Deep-PK: Theory

Absorption

1. Human colorectal adenocarcinoma cells (Caco-2)
    a. **Description**: The Caco-2 cell line is composed of human epithelial colorectal adenocarcinoma cells. The Caco-2 monolayer of cells is widely used as an *in vitro* model of the human intestinal mucosa to predict the absorption of orally administered drugs.
    b. **Interpretation**: The unit of Caco-2 permeability is the logarithm of the apparent permeability coefficient (log P_{app}; log cm/s). A compound is considered to have high Caco-2 permeability if its Papp is> 8 x 10^{-6} cm/s. For the predictive model, high Caco-2 permeability would translate into predicted values > 0.90.
    c. **Assay Type**: *In vitro.*
2. Human oral bioavailability (20 and 50%F)
    a. **Description**: Human oral bioavailability determines the proportion (%) of an orally administered drug that enters the systemic circulation. It can also be reported as a fraction (Fabs) of the delivered dose that is absorbed intact into the systemic circulation after bypassing first-pass metabolism in the liver.
    b. **Interpretation**: The probability of 20%F, for example, indicates that the chance of a drug bioavailability can be greater than 20% within the range of Class 0 and Class 1 for low and high, respectively.
    c. **Assay Type**: *In vitro.*
3. Human intestinal absorption (HIA)
    a. **Description**: The Intestine is normally the primary site for absorption of a drug from an orally administered solution. This method is built to predict the proportion of compounds that were absorbed (> 30%) through the human small intestine. This does not always reflect oral bioavailability which will be affected by first-pass metabolism.
    b. **Interpretation**: For a given compound it predicts the percentage of a delivered dose that will be absorbed into portal blood through the human intestine. A molecule with an absorbance of less than 30% is considered to be poorly absorbed.
    c. **Assay Type**: *In vitro.*
4. Madin-Darby Canine Kidney cells (MDCK)
    a. **Description**: MDCK cell based model was developed to measure the human intestine membrane permeability of chemicals.
    b. **Interpretation**: The unit of observed and predicted MDCK permeability coefficient (P_{app}) is nm/s. The permeability is classified into three classes: Low permeability < 4 nm/s, medium 4-70 nm/s, high > 70 nm/s. (adapted from Yazadanian et al (1998))
    c. Assay Type: *In vitro.*
5. P-glycoprotein inhibitor I

a. **Description**: 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.
b. **Interpretation**: The predictor will determine if a given compound is likely to be a P-glycoprotein I inhibitor.
c. **Assay Type:** *In vitro.*

6. P-glycoprotein substrate
    a. **Description**: The P-glycoprotein is an ATP-binding cassette (ABC) transporter. It functions as a biological barrier by extruding toxins and xenobiotics out of cells. P-glycoprotein transport screening is performed using transgenic MDR knockout mice and *in vitro* cell systems.
    b. **Interpretation**: The model predicts whether a given compound is likely to be a substrate of Pgp or not.
    c. **Assay Type**: *In vitro.*
7. Human skin permeability
    a. **Description**: Skin permeability is a significant consideration for the efficacy and systemic (blood) exposure of many consumer products applied topically to the skin, and is of interest for the development of transdermal drug delivery.
    b. **Interpretation**: It predicts if a given compound is likely to be skin permeable, expressed as the skin permeability constant logKp (cm/h). A compound is considered to have a relatively low skin permeability if it has a logKp > -2.5.
    c. **Assay Type**: *In vitro.*

Distribution

8. The blood-brain barrier (BBB)
    a. **Description**: The brain is protected from exogenous compounds by the blood-brain barrier (BBB). The ability of a drug to cross into the brain is an important parameter to consider to help reduce side effects and toxicities, or to improve the efficacy of drugs whose pharmacological activity is within the brain. Blood-brain permeability is measured in vivo in animal models using the permeability-surface area (PS) in its logarithmic ratio (logPS). Brain permeability, however, does not alone dictate efficacy, since many compounds can also be substrates for PgP mediated efflux from the brain.
    b. **Interpretation**: The distinction of positively (CNSp+) and negatively (CNSp−) classified molecules refers to compounds with logPS values ≥ −2 and ≤−3, respectively.
    c. Assay type: *In vivo.*
9. Breast cancer resistance protein inhibitor (BCRP)
    a. **Description**: The breast cancer resistance protein (BCRP) is a member of the ATP-binding cassette (ABC) transporter superfamily, which plays an important role in exporting xenobiotics out of cells.
    b. **Interpretation**: The predictor will determine if a given compound is likely to be a BCRP inhibitor or not.
    c. Assay type: *Unknown.*

