

Wissen statt Nichtwissen

Effizienteres Medikamentendesign durch
das Verständnis experimenteller Unsicherheit

Biologie ist naturgemäß komplex und selbst die Ergebnisse einfachster biochemischer Experimente sind mit nicht zu vernachlässigendem experimentellen Rauschen behaftet. Biochemische Messungen sind jedoch das Rückgrat moderner Pharmaforschung. Wird die experimentelle Unsicherheit bei der Planung neuer Wirkstoffkandidaten unterschätzt, lassen sich biochemische Daten sehr leicht überinterpretieren. Eine angemessene Berücksichtigung experimenteller Unsicherheit lässt sich mit wenig Zusatzaufwand erreichen und hilft, Wissen von Nichtwissen zu unterscheiden und zeit- und kostenintensive Irrwege zu vermeiden.

Experimentelle Unsicherheit in den Wissenschaften

In der Physik kann die Wahrscheinlichkeit neuer Erkenntnisse exakt quantifiziert werden: So erschien am 31.7.2012 auf dem Preprint Server *arXiv.org* der Beweis für die Existenz des letzten unbekanntes Elementarteilchens, des lange gesuchten Higgs Bosons – mit einer Signifikanz von 5,9 Standardabweichungen [1]. Auch in den Klimawissenschaften ist Unsicherheit Teil aller Vorhersagen. So hat die Arbeitsgruppe I des Intergovernmental Panel on Climate Change (IPCC; zwischenstaatlicher Ausschuss über Klimaveränderung) in ihrem vierten Bericht eine exakte Sprachregelung herausgegeben, mit der quantifizierte Unsicherheiten und Wahrscheinlichkeiten in Worte zu fassen sind [2]. Diese beiden Beispiele aus sehr unterschiedlichen Bereichen illustrieren, dass wissenschaftliche Erkenntnisse und Vorhersagen oft nicht in einem Ja-Nein-, Richtig-Falsch- oder Schwarz-Weiß-Schema dargestellt werden können, sondern mit einer gewissen Unsicherheit behaftet sind.

Auch in der Pharmaforschung unterliegen die allermeisten Studien- und Messergebnisse einer breiteren Verteilung: Die bekanntesten Fälle sind die Ergebnisse klinischer Studien, die danach bewertet werden, ob es den Probanden, die ein neues Medikament bekommen, signifikant besser geht als einer Vergleichsgruppe. Um diese Fragen, von deren Antwort Gesundheit und sehr viel Geld abhängen, sauber beantworten zu können, ist die Statistik gefragt, um echte Wirksamkeit von Scheinkorrelationen zu unterscheiden. Doch auch in den früheren nichtklinischen Phasen der medizinisch-chemischen Leitstrukturforschung und Leitstruktur-optimierung spielen Statistik und experimentelle Unsicherheit eine entscheidende Rolle.

Rationales computergestütztes Medikamentendesign

Klassisches Medikamentendesign läuft in iterativen Zyklen von chemischer Synthese und biologischer Testung,

die sich oft über Jahre hinziehen. In der frühen Phase des Drug Designprozesses ist es das Primärziel, eine Substanz mit hoher Affinität für das Zielprotein zu entwickeln. In den späteren Phasen kommen die Selektivität gegenüber anderen Proteinen, die toxische Nebenwirkungen vermitteln können, und die Aufnahme, Verteilung, Verstoffwechslung und Ausscheidung (engl. ADME: Absorption, Distribution, Metabolism, Excretion) des potenziellen Arzneistoffes als weitere Ziele hinzu.

Innerhalb eines Zyklus können die nächsten zu synthetisierenden Substanzen entweder nach einem trial-and-error-Prinzip oder aber rational, basierend auf allem chemischen und biochemischen Wissen, das man bereits über das Zielprotein und toxische Eigenschaften sowie ADME-Eigenschaften gesammelt hat, ausgewählt werden. Hier kommt das computergestützte Medikamentendesign zum Einsatz. Es hat die Aufgabe, alles bestehende Wissen zu bündeln und die vielversprechendsten chemischen Modifikationen herauszusuchen. Durch computergestütztes Design lässt sich die Anzahl der Designzyklen bedeutend verringern und viel Zeit und Geld sparen. Martin Karplus, Michael Levitt und Arieh Warshel haben für die Erforschung der Grundlagen, auf denen heutiges computergestütztes Medikamentendesign unter anderem aufbaut, 2013 den Nobelpreis für Chemie erhalten.

