Supplementary Material

Challenges of Antibacterial Drug Discovery

Folkert Reck, Johanna M. Jansen, and Heinz E. Moser

Novartis Institute for BioMedical Research, 5300 Chiron Way, Emeryville, CA 94608, U.S.A.

Email: heinz.moser@novartis.com

Table of Contents

Calculation of logD values (Figure 1) S2
Additional information to Figure 4 ..... S2
Additional information to Figure 5 ..... S2
Experimental determination of logD values S2
Calculated versus experimentally determined logD values S3
References S6

Calculation of logD values (Figure 1)

Calculations for logD at pH 7.4 use the Biobyte clogP (version 4.3, http://www.biobyte.com/index.html) and a pKa prediction based on an internally retrained MoKa model (Gedeck et al, 2015, reference 14 in main paper) for pKA, using MoKa version 2.5.4 (http://www.moldiscovery.com/software/moka/). For 18 molecules that failed the retrained MoKa pKa calculation (because of molecular size), we inserted a logD prediction from Pipeline Pilot (current software available here: http://www.3dsbiovia.com/products/collaborative-science/biovia-pipeline-pilot/). All 210 molecules in Figure 1 are included in the Excel dataset that is also made available as supplementary material.

1. Additional information to Figure 4

Primary screening data was collected for three high throughput screening campaigns against *P. aeruginosa*, yeast and human HaCat cells. The *P.aeruginosa* strain was an efflux defective derivative of strain PAO1 (*ΔmexAB-oprM* strain K1119[1]) and the screen was run at 40 µM for 800k compounds. The readout from the screen was growth inhibition. This screening campaign was compared against a primary screen in yeast (run at 20 µM for 800k compounds) and in a human HaCat cell line (run at 24 µM for 800k compounds). Hits were called at > 50% inhibition for all three campaigns. Hit-rates were calculated for sets of compounds binned by their clogD. Calculations for logD at pH 7.4 were done in Pipeline Pilot during 2015; current software available here: http://www.3dsbiovia.com/products/collaborative-science/biovia-pipeline-pilot/

2. Additional information to Figure 5

We collected a set of 838 validated inhibitors of wildtype Gram-negative (GN) bacteria that had been assessed for cytotoxic activity against at least one of two mammalian cell-lines (HepG2 and K562). A validated inhibitor was defined as a compound with an EC_{50} < 20 µM against at least one of the following bacteria: *E. coli* (Eco01, ATCC strain 25922), *K. pneumoniae* (Kpn02, ATCC43815) or *P. aeruginosa* (PAO1[2]). These compounds were collected from our historical knowledgebase spanning many years and many projects. Compounds were labeled as not cytotoxic if cytotoxicity EC_{50} / GN EC_{50} ≥ 5 OR if cytotoxicity EC_{50} has a ">" qualifier. This condition had to be met for all bacterial strains that were tested (up to 13, including efflux mutants) in order to be labeled as not cytotoxic; if a compound was cytotoxic against one but not both mammalian cell-lines, it was still labeled as cytotoxic. In this data set, 83% of all compounds with clogD > 3 are cytotoxic (241 out of 292), and only 38% of the compounds in the full set are free of cytotoxicity (316 out of 838). Values for logD were calculated using the in-house model described for Figure 1.

3. Experimental determination of logD values

10 µl of a 10 mM DMSO stock solution for a given compound is placed in a well of a 96 well plate. 10 µl of 2 mM Halodipine (logD = 3.0) in DMSO is added to each well as an internal standard followed by mixing. The DMSO is removed by lyophilization overnight. 250 µl of octanol-saturated PBS buffer (pH 7.4) and 250 µl of PBS buffer (pH 7.4) saturated octanol is added to each well. The plate is sealed and vortexed at 600 rpm overnight. 10 µl of the octanol phase is removed, diluted 1:100 with DMSO and 1 µl is injected into the LC/UV/qTOF using a reversed phase column with an 98 to 2% gradient of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). Gradient: 1 minute at 98% solvent A, linearly changing in 0.6 minutes to 98% solvent B and keeping it for another 0.8 minutes before switching back to 98% solvent A. 10 µl of the aqueous phase is injected for analysis as well. Injection volumes and concentrations are dependent on compound distribution between the two phases and might have to be adapted for more polar compounds.