10. Plasma protein binding (PPB)

a. **Description**: The plasma protein binding (PPB) measures the non-specific binding of a drug to plasma proteins, which can affect the amount of free drug in the body.
b. **Interpretation**: The predicted value in percentage is the fraction of compounds bound to proteins (Fu).
c. Assay type: *In vivo.*

11. Fraction Unbound (Human)

a. Description: Fraction unbound (FU) refers to the proportion of a small molecule drug that is not bound to proteins in the bloodstream of humans. FU is an important pharmacokinetic property because only the unbound fraction of a drug is typically available to exert pharmacological effects or be metabolized and eliminated from the body. Therefore, it directly influences the drug's potency, efficacy, and potential for adverse effects.
b. Assay type: *In vivo*

12. Steady State Volume of Distribution (SSVD)

a. Description: SSVD determines the volume of distribution during steady state conditions, i.e. when there is a stable drug concentration.
b. Assay Type: *In vivo.*

<u>Metabolism</u>

13. Cytochrome P450 inhibitors

a. **Description**: Cytochrome P450s are an important class of detoxification enzymes in the body, mainly found in the liver. They are responsible for metabolising xenobiotics to facilitate their excretion via the urine. While most drugs are inactivated by the cytochrome P450's, in some cases, P450's are employed to activate 'prodrugs'. Inhibitors of this enzyme, such as grapefruit juice, can affect drug metabolism and are therefore contraindicated for some drugs who's pharmacokinetics are critically driven by P450 activity. It is therefore important to assess a compound's ability to inhibit cytochrome P450's. Models for major P450 isoforms (CYP1A2/CYP2C19/CYP2C9/CYP2D6/CYP3A4) were built based on compounds whose ability to inhibit these cytochrome P450 were known.
b. **Interpretation**: A compound is considered to be a cytochrome P450 inhibitor (Class 1) if the concentration required to inhibit P450 activity by 50% is less than 10 uM.
c. Assay Type: *In vitro.*

14. Cytochrome P450 substrates

a. **Description**: Cytochrome P450 substrates represent drugs who's pharmacokinetics are largely driven by their metabolism by P450's. The main isoforms responsible for drug metabolism are CYP1A2, CYP2C19, CYP2C9, CYP2D6 and CYP3A4.

b. **Interpretation**: The predictor will assess whether a given molecule is likely to be metabolised by one of these P450's (Class1) or not (Class 0).
c. **Assay Type:** *In vitro.*

15. Organic anion transporting polypeptides (OATP) inhibitors
    a. **Description**: Both OATP1B1 and OATP1B3 are transmembrane transporters on the basolateral membrane of the liver, which play important roles in exporting xenobiotics into hepatocytes. For some drugs, this can be a key step to bringing those drugs into contact with liver enzymes such as P450's.
    b. **Interpretation:** The predictor will determine if a given compound is likely to be an OATP inhibitor or non-inhibitor.
    c. **Assay type:** *In vitro.*

Excretion

16. Clearance
    a. **Description**: Drug clearance (from plasma/serum) is measured by the proportionality constant (CLtot), and occurs primarily as a combination of hepatic clearance (metabolism in the liver and biliary clearance) and renal clearance (excretion via the kidneys). It is the major pharmacokinetic parameter that describes how long a drug will remain in the blood. It is related to the systemically available dose, and is important for determining dosing rates to achieve steady-state concentrations.
    b. **Interpretation**: The predicted total clearance (CLtot) of a given compound is given in ml/min/kg body weight.
    c. **Assay Type:** *In vivo.*
17. Half-life
    a. **Description**: Half-life (t1/2) is the time required to decrease a given concentration of a drug by half in plasma/serum. Reported half-life represents the half life calculated during the final 'elimination' phase of the drugs plasma concentration vs time profile, after absorption and tissue distribution have been completed.
    b. **Interpretation**: The predictor will determine if a given compound is likely to have an elimination half-life that is greater than or equal to 3 hours (Class 1) or less than 3 hours (Class 0).
    c. **Assay Type:** *In vivo.*
18. Organic cation transporter-2 (OCT2)
    a. **Description**: Renal organic cation transporter 2 (OCT2) is a key transporter located on the basolateral membrane of the renal proximal tubule. The transporter plays an important role in the active secretion of cationic drugs into urine.
    b. **Interpretation**: The predictor will determine if a given compound is likely to be an OCT2 inhibitor (Class 1) or non-inhibitor (Class 0).
    c. **Assay Type:** *In vivo.*

