

Knowledge instead of ignorance

More efficient drug design through
the understanding of experimental uncertainty



**Prof. Dr Christian Kramer,
Institute of General, Inorganic
and Theoretical Chemistry,
University of Innsbruck, Austria**

Biology is naturally complex, and even the results of the simplest biochemical experiments are afflicted with experimental noise that cannot be ignored. However, biochemical measurements are the backbone of modern pharmaceutical research. If the experimental uncertainty is underestimated, biochemical data can be very easily over-interpreted. An appropriate consideration of experimental uncertainty can be achieved with very little additional effort, and this helps to differentiate knowledge from ignorance and to avoid taking wrong tracks that can be time-consuming and expensive.

Experimental uncertainty in the sciences

The probability of new knowledge can be accurately quantified in physics: On 31.7.2012, documented evidence of the existence of the last unknown elementary particle, the long-sought-after Higgs Boson appeared on the Preprint Server arXiv.org – with a significance of 5.9 standard deviations [1]. In climate science, too, experimental uncertainty is part of all predictive knowledge: In its fourth report, the Working Group I of the Intergovernmental Panel on Climate Change (IPCC) published the exact terminology to be used for translating quantified uncertainties and probabilities into words [2]. These two examples from very different areas illustrate that scientific knowledge and forecasts often cannot be presented in a yes/no, right/wrong, black/white way, but are subject to a particular degree of uncertainty.

In pharmaceutical research, the vast majority of results from studies and measurements are part of a broader distribution: The best-known cases are the results of clinical studies that are assessed according to whether the health of test persons receiving a new medication improves significantly when compared to a group receiving placebos or the current therapeutical gold standard. In order to be able to accurately answer these questions – questions on whose answers depends health and a great deal of money – statistics is required to differentiate true effectiveness from spurious correlations. In fact, statistics and experimental uncertainty play a crucial part in the non-clinical hit-to-lead and lead optimisation phases.

Rational computer-based drug design

Classic drug design runs in iterative cycles of chemical synthesis and biological testing, which often extend over

many years. In the early phase of the drug design process, the primary aim is to develop a substance that binds to the target protein with a high affinity. In later phases, other properties such as selectivity towards other proteins that can impart toxic side-effects, and the absorption, distribution, metabolism and excretion (ADME) properties of the potential drug are also optimized.

Within one cycle, the next substances to be synthesised and tested are either selected by trial-and-error, or they are selected in a rational way. Rational selection is ideally based on all the chemical and biological knowledge previously gathered on the target protein and ADME-Tox properties. Rational selection allows developing optimized clinical candidates more quickly than pure trial-and-error. A key component of rational selection strategies is Computer-Aided Drug Design. Its job is to bundle all of the existing knowledge and to search for the most promising chemical modifications. Through computer-aided design, the number of design cycles can be significantly reduced and a great deal of time and money saved. In 2013, Martin Karplus, Michael Levitt and Arieh Warshel received the Nobel Prize for Chemistry in 2013 for their contributions to understanding the chemical forces that drive protein-ligand recognition, which is nowadays part of the basis for computer-based drug design.

Binding constants: The number-one criterion in rational drug design

Most of the important properties to be optimised during the different drug design phases are measured using protein-ligand binding constants. The dissociation constant K_d of the protein-ligand complex is linked to the Gibbs free energy for binding ΔG^0 according to

$$\Delta G^0 = -RT \cdot \ln K_d$$

where T stands for the temperature, and R is the ideal gas constant. In biochemical assays, it is frequently not the K_d values that are determined but IC_{50} or K_i values. IC_{50} values are the ligand concentration the function of the protein is reduced by 50%. With a few constraints, K_i values, the dissociation constants of enzyme inhibitors, can be calculated from IC_{50} values.

At room temperature, the difference in $K_{i/d}$ by a factor of ten corresponds to a difference in binding energy of around 1.4 kcal/mol. Often, only small differences far below 1.4 kcal/mol are achieved through modifications in the chemical structure. These are marginal cases, where rules of thumb are used to assess the significance of observed differences. The individual rules of thumb vary strongly, depending on the past history of the user.