The obtained data is processed with ProfileLynx. Mass chromatograms are integrated and the ratio between peaks from aqueous and organic phase determined after correction for dilution and injection volumes and area ratio of the internal standard. Typically logD values are determined reliably between values of -2 and +5.

4. Calculated versus experimentally determined logD values

The following Figures represent logD values at pH 7.4 either determined experimentally or calculated by the two methods indicated. The "Standard MoKa" calculation uses the MoKa software as provided by the vendor (http://www.moldiscovery.com/software/moka/). The "NIBR MoKa" calculation is the same as described under item 1 above and uses the retrained MoKa pKa model. This includes the experimental pKa values from the NIBR knowledgebase added to the training set for the Standard MoKa calculation (Gedeck et al, 2015, reference 14 in main paper). All compounds are represented (both CP targets) for which the experimental values were determined, lacking a qualifier.

The NIBR MoKa model results in tighter r^2 between experimental and predicted logD for both projects compared to Standard MoKa. For the first CP target, the logD prediction using NIBR MoKa is offset, but r^2 is better than for Standard MoKa, so it is better for comparative assessment between molecules (rank ordering). This analysis therefore shows that expansion of the training set for pKa prediction improves the performance of the subsequent logD prediction for the projects in this paper.

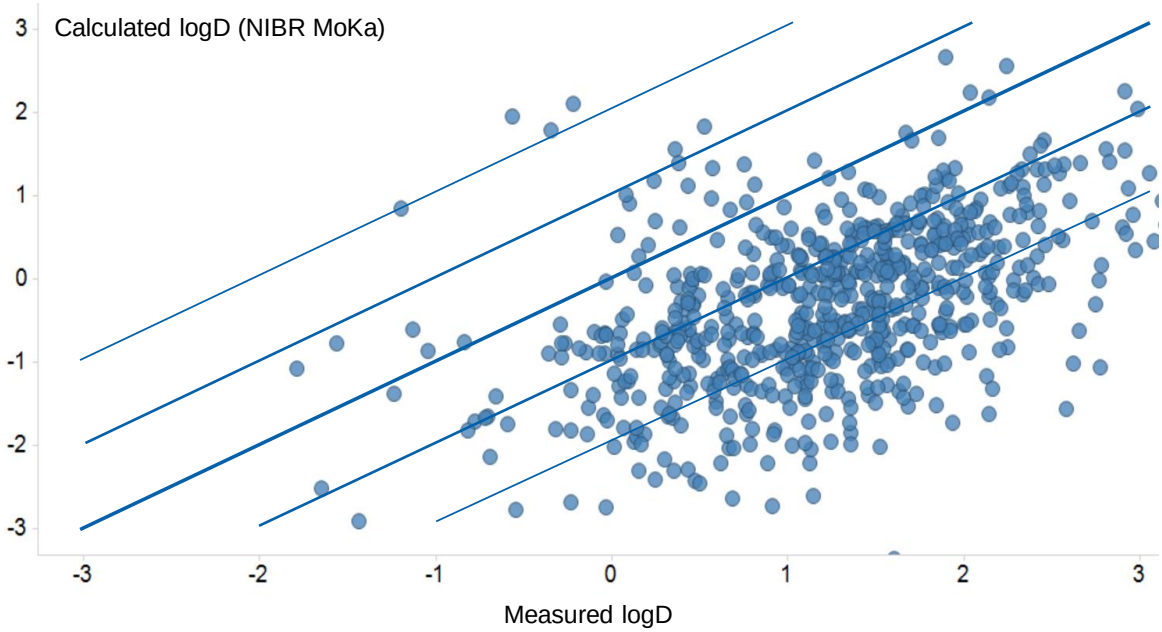

Figure S1. Calculated (NIBR MoKa) versus measured logD values at pH 7.4 for 693 compounds (first CP target, Figure 2b).

