule***

- Molecules containing these substructures may be non-biodegradable. There are 19 substructures in this endpoint.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

#### ***SureChEMBL Rule***

- Molecules matching one or more structural alerts are considered to have MedChem unfriendly status. There are 164 substructures in this endpoint.
- Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.

#### ***FAF-Drugs4 Rule***

- Molecules containing these substructures may be toxic. There are 154 substructures collected form FAF-Drugs4 webserver in this endpoint.  
Results interpretation: If the number of alerts is not zero, the users could check the substructures by the DETIAL button.



- [1] Bickerton G R, Paolini G V, Besnard J, et al. Quantifying the chemical beauty of drugs[J]. *Nat Chem*, 2012, 4(2): 90-8.
- [2] Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions[J]. *J Cheminform*, 2009, 1(1): 8.
- [3] Lovering F, Bikker J, Humblet C. Escape from flatland: increasing saturation as an approach to improving clinical success[J]. *J Med Chem*, 2009, 52(21): 6752-6.
- [4] Ivanenkov Y A, Zagribelnyy B A, Aladinskiy V A. Are We Opening the Door to a New Era of Medicinal Chemistry or Being Collapsed to a Chemical Singularity?[J]. *J Med Chem*, 2019, 62(22): 10026-10043.
- [5] Ertl P, Roggo S, Schuffenhauer A. Natural product-likeness score and its application for prioritization of compound libraries[J]. *J Chem Inf Model*, 2008, 48(1): 68-74.
- [6] Lipinski C A, Lombardo F, Dominy B W, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings[J]. *Adv Drug Deliv Rev*, 2001, 46(1-3): 3-26.
- [7] Hughes J D, Blagg J, Price D A, et al. Physicochemical drug properties associated with in vivo toxicological outcomes[J]. *Bioorg Med Chem Lett*, 2008, 18(17): 4872-5.
- [8] Gleeson M P. Generation of a set of simple, interpretable ADMET rules of thumb[J]. *J Med Chem*, 2008, 51(4): 817-34.
- [9] Johnson T W, Dress K R, Edwards M. Using the Golden Triangle to optimize clearance and oral absorption[J]. *Bioorg Med Chem Lett*, 2009, 19(19): 5560-4.
- [10] 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]. *J Med Chem*, 2010, 53(7): 2719-40.
- [11] Huth J R, Mendoza R, Olejniczak E T, et al. ALARM NMR: a rapid and robust experimental method to detect reactive false positives in biochemical screens[J]. *J Am Chem Soc*, 2005, 127(1): 217-24.
- [12] Pearce B C, Sofia M J, Good A C, et al. An empirical process for the design of high-throughput screening deck filters[J]. *J Chem Inf Model*, 2006, 46(3): 1060-8.
- [13] Agrawal A, Johnson S L, Jacobsen J A, et al. Chelator fragment libraries for targeting metalloproteinases[J]. *ChemMedChem*, 2010, 5(2): 195-9