1. **QSAR ANALYSIS**
	1. **Fit all Models – 3 and 4 parameters**

|  |  |  |
| --- | --- | --- |
| No.Des. | Descriptors | R2 |
| 3 | ALOGPs | ALOGPs-sq | Mv | – | 0.7901 |  |
| 3 | ALOGPs | ALOGPs-sq | J3D | – | 0.7892 |  |
| 3 | ALOGPs | ALOGPs-sq | Mp | – | 0.7836 |  |
| 3 | ALOGPs | ALOGPs-sq | nH | – | 0.7822 |  |
| 3 | ALOGPs | ALOGPs-sq | AMW | – | 0.7768 |  |
| 3 | ALOGPs | ALOGPs-sq | J | – | 0.7680 |  |
| 3 | ALOGPs | ALOGPs-sq | E3u | – | 0.7672 |  |
| 3 | ALOGPs | ALOGPs-sq | ARR | – | 0.7654 |  |
| 3 | ALOGPs | ALOGPs-sq | Density (g/cm3) | – | 0.7615 |  |
| 3 | ALOGPs | ALOGPs-sq | Surface tension (dyne/cm) | – | 0.7571 |  |
| 3 | ALOGPs | ALOGPs-sq | Lop | – | 0.7560 |  |
| 3 | ALOGPs | ALOGPs-sq | Du | – | 0.7558 |  |
| 3 | ALOGPs | ALOGPs-sq | E3e | – | 0.7554 |  |
| 3 | ALOGPs | ALOGPs-sq | Ui | – | 0.7533 |  |
| 3 | ALOGPs | ALOGPs-sq | LogD duodenum (pH 4.6) | – | 0.7516 |  |
| 3 | ALOGPs | ALOGPs-sq | LogD Jejunum & Ileum (pH 6.5) | – | 0.7513 |  |
| 3 | ALOGPs | ALOGPs-sq | LogD (pH 7.2) blood | – | 0.7503 |  |
| 3 | ALOGPs | ALOGPs-sq | LogD (pH 7.4) blood | – | 0.7497 |  |
| 3 | ALOGPs | ALOGPs-sq | LogD (pH 8) colon | – | 0.7458 |  |
| 3 | ALOGPs | ALOGPs-sq | Me | – | 0.7394 |  |
| 4 | ALOGPs | ALOGPs-sq | nCIC | J3D | 0.8160 |  |
| 4 | ALOGPs | ALOGPs-sq | nH | J | 0.8152 |  |
| 4 | ALOGPs | ALOGPs-sq | AMW | J | 0.8151 |  |
| 4 | ALOGPs | ALOGPs-sq | AMW | J3D | 0.8141 |  |
| 4 | ALOGPs | ALOGPs-sq | J3D | Ui | 0.8140 |  |
| 4 | ALOGPs | ALOGPs-sq | Density (g/cm3) | J3D | 0.8138 |  |
| 4 | ALOGPs | ALOGPs-sq | Density (g/cm3) | J | 0.8121 |  |
| 4 | ALOGPs | ALOGPs-sq | Parachor (cm3) | nH | 0.8099 |  |
| 4 | ALOGPs | ALOGPs-sq | Molar refractivity (cm3) | nH | 0.8085 |  |
| 4 | ALOGPs | ALOGPs-sq | Polarizability (cm3) | nH | 0.8084 |  |
| 4 | ALOGPs | ALOGPs-sq | Molecular Volume (cm3) | nH | 0.8084 |  |
| 4 | ALOGPs | ALOGPs-sq | Me | Mp | 0.8080 |  |
| 4 | ALOGPs | ALOGPs-sq | LogD (pH 6.5) | AMW | 0.8069 |  |
| 4 | ALOGPs | ALOGPs-sq | LogD (pH 7.2) | AMW | 0.8068 |  |
| 4 | ALOGPs | ALOGPs-sq | LogD (pH 7.4) | AMW | 0.8068 |  |
| 4 | ALOGPs | ALOGPs-sq | LogD (pH 8) | nH | 0.8063 |  |
| 4 | ALOGPs | ALOGPs-sq | LogD (pH 8) | AMW | 0.8062 |  |
| 4 | ALOGPs | ALOGPs-sq | LogD (pH 4.6) | AMW | 0.8062 |  |
| 4 | ALOGPs | ALOGPs-sq | LogD (pH 7.4) | nH | 0.8056 |  |
| 4 | ALOGPs | ALOGPs-sq | LogD (pH 7.2) | nH | 0.8054 |  |

Table S2 - 20 Best fitted 3 and 4 parameter models, ranked by R2 values.

4 parameter models are fitted with a subset of descriptors.

* 1. **Descriptor Definitions**

ALOGPs – octanol/water partition coefficient

ALOGPs-sq – octanol/water partition coefficient squared

AMW - average molecular weight

ARR – aromatic ratio

De - D total accessibility index / weighted by Sanderson electronegativity

Density (g/cm3)

Du - D total accessibility index / unweighted

E3e - 3rd component accessibility directional WHIM index / weighted by Sanderson electronegativity

E3u - 3rd component accessibility directional WHIM index / unweighted

J - Balaban distance connectivity index

J3D - 3D-Balaban index

LogD (pH 7.2) – distribution constant, calculated at pH 7.2

LogD (pH 7.4) blood – distribution constant, calculated at pH 7.4

LogD (pH 8) colon – distribution constant, calculated at pH 8

LogD duodenum (pH 4.6) – distribution constant, calculated at pH 4.6

LogD Jejunum & Ileum (pH 6.5) – distribution constant, calculated at pH 6.5

Lop - Lopping centric index

Molecular Volume (cm3)

Mp - mean atomic polarizability (scaled on Carbon atom)