Bindekonstanten: das Kriterium Nummer eins im rationalen Drug Design

Die weitaus meisten zu optimierenden Bindeeigenschaften werden über Bindekonstanten an Proteine gemessen. Die Dissoziationskonstante K_d der Bindung an einen Rezeptor ist mit der Gibbs'schen freien Bindungsenergie ΔG^0 verknüpft

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

T steht für die Temperatur, R ist die allgemeine Gaskonstante. Häufig werden in biochemischen Assays nicht die K_d Werte bestimmt, sondern IC_{50} oder K_i Werte. IC_{50} Werte

beschreiben die Konzentration an Wirkstoffen, bei der die Funktion des Proteins um die Hälfte verringert ist. Aus IC_{50} Werten lassen sich mit einigen Nebenbedingungen K_i Werte, die Dissoziationskonstanten von Enzyminhibitoren, errechnen.

Bei Raumtemperatur entspricht die Änderung von K_d um eine Zehnerpotenz einem Unterschied in der Bindungsenergie von etwa 1,4 kcal/mol. Oft werden durch Modifikationen der chemischen Struktur nur kleine Verbesserungen in der Bindung weit unterhalb einer Größenordnung erzielt. Dies sind Grenzfälle, in denen Faustregeln zur Beurteilung der Signifikanz der beobachteten Unterschiede benutzt werden. Diese variieren je nach Vorgeschichte des Anwenders mitunter stark.

Da Biologie per se komplex ist und in biochemische Versuche sehr viele Faktoren mit hineinspielen, sind die gemessenen Bindekonstanten mit einiger experimenteller Unsicherheit versehen. Wenn die experimentelle Unsicherheit unterschätzt wird, besteht die Gefahr, kleine Unterschiede in Bindekonstanten zu überinterpretieren und Struktur-Aktivitäts-Beziehungen abzuleiten, wo keine existieren. Wird die experimentelle Unsicherheit überschätzt, werden Signale, die in den Daten vorhanden sind, nicht optimal genutzt. Beide Situationen, das Über- und Unterschätzen der experimentellen Ungenauigkeit, kosten Geld und Zeit. Sie verlangsamen das Medikamentendesign, weil sich das Projektteam bildlich gesprochen in Kämpfen mit Windmühlen verwickelt.

Wie groß ist die experimentelle Ungenauigkeit in Bindekonstanten?

In der wissenschaftlichen Literatur angegebene Bindekonstanten schwanken beträchtlich, unterschätzen aber meist die tatsächliche Variation der Messwerte. Einen Eindruck über publizierte Ungenauigkeiten kann man aus dem CSAR NRC-HiQ Datensatz (www.csardock.org) gewinnen. Hier wurden publizierte biochemische Affinitäten inklusive Standardabweichung zu 157 chemisch- und biologisch diversen Proteinligand-Systemen gesammelt. Der Median der publizierten Standardabweichungen ist $0,044 \log K_{d/i}$, mit den kleinsten Werten von 0,001 und $0,002 \log K_{d/i}$. Für jeden Wissenschaftler, der selbst schon einmal K_i Werte aus der Literatur zu reproduzieren versucht hat, ist klar, dass dies viel zu geringe experimentelle Unsicherheiten sind.

Ein realistischerer Eindruck über die experimentelle Unsicherheit lässt sich gewinnen, wenn K_i Werte, die von verschiedenen Wissenschaftlern in verschiedenen Labors, aber an denselben Protein-Ligand-System gemessen wurden, miteinander verglichen werden. Abbildung 1 zeigt einen solchen Vergleich für alle unabhängig voneinander gemessenen K_i Werte aus der ChEMBL Datenbank [3].