Toxicity

19. AMES Mutagenesis
    a. **Description:** Toxicity is defined by genotoxicity tests, and it is mapped in this endpoint by the Ames test, a widely used method to assess compound mutagenic using bacteria.
    b. **Interpretation:** A positive case indicates mutagenic and therefore may act as a carcinogen.
    c. **Assay Type:** *In vitro.*
20. Avian
    a. **Description:** As avian species are quite sensitive to pesticides and chemicals, they are hence suitable for toxicity studies. In this endpoint, toxicity is mapped across 17 avian species, where data from PA Ecotox database is employed.
    b. **Interpretation:** Lethal concentration 50 (LC50) values are used to determine the toxicity of a given compound. Based on these LC50 values, compounds are separated between toxic and non-toxic**.**
    c. **Assay Type:** *In vivo.*
21. Biodegradation
    a. **Description:** Biodegradation is the process of degradation of organic components. Biodegradable profiles of chemicals are important as these compounds can stay for a long time in the environment, causing distinct adverse effects. In this endpoint, biodegradation is analysed on chemical compounds following the Japanese Ministry of International Trade and Industry (MITI) protocol.
    b. **Interpretation:** Compounds are distinguished between having and not having ready biodegradability based on biological oxygen demand (BOD) values.
    c. **Assay Type:** *In vitro.*
22. Carcinogenesis
    a. **Description:** Toxicity is defined by genotoxicity tests, and it is mapped in this endpoint by the Ames test, a widely used method to assess compound mutagenic using bacteria.
    b. **Interpretation:** A positive case indicates mutagenic and therefore may act as a carcinogen.
    c. **Assay Type:** *In vivo.*
23. Crustacean
    a. **Description:** In this endpoint, toxicity is defined based on the ECOTOX database, a large data set containing diverse compounds with experimental 48h EC50 values on crustaceans. EC50 values refer to the half-maximum effective concentration of a chemical compound, inducing half of its maximum effect.
    b. **Interpretation:** A threshold on EC50 values defines whether the compound would be toxic or not.
    c. **Assay Type:** *In vivo.*
24. Eye Corrosion and Irritation
    a. **Description:** Evaluating the eye irritation or corrosion of chemicals is a valuable risk assessment during compound development and optimisation. For this endpoint, compounds were collected from 20 databases of QSAR Toolbox 3.3.

All the compound IDs of toxic chemicals (corrosives/irritants) were double checked with ChemIDplus Advanced. On the other hand, compounds that do not cause eye irritation/corrosion were retrieved from Verma and Matthew (2015).

b. **Assay Type:** *In vivo.*

25. Fathead Minnow
    a. **Description:** In this endpoint, pesticides or pesticide-like molecules with pLC50 (mg/L) value were collected from the EPA Fathead Minnow Acute Toxicity Database (EPAFHM). It is worth noting that the targets were transformed, where their logarithm values were employed instead of their original values.
    b. **Assay Type:** *In vivo.*
26. hERG Inhibitors
    a. **Description:** The inhibition of the potassium channels encoded by hERG (human ether-go-go gene) is one of the main causes of the emergence of acquired long QT syndrome, which can lead to fatal ventricular arrhythmia. Therefore, analysing if a chemical compound can affect or not these channels is essential for heart toxicity purposes.
    b. Assay Type: *In vitro.*
27. Honey Bee
    a. **Description:** In this endpoint data, pesticides or pesticide-like compounds are assessed in terms of presenting high or low toxicity to honey bees. This dataset was retrieved by Yang et al. from the US EPA ECOTOX Database.
    b. **Interpretation:** To define the categorical values defining toxicity, the authors used the value of Lethal Dose 50% (LD50), where more than 100 μg/bee was considered a high acute toxic compound, whereas values lower than this value were considered as low acute toxic compound.
    c. Assay Type: *In vivo.*
28. Liver Injury I (DILI)
    a. **Description:** Molecular compounds inducing liver injury present major safety concerns while assessing ADMET properties for drugs. This endpoint data is defined upon the analysis of this aspect, mapping when one pathological or physiological liver event (i.e., an event that causes disruption of the normal function of the liver) is strongly associated with the use of a compound. More specifically, this endpoint corresponds to Drug-Induced Liver Injury (DILI) data.
    b. **Assay Type:** *In vivo.*
29. Liver Injury II
    a. **Description:** Molecular compounds inducing liver injury present major safety concerns while assessing ADMET properties for drugs. In this endpoint, a large dataset of *in vivo* hepatotoxicity was retrieved from the literature (e.g., DrugBank) by the authors. Clinical information was accessed through PharmaPendium and enhanced by Leadscope. Only information about hepatotoxicity or non-hepatotoxicity in humans was considered.
    b. **Assay Type:** *In vivo.*
30. Maximum Tolerated Dose