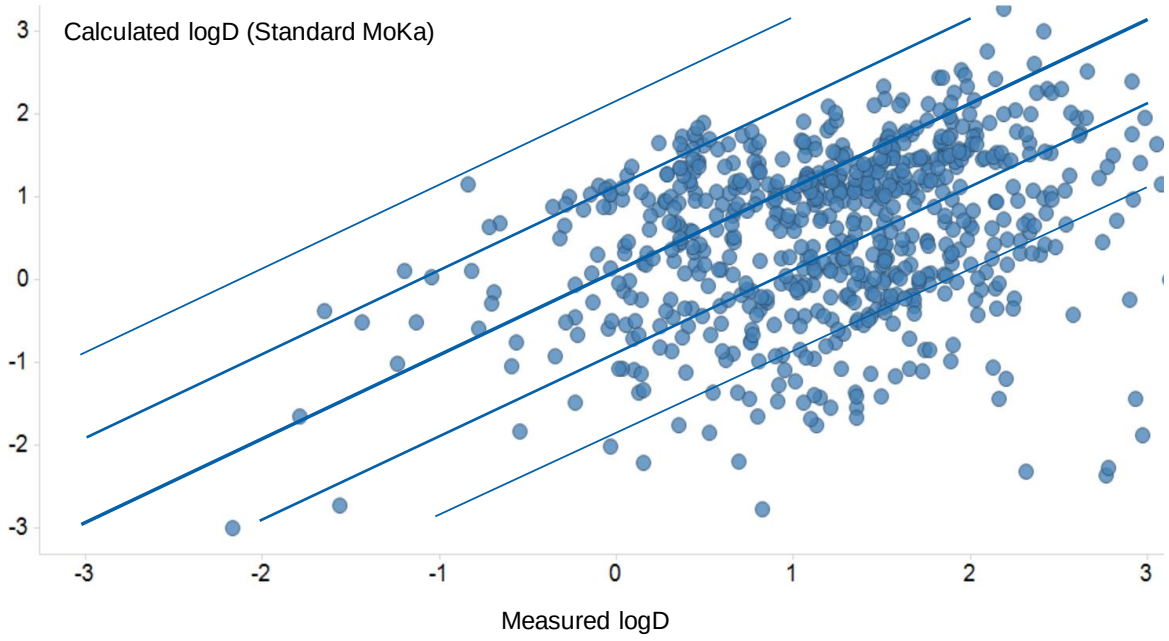

Figure S2. Calculated (Standard MoKa) versus measured logD values at pH 7.4 for 693 compounds (first CP target, Figure 2b).

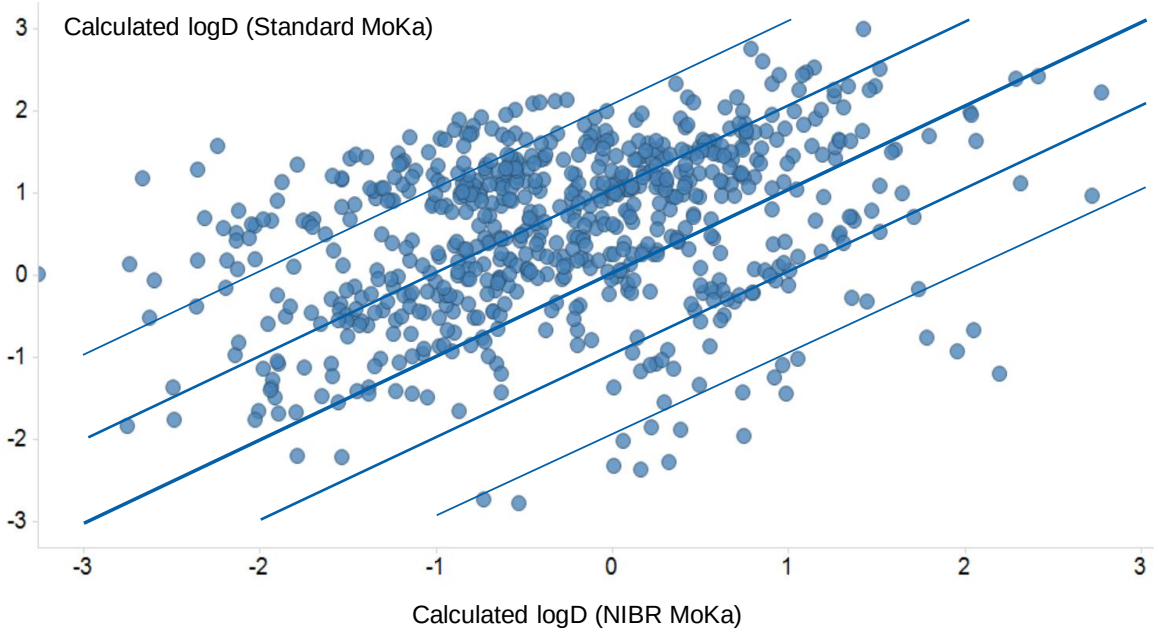

Figure S3. Calculated (Standard MoKa) versus calculated (NIBR MoKa) logD values at pH 7.4 for 693 compounds (first CP target, Figure 2b).