As biology is complex per se and many factors influence the outcome of biochemical assays, the measured binding constants contain a certain amount of experimental uncertainty. If the experimental uncertainty is underestimated, there is a risk that small differences in binding constants are over-interpreted and structure activity relationships are deduced where none exist. On the other hand, if the experimental uncertainty is overestimated, signals present in the data will not be optimally used. The two situations of over- and under-estimating experimental uncertainty cost time and money. They slow down the design of drugs because the project team is either not using all the information contained in the data, or focussing on spurious facts.

How much experimental uncertainty do binding constants contain?

The binding constants stated in the scientific literature fluctuate considerably, but generally underestimate the actual variation in measured values. An impression of published inaccuracies can be gained from the CSAR NRC-HiQ dataset (www.csardock.org). Here, published biochemical affinities, including standard deviations, have been gathered for 157 diverse chemical and biochemical protein-ligand systems. The median of the published standard deviations is $0.044 \log K_{d/i}$, with the smallest values being 0.001 and $0.002 \log K_{d/i}$. For every scientist who has already tried to reproduce the K_i values from the literature, it is clear that these experimental uncertainties are much too low.

A more realistic idea about the experimental uncertainty can be obtained from comparing K_i values which have been independently measured for the same protein- ligand systems. Figure 1 shows such a comparison of all independently measured K_i values from the ChEMBL database [3].

Assuming a simple normal distribution for the experimental errors, an experimental inaccuracy for heterogeneous K_i values of $0.54 \log K_i$ units can be calculated from this comparison [4]. This means that two independent K_i measurements for the same protein-ligand system can be found with around 68% probability within an interval of $\pm 0.54 \log K_i$ units. K_i values have to be comparable because they are physical binding constants. For the more frequently measured IC_{50} values, the standard deviation of the experimental variation is 0.69 log units [5]. IC_{50} values measured in different experimental setups do not have to be comparable. Nevertheless, in practice they are frequently compared with each other, for example in selectivity considerations. For chemical standards, which were frequently measured in the same assay at Novartis in Basel, we calculated an experimental uncertainty with a standard deviation of 0.18 to 0.35 log units, depending on the system and experimental structure [5]. This is equivalent to a factor of 1.5 to 2.2. From a scientific point of view, the reasons for the comparatively high experimental uncertainty are rather poorly understood. Faults in the measuring devices that assess the biological signal appear to be the least significant problem, as can be

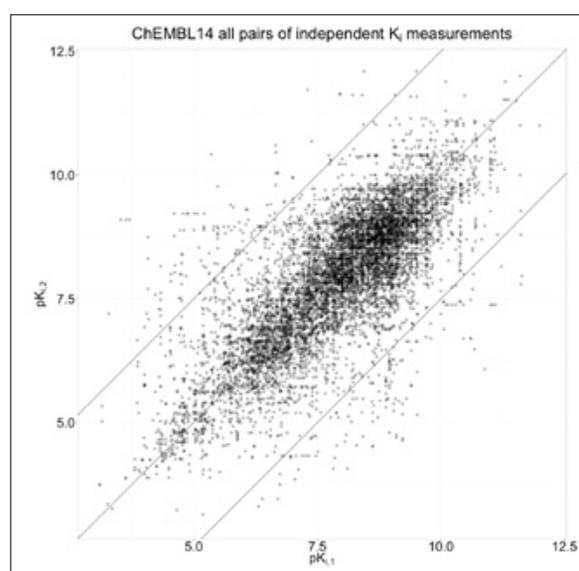


Fig. 1 Pairs of independently measured pK_i values on the same protein-ligand system from ChEMBL14. In total, there are 8.524 pairs for 2.046 protein-ligand systems in ChEMBL14. The diagonals indicate the line of identical measurements and the boundaries at which the differences are more than 2.5 log units [4]. The correlation for all pairs with less than 2.5 log units is $R^2 = 0.66$.