a. **Description:** This endpoint is associated with the definition of the Maximum Recommended Tolerated Dose (MRTD) of chemical compounds in humans. Data comprehends human clinical trials, where the value of the logarithm of MRTD [log(mg/kg/day)] is employed to define the toxicity.
b. **Interpretation:** In these terms, a MRTD of less than or equal to 0.477 [log(mg/kg/day)] is considered to be low (or non-toxic), whereas a value greater than 0.477 [log(mg/kg/day)] is considered to be high (or toxic).
c. **Assay Type:** *In vivo.*

31. Micronucleus
    a. **Description:** Toxicity is defined by genotoxicity tests, where compounds having adverse effects on the process of heredity are detected. In this particular endpoint, the in vivo micronucleus assay, a genotoxicity test method, is used to evaluate the presence/extent or absence of chromosomal damage in human beings.
    b. **Assay Type:** *In vivo*
32. NR-AhR
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Aryl hydrocarbon Receptor (AhR).
    b. Assay type: *In vitro.*
33. NR-AR
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Androgen Receptor (AR).
    b. **Assay type:** *In vitro.*
34. NR-AR-LBD
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Androgen Receptor (AR) Ligand-Binding Domain (LBD).
    b. **Assay type:** *In vitro.*
35. NR-Aromatase
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Aromatase receptor.
    b. **Assay type:** *In vitro.*
36. NR-ER
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Estrogen Receptor (ER).
    b. **Assay type:** *In vitro.*
37. NR-ER-LBD
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Estrogen Receptor (ER) Ligand-Binding Domain (LBD).
    b. **Assay type:** *In vitro.*
38. NR-GR
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Glucocorticoid Receptor (GR).

39. NR-PPAR-gamma
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Peroxisome Proliferator-Activated Receptor Gamma (PPAR-gamma).
    b. **Assay type:** *In vitro.*
40. NR-TR
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Thyroid Receptor (TR).
    b. **Assay type:** *In vitro.*
41. Rat (Acute, LD50)
    a. **Description:** In this endpoint, it is evaluated the toxic potency of a compound, by evaluating its lethal dosage (LD50). This is a standard measurement of acute toxicity. The LD50 is the amount of a compound given to all at once that causes the death of 50% of a group of test animals (rats, in this particular case).
    b. **Assay Type:** *In vivo*
42. Rat (Chronic Oral, LOAEL)
    a. **Description:** Exposure to low-moderate doses of chemicals over long periods of time may be concerning in treatments. LOAEL identifies the minimal necessary dose of a compound that yields an observed adverse effect. NOAEL, in contrast, maps cases with no observed adverse effect. In this endpoint, oral chronic adverse effects are assessed in rats.
    b. **Assay Type:** *In vivo*
43. Respiratory Disease
    a. **Description:** Chemical respiratory toxicity might cause severe harm to the human body. Therefore, it is necessary to better understand it to identify these potential respiratory toxic compounds in advance. In this context, this endpoint encompasses three main databases to define toxicity or safety in the respiratory system: the PNEUMOTOX database (released on March 09, 2020), the ADReCS database (released on May 15, 2019) and the Hazardous Chemical Information System (released on May 09, 2018). Other papers were also used to determine respiratory toxicity or safety. For this data, toxic compounds comprehend cases where adverse effects on the human respiratory system were identified, whereas safe compounds are those considered harmless to the respiratory system, including respiratory non-sensitizers and skin non-sensitizers.
    b. **Assay Type:** *In vivo.*
44. Skin Sensitisation
    a. **Description:** Skin sensitisation is a commonly evaluated adverse effect for dermally applied products. This endpoint evaluates if a compound will cause/induce an allergic reaction while in contact with the dermatitis. Therefore, it will provide guidance for classifying if a compound will yield skin sensitisation or not.
    b. **Assay Type:** *In vivo.*
45. SR-ARE

a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Antioxidant Responsive Element (ARE).
b. **Assay type:** *In vitro.*