Wenn man eine einfache Normalverteilung für den experimentellen Fehler annimmt, lässt sich aus diesem Vergleich eine experimentelle Ungenauigkeit für heterogene K_i Werte von $0,54 \log K_i$ Einheiten errechnen [4]. Das bedeutet, dass sich zwei unabhängige K_i Messungen für dasselbe Protein-Ligand-System mit etwa 68 % Wahrscheinlichkeit innerhalb eines Abstandes von $\pm 0,54 \log K_i$ Einheiten finden. K_i Werte müssen vergleichbar sein, weil sie physikalische Bindekonstanten sind. Für die häufiger gemessenen IC_{50} Werte beträgt die Standardabweichung der experimentellen Variation $0,69 \log$ Einheiten [5]. Obwohl IC_{50} Werte aus verschiedenen Versuchsanordnungen und für verschiedene Protein-Ligand-Systeme nicht vergleichbar sein müssen, werden diese in der Praxis häufig miteinander verglichen, z.B. bei Selektivitätsbetrachtungen. Für chemische Standards, die häufig im selben Versuch bei Novartis/Basel gemessen wurden, haben wir je nach System und Versuchsaufbau eine experimentelle Unsicherheit mit einer Standardabweichung von 0,18 bis $0,35 \log$ Einheiten errechnet [5]. Das entspricht einem Faktor von 1,5 bis 2,2.

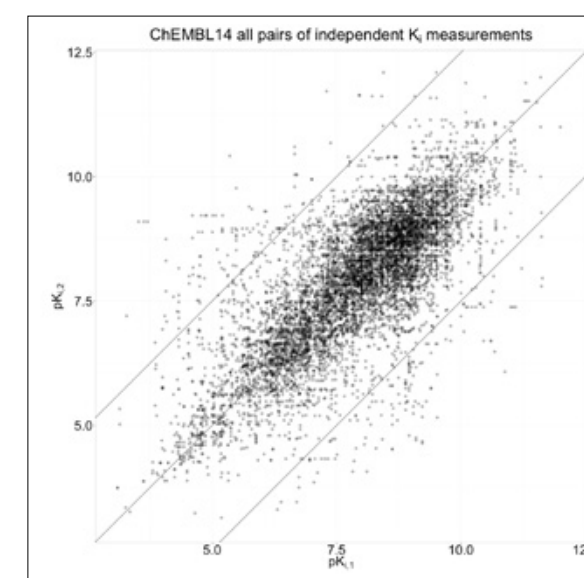


Abb. 1 Paare unabhängig gemessener pK_i Werte am selben Protein-Ligand-System aus ChEMBL14. Insgesamt finden sich in ChEMBL14 8524 Paare für 2046 Proteinligand-Systeme. Die Diagonalen kennzeichnen die Linie der identischen Messungen und die Grenzen, ab der die Unterschiede mehr als $2,5 \log$ Einheiten ausmachen [4]. Die Korrelation für alle Paare mit weniger als $2,5 \log$ Einheiten Unterschied beträgt $R^2 = 0,66$.



Christian Kramer, Jg. 1980, studierte Molecular Sciences in Erlangen und Zürich. Er promovierte von 2007–2009 an der Universität Erlangen in enger Zusammenarbeit mit Boehringer-Ingelheim/Biberach über neue QSAR- und QSPR-Methoden zu statistischen Vorhersagen von physikochemischen und biochemischen Eigenschaften. Von 2010 bis 2013 war er Presidential PostDoc bei Novartis in Basel mit Peter Gedeck in der Computer-Aided Drug Design-Gruppe. Dort beschäftigte er sich mit statistischen Ansätzen im Docking/Scoring, der Integrität von biochemischen Daten und Multipolkräftfeldern. Seit 2013 ist er Assistant Professor an der Universität Innsbruck. Seine Forschungsinteressen gelten der computergestützten Transformation chemischer und biochemischer Daten in wissenschaftliche Erkenntnisse für das molekulare Design. Schwerpunkte seiner Arbeit sind die Vernetzung von quantenmechanischen Methoden und Kraftfeldmethoden mit statistischer Auswertung von Datenbanken in der Leitstrukturfundung und -optimierung. Er ist Autor von 24 wissenschaftlichen Publikationen und Träger mehrerer Auszeichnungen, unter anderem eines Carl-Duisberg Scholarships, eines Novartis Presidential PostDoc Fellowships und zweier interner Novartis Auszeichnungen.