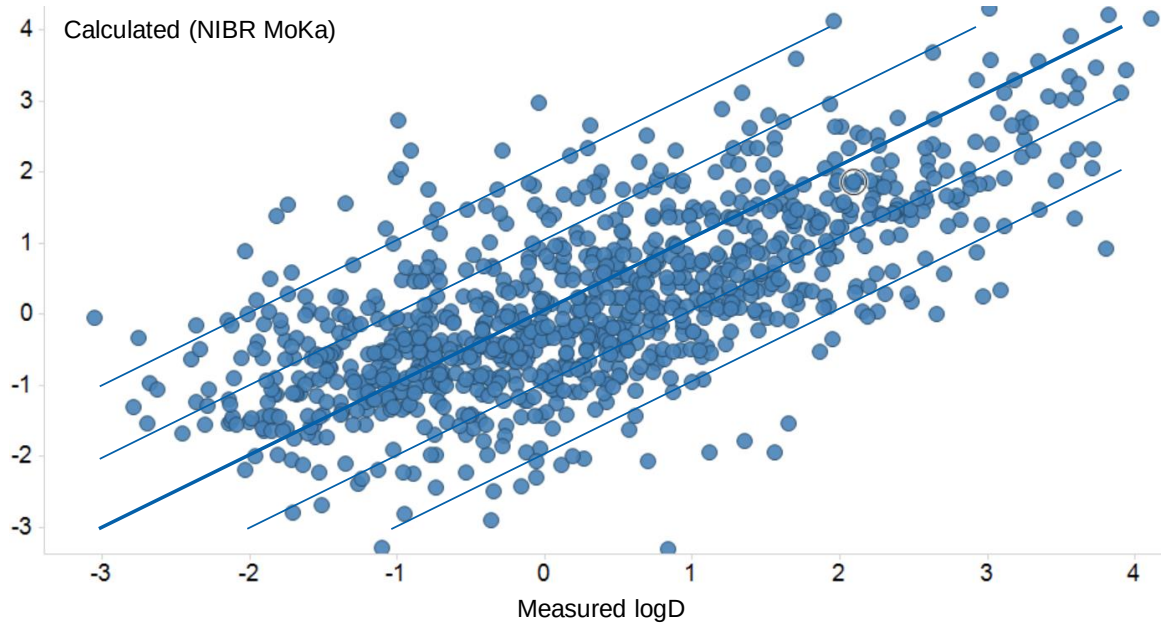

Figure S4. Calculated (NIBR MoKa) versus measured logD values at pH 7.4 for 1032 compounds (second CP target, Figure 2c and Figure 3).