Christian Kramer, born in 1980, studied Molecular Sciences in Erlangen and Zürich. He did his doctorate between 2007–2009 at the University of Erlangen in close collaboration with Boehringer-Ingelheim/Biberach, developing novel QSAR and QSPR methods for the statistical prediction of physicochemical and biochemical properties. From 2010 to 2013, he was Presidential PostDoc at Novartis in Basel, with Peter Gedeck in the Computer-Aided Drug Design Group. Here he was involved in developing statistical approaches in docking/scoring, multipole force fields and investigating the integrity of biochemical data and. He is Assistant Professor at the University of Innsbruck since 2013. His research interests are focused on the computer-based transformation of chemical and biochemical data into scientific knowledge for molecular design. The main emphases of his work are the linking of quantum-mechanical- and force-field methods with the statistical evaluation of databases in the discovery and optimization of leads. He is the author of 24 scientific publications and holds many awards, including a Carl Duisberg Scholarship, a Novartis Presidential PostDoc Fellowship and two internal Novartis awards.

deduced from the small uncertainties (repeatability) reported in the literature. Other possible reasons for high levels of uncertainty include the quality and stability of the biological material, the purity of the measured chemical substances, aggregation of the active ingredients and variations in temperature, air humidity and pressure. One further source of errors, which should not be underestimated, are errors in the dilution series. Some badly soluble substances remain adhering to the walls of the pipette during the dilution process, which leads to concentrations that are too low by orders of magnitude, especially for higher dilutions. Ekins et al. have recently shown that structural interpretations based on such data may be completely incorrect [6].

Tightening the thresholds: How experimental uncertainty influences modelling

Here, I will use two examples to show how the experimental uncertainty can appropriately be taken into account in data analysis and modelling.

A standard application in computer-based drug design involves QSAR and docking models. Here, various structural chemical and biochemical properties are correlated with the measured activity. The quality of such models is usually quantified using the R^2 , the fraction of the explained variance of the measured data. If part of the measured variance consists of experimental uncertainty (noise), the maximum explainable part of the variance R^2_{max} can be calculated according to

$$R^2_{max} = 1 - \frac{\sigma_{noise}^2}{\sigma_{tot}^2}$$

where σ_{noise} is the standard deviation of experimental uncertainty and σ_{tot} is the standard deviation of the entire measured data. Thus, R^2 can only be interpreted if the experimental uncertainty of the data is known. Depending on the signal-to-noise ratio, R^2_{max} can become very small.

A second example for the importance of experimental uncertainty is matched molecular pair analysis (MMPA). MMPA is a method for chemical knowledge extraction from huge databases and is increasingly being applied in lead optimisation. Here, activity differences between two molecules are compared with the differences in chemical substitutions. For carrying out MMPA, a large set of binding data is assembled for molecule pairs which all differ by the same exchange of a functional group.

From the distribution of the activity differences, predictions about the future effects of the same functional group exchange on new molecules are made. As an example, Figure 2 shows the distribution of the affinity differences from the hERG channel for all pairs present in ChEMBL14, where a fluorine is converted into a chlorine.

The accuracy of MMPA predictions crucially depends on the standard deviation of the activity differences. The smaller the standard deviation, the more accurate the prediction. However, the standard deviation can never be zero due to the omnipresent experimental uncertainty. The minimal standard deviation for the pairs $\sigma_{pairs,min}$, which is to be expected due to the experimental uncertainty, can be calculated[7] according to.

$$\sigma_{pairs,min} = \sqrt{2} \cdot \sigma_{noise}$$

Assuming an experimental uncertainty $\sigma_{noise} = 0.2$ log units for hERG measurements from the same laboratory, this results in a minimum standard deviation $\sigma_{pairs,min}$ of 0.28 log units for the pairs – very close to the observed standard deviation of the hERG affinity differences of 0.33 for the F>>Cl transformation. The observed standard deviation can thus be almost completely explained through experimental inaccuracy and, unlike other transformations with higher standard deviations, there is no indication from the database that specific environmental effects influence the differences. The next exciting step is to now verify the binding constants of the pairs with the highest and lowest difference in order to examine the theory.