Die Gründe für die vergleichsweise hohe experimentelle Unsicherheit sind wissenschaftlich eher schlecht verstanden. Die Fehler in den Messgeräten, die das biologische Signal auswerten, scheinen das geringste Problem zu sein, wie sich aus den berichteten Unsicherheiten in der

Literatur ablesen lässt. Mögliche andere Gründe für die hohe Unsicherheit sind die Qualität und Stabilität des biologischen Materials, die Reinheit der vermessenen chemischen Substanzen, die Aggregation der Wirkstoffe und Variationen in Temperatur, Luftfeuchte und Druck. Eine weitere, nicht zu unterschätzende Fehlerquelle sind Fehler in den Verdünnungsreihen. Manche schlecht lösliche Substanzen bleiben während des Verdünnungsprozesses an den Pipettenwänden kleben, wodurch sich bei höheren Verdünnungen um Größenordnungen zu geringe Konzentrationen ergeben. Ekins et al. haben kürzlich gezeigt, dass strukturelle Interpretationen, die auf solchen Daten basieren, komplett falsch sein können [6].

Die Grenzwerte straffen: Wie experimentelle Unsicherheit Modellbildung beeinflusst

An zwei Beispielen sei hier gezeigt, wie die experimentelle Unsicherheit in der Modellbildung angemessen berücksichtigt werden kann.

Eine Standardanwendung im computergestützten Medikamentendesign sind QSAR- und Docking-Modelle. Hier werden verschiedene, strukturelle chemische und biochemische Eigenschaften mit der gemessenen Aktivität korreliert. Die Qualität solcher Modelle wird häufig mit dem R^2 quantifiziert, dem Bruchteil der erklärten Varianz der gemessenen Daten. Wenn nun ein Teil der gemessenen Varianz aus experimenteller Unsicherheit (Rauschen) besteht, kann der maximal erklärbare Teil der Varianz R^2_{max} berechnet werden nach

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

wobei hier σ_{noise} die Standardabweichung der experimentellen Unsicherheit ist und σ_{tot} die Standardabweichung der gesamten gemessenen Daten. Die Gleichung ergibt, dass R^2 als Maß nur dann interpretiert werden kann, wenn die experimentelle Unsicherheit der Daten bekannt ist. Je nach Verhältnis zwischen Signal und Rauschen kann R^2_{max} sehr klein werden.

Ein zweites Beispiel für die Bedeutung der experimentellen Unsicherheit ist die molekulare Paaranalyse, die in zunehmendem Maße in der Leitstrukturoptimierung angewandt wird. Hier werden Aktivitätsunterschiede zwischen zwei Molekülen mit dem chemischen Unterschied verglichen. Dafür wird aus Datenbanken ein großer Satz Bindungsdaten von Molekülpaaren, die sich alle durch denselben Austausch einer funktionellen Gruppe

unterscheiden, zusammengesucht. Aus der Verteilung der bekannten Aktivitätsunterschiede lässt sich eine Vorhersage über die Auswirkung desselben Austausches bei einem neuen Molekül machen. In Abbildung 2 ist beispielhaft die Verteilung der Affinitätsdifferenzen zum hERG-Kanal für alle in ChEMBL14 vorhandenen Paare gezeigt, bei denen ein Fluor in ein Chlor umgewandelt wurde.

Entscheidend für die Genauigkeit ist hierbei die Standardabweichung der bekannten Aktivitätsunterschiede. Je kleiner die Standardabweichung, desto genauer die Vorhersage. Durch die omnipräsente experimentelle Unsicherheit kann die Standardabweichung aber niemals null werden. Die minimale Standardabweichung für die Paare $\sigma_{pairs,min}$, die sich alleine aus der experimentellen Unsicherheit heraus erwarten lässt, lässt sich nach