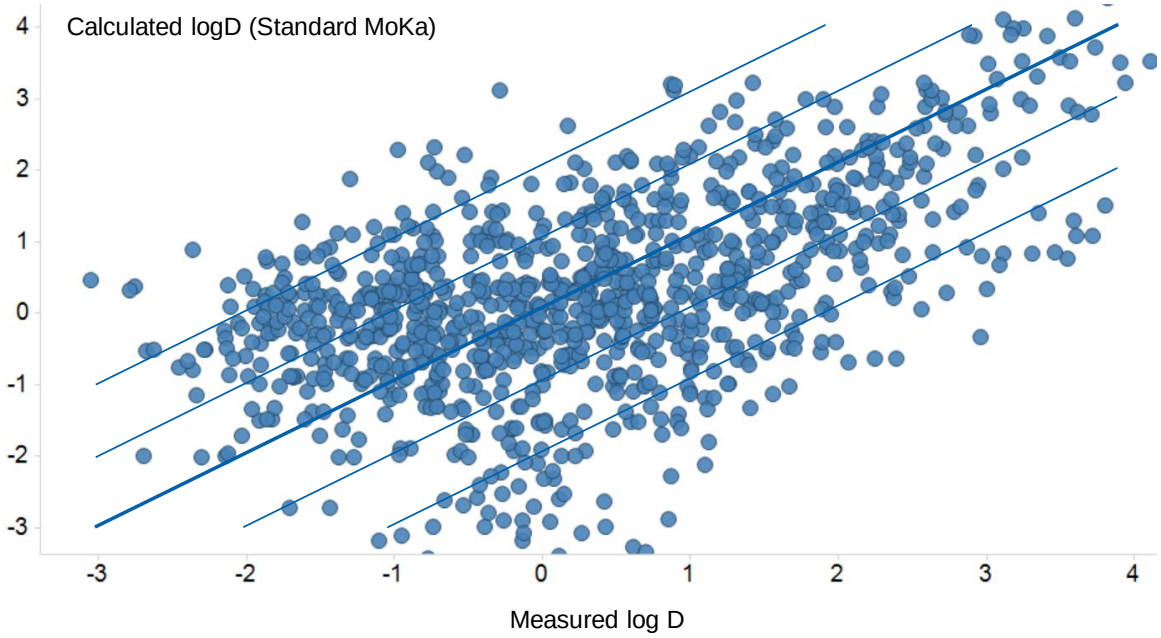

Figure S5. Calculated (Standard MoKa) versus measured logD values at pH 7.4 for 1032 compounds (second CP target, Figure 2c and Figure 3).

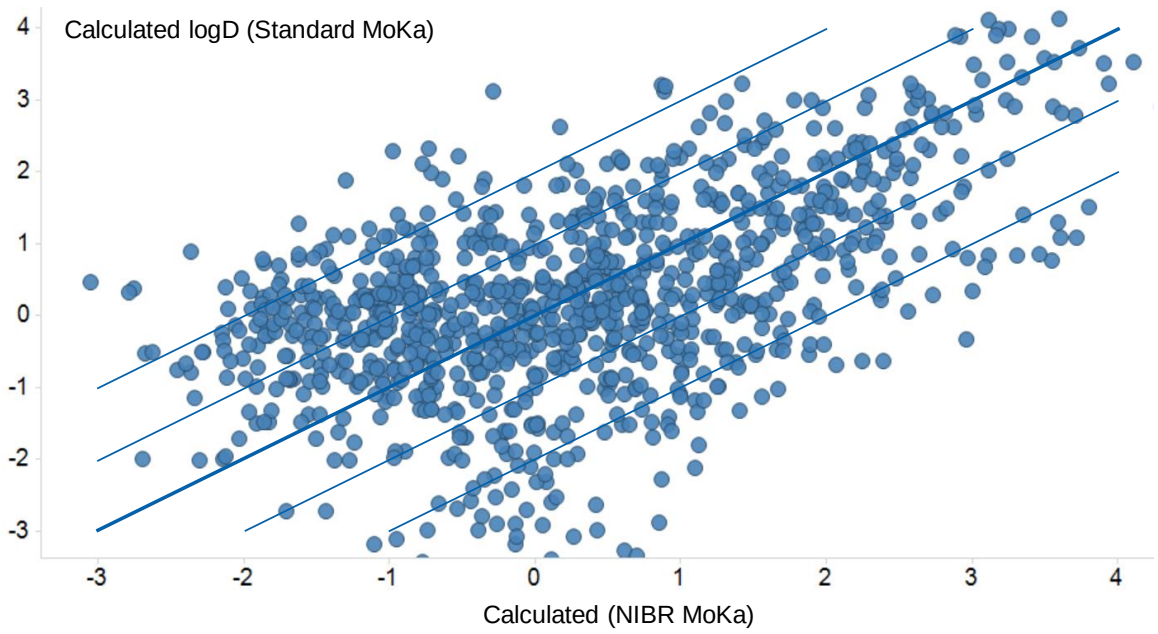

Figure S6. Calculated (Standard MoKa) versus calculated (NIBR MoKa) logD values at pH 7.4 for 1032 compounds (second CP target, Figure 2c and Figure 3).

References

1. Li, X.Z.; Zhang, L.; Srikumar, R.; Poole, K. *Antimicrob. Agents Chemother.* **1998**, *42*, 399.
2. Masuda, N.; Ohya, S. *Antimicrob. Agents Chemother.* **1992**, *36*, 1847.