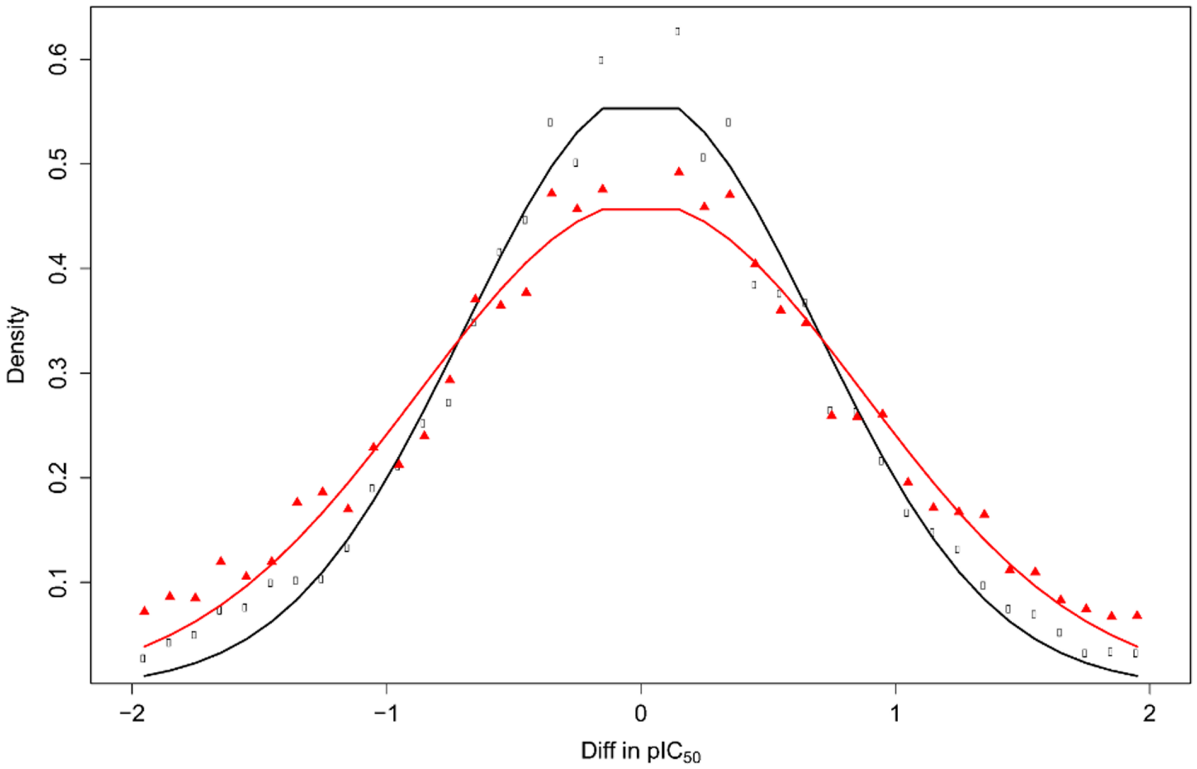

Figure 5. Fitted Gaussian distribution of ΔpIC_{50} (red) and ΔpK_i (black). The Gaussian distributions shown were fitted to all ΔpActivity values with an upper threshold ΔpActivity = 2.0. Standard deviations for the fitted Gaussian distributions are $\sigma_{pIC50} = 0.87$ and $\sigma_{pKi} = 0.69$. Note that since the σ here is calculated from pairs of measurements each containing experimental uncertainty and other sources of variability, it has to be divided by $\sqrt{2}$ in order to obtain the true σ of the individual measurements [13].
doi:10.1371/journal.pone.0061007.g005

Based on the Cheng-Prusoff equation and under the assumption of a competitive mechanism of action, pK_i values are larger or equal to pIC_{50} values. However due to unknown mechanism, experimental uncertainty and some database annotation errors in the data, there are a significant number of pairs where the pIC_{50} is larger than the pK_i. On average, the measured pK_i values are 0.355 log units larger than the measured pIC_{50} values, corresponding to a factor of 2.3. A factor of 2 is in agreement with a balanced assay condition in which the substrate concentration is equal to the K_m value. This is often used in order to allow the detection of inhibitors with different mechanism of action.

After subtracting 0.35 log units from the pK_i values and correcting by $\sqrt{2}$, pK_i and pIC_{50} values agree with an $R^2 = 0.46$, $\sigma = 0.68$, $MUE = 0.54$ and $M_{ed}UE = 0.43$. The standard deviations of Gaussian distributions fitted to the inner part with an upper threshold of 1.5, 2.0 and 2.5 ΔpActivity units are 0.79, 0.83, and 0.85.

Overall, this is close to or even slightly better than the agreement obtained for pIC_{50} values with themselves. Therefore we can conclude that pK_i values can be used to augment pIC_{50} values without any loss of quality, if they are corrected by an offset. In the absence of assay information, the best guess for the conversion factor between K_i into IC_{50} is extrapolated from the average offset calculated from the heterogeneous ChEMBL data, i.e. a factor of 2.3, corresponding to 0.35 pActivity units.

Discussion

In this contribution we show how the comparability of IC_{50} data can be analyzed using the public ChEMBL database. We find that when comparing all independently measured pIC_{50} data, the variability found for pIC_{50} data is approximately 25% larger than the variability found for pK_i data, with $\sigma_{pIC50} = 0.68$, $MUE_{pIC50} = 0.55$ and $M_{ed}UE_{pIC50} = 0.43$. These values correspond to the most probable variability of pIC_{50} data mixing from different (unknown) assays.

We want to stress that pIC_{50} data from different assays can only be compared under certain conditions. However, as discussed in the introduction, this is often done in large-scale data analysis. A standard deviation of 0.68 corresponds to a factor of 4.8, meaning that 68.2% of all IC_{50} measurements agree within a factor of 4.8, even when measured in different laboratories under potentially different assay conditions. One reason why the variability of IC_{50} data is found only moderately higher than the variability of K_i data might be that practically most of the IC_{50} assays may have been run using very similar assay protocols. Unfortunately, the assay descriptions available within ChEMBL are too terse to permit analyzing this any further.

IC_{50} values measured in the same laboratory usually show a better reproducibility. From our in-house database, we extracted series of reference pIC_{50} values measured for assay standards. The plots in Figure 9 show the pIC_{50} values measured for rolipram on PDE4D and cilostamide on PDE3. The standard deviation of the pIC_{50} values are $\sigma = 0.22$ for rolipram/PDE4D and $\sigma = 0.17$ for cilostamide/PDE3.

There is some variation over time which could indicate changes in the assay conditions and solution handling. We also tried to find public series of at least ten compounds that have been measured in independent parallel assays. However, such series did not exist within ChEMBL as all the series we found were either measured in