Mv - mean atomic van der Waals volume (scaled on Carbon atom)

nCIC - number of rings (cyclomatic number)

nH - number of Hydrogen atoms

Parachor (cm3)

Polarizability (cm3)

Surface Tension (dyne/cm) - surface free energy

Ui - unsaturation index

* 1. **Model Fits**

|  |  |  |
| --- | --- | --- |
|  |  | **Coefficients** |
|  |  |  | **Linear Fit** | **Bootstrap** |  | **Linear Fit** | **Bootstrap** |  | **Linear Fit** | **Bootstrap** |  | **Linear Fit** | **Bootstrap** |  | **Linear Fit** | **Bootstrap** |
| **Model parameters** | **R^2 value** | **Intercept** | **2.5% CI** | **97.5%** **CI** | **2.5% CI** | **97.5%** **CI** | **ALogPs** | **2.5% CI** | **97.5%** **CI** | **2.5% CI** | **97.5%** **CI** | **ALogPs****^2** | **2.5% CI** | **97.5%** **CI** | **2.5% CI** | **97.5%** **CI** | **3rd Param** | **2.5% CI** | **97.5%** **CI** | **2.5% CI** | **97.5%** **CI** | **4th param** | **2.5%** **CI** | **97.5%** **CI** | **2.5%** **CI** | **97.5%** **CI** |
| ALogPs, ALogPs^2 **(1)** | 0.6292 | -0.579 | -1.165 | 0.008 | -1.108 | -0.086 | 1.203 | 0.903 | 1.504 | 0.904 | 1.470 | -0.133 | -0.168 | -0.098 | -0.166 | -0.093 |  |  |  |  |  |  |  |  |  |  |
| ALogPs, ALogPs^2,nH **(2)** | 0.782 | -1.659 | -2.276 | -1.041 | -2.454 | -1.033 | 1.310 | 1.073 | 1.547 | 1.038 | 1.601 | -0.176 | -0.207 | -0.144 | -0.213 | -0.130 | 0.065 | 0.040 | 0.090 | 0.043 | 0.087 |  |  |  |  |  |
| ALogPs, ALogPs^2, Mv **(3)** | 0.7901 | 3.362 | 1.838 | 4.887 | 2.159 | 4.419 | 1.372 | 1.135 | 1.61 | 1.126 | 1.579 | -0.158 | -0.186 | -0.129 | -0.19 | -0.123 | -6.616 | -9.063 | -4.168 | -8.432 | -4.473 |  |  |  |  |  |
| ALogPs, ALogPs^2, J3D**(4)** | 0.789 | -2.372 | -3.176 | -1.569 | -3.071 | -1.648 | 1.339 | 1.104 | 1.574 | 1.109 | 1.556 | -0.148 | -0.175 | -0.120 | -0.177 | -0.115 | 0.733 | 0.461 | 1.006 | 0.495 | 0.960 |  |  |  |  |  |
| ALogPs, ALogPs^2, ARR **(5)** | 0.765 | -0.222 | -0.719 | 0.274 | -0.647 | 0.070 | 1.366 | 1.114 | 1.618 | 1.117 | 1.605 | -0.153 | -0.182 | -0.123 | -0.186 | -0.116 | -1.675 | -2.387 | -0.963 | -2.290 | -1.065 |  |  |  |  |  |
| AlogPs, ALogPs^2, AMW **(6)** | 0.777 | 0.731 | 0.035 | 1.427 | -0.113 | 1.368 | 1.330 | 1.089 | 1.572 | 1.104 | 1.536 | -0.154 | -0.183 | -0.125 | -0.182 | -0.121 | -0.200 | -0.279 | -0.120 | -0.293 | -0.133 |  |  |  |  |  |
| ALogPs, ALogPs^2, nCIC, J3D (**7**) | 0.816 | -5.105 | -7.579 | -2.632 | -7.681 | -2.694 | 1.284 | 1.056 | 1.511 | 1.087 | 1.493 | -0.146 | -0.172 | -0.120 | -0.173 | -0.116 | 0.411 | 0.057 | 0.764 | 0.064 | 0.796 | 1.587 | 0.808 | 2.367 | 0.796 | 2.330 |
| ALogPs, ALogP^2, AMW, J3D (**8**) | 0.814 | -1.018 | -2.453 | 0.417 | -2.309 | 0.429 | 1.361 | 1.137 | 1.586 | 1.123 | 1.591 | -0.154 | -0.181 | -0.127 | -0.185 | -0.119 | -0.110 | -0.209 | -0.011 | -0.218 | -0.051 | 0.475 | 0.127 | 0.823 | 0.148 | 0.781 |
| ALogPs, ALogPs^2, nH, J (**9**) | 0.815 | -3.067 | -4.304 | -1.830 | -4.133 | -1.767 | 1.432 | 1.191 | 1.673 | 1.196 | 1.653 | -0.172 | -0.202 | -0.142 | -0.205 | -0.128 | 0.044 | 0.015 | 0.728 | 0.014 | 0.071 | 0.904 | 0.202 | 1.607 | 0.173 | 1.475 |
| ALogPs, ALogPs^2, Polarizability, nH (**10**) | 0.808 | -0.983 | -1.822 | -0.144 | -1.736 | 0.078 | 1.307 | 1.082 | 1.532 | 1.032 | 1.519 | -0.165 | -0.197 | -0.132 | -0.199 | -0.110 | -0.022 | -0.041 | -0.002 | -0.036 | 0.005 | 0.063 | 0.039 | 0.087 | 0.040 | 0.087 |

Table S3. Coefficients & Confidence intervals for Linear and Bootstrap fits for 2, 3 and 4 parameter models.

* 1. **Model fit plots**



Figure S166. Plots showing the predicted vs actual values for Log(1/EC50) for 10 of the models fitted