46. SR-ATAD5
    a. **Description:** Toxicity is mapped in terms of the molecular compound affecting/not affecting the ATAD5 (ATPase Family AAA Domain Containing 5) gene.
    b. **Assay type:** *In vitro.*
47. SR-HSE
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the Heat Shock Sequence (HSE) elements.
    b. **Assay type:** *In vitro.*
48. SR-MMP
    a. **Description:** Toxicity is mapped in terms of the molecular compound changing/not changing the Mitochondrial Membrane Potential (MMP).
    b. **Assay type:** *In vitro.*
49. SR-p53
    a. **Description:** Toxicity is mapped in terms of the molecular compound activating/not activating the p53 pathway.
    b. **Assay type:** *In vitro.*
50. T. Pyriformis
    a. **Description:** In this endpoint, toxicity is determined when it affects a protozoa bacteria, T. Pyriformis. The concentration of compounds on T. Pyriformis are assessed in terms of inhibition of 50% of growth (IGC50).
    b. **Assay Type:** *In vivo.*
51. Bioconcentration Factor
    a. **Description:** In this endpoint, the molecules are assessed in terms of the ratio of their chemical concentration in biota considering the absorption through the respiratory surface in water at the steady state. The built model on this endpoint data will help in understanding secondary poisoning potential. Its unit is log10(L/kg).
    b. **Assay Type:** *In vivo*
52. Daphnia Magna
    a. **Description:** In this endpoint, compounds are tested by applying them to water and verifying if that results in the death of 50% of Daphnia Magna after 48 hours, i.e., the lethal concentration at 50% in mg/L. The used unit characterising LC 50% is $-\log10[(mg/L)/(1000*MW)]$.
    b. **Assay Type:** *In vivo.*

<u>General Properties</u>

53. Boiling point
    a. **Description**: The temperature (℃) at which a liquid becomes a gas by heating.
    b. **Interpretation**: The unit of boiling point of a molecule is given in ℃.
    c. **Assay type:** *In vitro.*

54. Hydration free energy
    a. **Description**: Hydration free energy measures the solvation free energies between a solute and solvent at ambient temperature, which gives an indication of the drug's aqueous solubility.
    b. **Assay type:** *In vitro.*
55. log D7.4
    a. **Description**: The logarithm of the n-octanol/water distribution coefficients represents the lipophilicity of a molecule at pH=7.4, which can affect aqueous solubility and membrane permeability.
    b. **Interpretation**: The optimal logD7.4 (log mol/L) for orally administered drugs is between 1 and 3 according to Arnott et al., (2012).
    c. **Assay type:** *In vitro.*
56. log P
    a. **Description**: The logarithm of the n-octanol/water distribution coefficients represents the pH-independent lipophilicity of a molecule. Thus, un-ionizable compounds will have similar logP and logD7.4 values.
    b. **Interpretation**: The optimal logP (log mol/L) for orally administered drugs is between 1 and 3 according to Arnott et al., (2012).
    c. **Assay type:** *In vitro.*
57. log S
    a. **Description**: The logarithm of aqueous solubility at a temperature of 20-25 ℃ in log mol/L.
    b. **Interpretation**: More than 80% of the drugs on the market have a log S > -4.
    c. **Assay type:** *In vitro.*
58. log VP
    a. **Description**: The logarithm of the vapour pressure represents the volatility of a molecule at 25 ℃.
    b. **Interpretation**: Gas phase : log VP < 4 and Solid phase > log VP > 8.
    c. **Assay type:** *In vitro*
59. Melting point (MP)
    a. **Description**: The temperature (℃) at which a solid becomes a liquid by heating.
    b. **Interpretation**: The unit of melting point of a molecule is ℃.
    c. **Assay type:** *In vitro.*
60. pKa acid
    a. **Description**: The negative logarithm of the acid dissociation constant (K_a) represents the ratio of conjugate acid to the conjugate base of a molecule that can control its pharmacokinetic properties.
    b. **Interpretation**: -log K_a
61. pKa basic
    a. **Description**: The negative logarithm of the base dissociation constant represents the ratio of conjugate acid to conjugate base of a molecule that can control its pharmacokinetic properties.
    b. **Interpretation**: -log K_b
    c. **Assay type:** *In vitro.*