Outlook: Control through understanding

Experimental uncertainties in biochemical measurements can have a great influence on the interpretation of data and thus the number of optimisation cycles in drug design. At the same time, the origin of the experimental uncertainty is relatively poorly understood from a scientific point of view. Important steps in assessing the source of the observed variations include a deeper understanding of the dilution-series errors and the variability of the biological material, and a routine inspection of the chemical purity of the measured substances.

Existing uncertainties can be estimated from independently repeated measurements. In order to understand

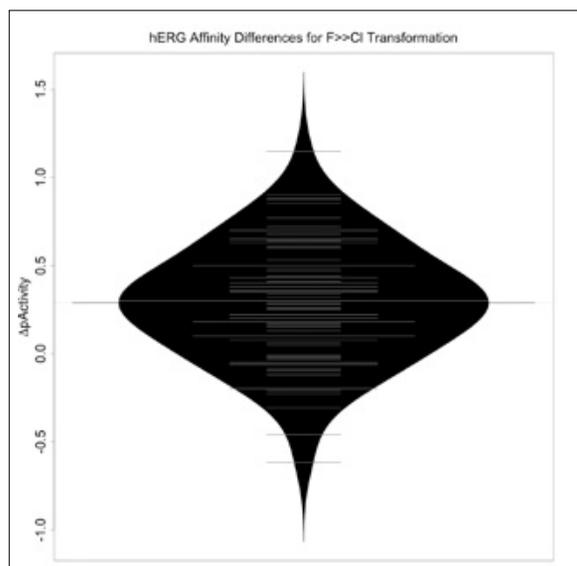


Fig. 2 Distribution of hERG binding affinity differences for all F>>Cl transformations from molecule pairs measured in the same laboratory and assay. The standard deviation of the distribution is 0.33 log units; the average increase of hERG affinity is 0.29 log units.

experimental uncertainty and to be able to trace differences in activity back to specific protein-ligand interactions, it is important that multiple measurements are carried out in a way that is completely independent. With better data from systematically repeated independent measurements, the error models can be refined in a subsequent step: for example, it is likely that the experimental uncertainty depends on the measurement range (very low and very high activity is measured more poorly than average activity) and on substance properties such as solubility and lipophilicity.

A further fundamental improvement to the understanding of experimental inaccuracy and test results can also be achieved by consulting statistical experts in the development of new assays. This is already the case in some pharmaceutical companies.

■ christian.kramer@uibk.ac.at

Literature

- [1] Aad, G. et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* 716, 1–29 (2012).
- [2] Intergovernmental Panel on Climate Change & Intergovernmental Panel on Climate Change. *Climate change 2007: the physical science basis: contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* (Cambridge University Press, 2007).
- [3] Gaulton, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107 (2011).
- [4] Kramer, C., Kalliokoski, T., Gedeck, P. & Vulpetti, A. The Experimental Uncertainty of Heterogeneous Public Ki Data. *J. Med. Chem.* 55, 5165–5173 (2012).
- [5] Kalliokoski, T., Kramer, C., Vulpetti, A. & Gedeck, P. Comparability of Mixed IC50 Data – A Statistical Analysis. *Plos One* 8, e61007 (2013).
- [6] Ekins, S., Olechno, J. & Williams, A. J. Dispensing Processes Impact Apparent Biological Activity as Determined by Computational and Statistical Analyses. *Plos One* 8, e62325 (2013).
- [7] Kramer, C.; Fuchs, J.; Gedeck, P.; Liedl, K. Matched Molecular Pair Analysis: Significance and the Impact of Experimental Uncertainty. Submitted

Photo: © istockphoto.com/ TimArbaev, Iuskiv