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

berechnen. [7] Wenn man für hERG-Messungen aus demselben Labor eine experimentelle Unsicherheit $\sigma_{noise} = 0,2 \log$ Einheiten annimmt, ergibt sich für die Paare eine minimale Standardabweichung $\sigma_{pairs,min}$ von $0,28 \log$ Einheiten – sehr nahe an der beobachteten Standardabweichung der hERG-Affinitätsdifferenzen von $0,33$ für die F>>Cl Transformation. Die beobachtete Standardabweichung kann also fast vollständig mit der experimentellen Ungenauigkeit erklärt werden und es gibt im Gegensatz zu anderen Transformationen mit höherer Standardabweichung keinen Hinweis aus der Datenbank, dass hier spezifische Umgebungseffekte eine Rolle spielen. Ein spannender nächster Schritt besteht nun darin, die Bindungskonstanten der Paare mit der höchsten und der niedrigsten Differenz nachzumessen, um die Theorie zu überprüfen.

Ausblick: Kontrolle durch Verständnis

Experimentelle Unsicherheiten in biochemischen Messungen können einen großen Einfluss auf die Interpretation der Daten und damit der Zahl von Zyklen im Medikamentendesign haben. Gleichzeitig ist der Ursprung der Unsicherheit wissenschaftlich vergleichsweise schlecht verstanden. Wichtige Schritte zur inhaltlichen Beurteilung der beobachteten Variationen gehen über ein tieferes Verständnis der Verdünnungsreihenfehler und der Variabilität des biologischen Materials und eine routinemäßige Prüfung der chemischen Reinheit der gemessenen Substanzen.

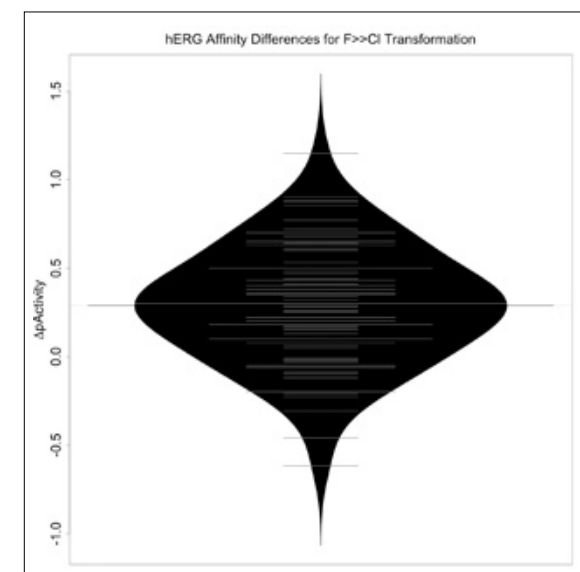


Abb. 2 Verteilung der hERG-Bindungsaffinitätsdifferenzen für alle F>>Cl Transformationen von im selben Labor und Assay gemessenen Molekülpaaren. Die Standardabweichung der Verteilung beträgt $0,33 \log$ Einheiten, die mittlere Zunahme der hERG-Affinität ist $0,29 \log$ Einheiten.

Existierenden Unsicherheiten lassen sich aus Mehrfachmessungen abschätzen. Um experimentelle Unsicherheit zu verstehen und Aktivitätsunterschiede auf spezifische Protein-Ligand-Wechselwirkungen zurückführen zu können, ist es wichtig, dass Mehrfachmessungen vollständig unabhängig voneinander durchgeführt werden. Mit besseren Daten aus systematischen Mehrfachmessungen lassen sich in einem nächsten Schritt die Fehlermodelle verfeinern: Es ist zum Beispiel nicht unwahrscheinlich, dass die experimentelle Unsicherheit vom Messbereich (sehr niedrige und sehr hohe Aktivität wird schlechter gemessen als mittlere Aktivität) und von Substanzeigenschaften wie Löslichkeit und Lipophilie abhängen.

Eine weitere grundlegende Verbesserung im Verständnis der experimentellen Ungenauigkeit und der Versuchsergebnisse ließe sich auch dadurch erreichen, dass Statistikexperten bei der Entwicklung von neuen Versuchen zurate gezogen werden. In einigen Pharmafirmen ist dies bereits der Fall.

■ christian.kramer@uibk.ac.at

Literatur
 [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