

Toxizitätsmodellierung mit kommerziellen (Q)SAR-Systemen: praktische Aspekte für die regulatorische Anwendung

Brigitte Simon-Hettich (brigitte.simon-hettich@merck.de), Darmstadt

In der neuen europäischen Chemikaliengesetzgebung REACH sind für Chemikalien mit geringen Produktionsvolumina vorwiegend tierversuchsfreie Alternativen zur Erfassung toxischer Wirkungen vorgesehen. In Anhang IX werden (Q)SAR-Modelle zur Vorhersage bestimmter toxischer Eigenschaften genannt (EU, 2004). Solche Vorhersagen bedürften nur dann keiner Bestätigung durch eine experimentelle Prüfung, wenn die Vorhersagen von validierten (Q)SAR-Systemen stammen, die Ergebnisse geeignet für Einstufung und Kennzeichnungszwecke sind und eine adäquate sowie verlässliche Dokumentation der Ergebnisse vorgelegt wird. Daraus ergibt sich die Frage, in wie weit die derzeit verfügbaren (Q)SAR-Systeme zur Vorhersage von toxischen Wirkungen unter diesen Voraussetzungen zur Beurteilung von Chemikalien unter REACH Verwendung finden können. Eine breite Akzeptanz von (Q)SARs würde in einer Reduktion von Tierversuchen sowie in Einsparungen von Zeit und Kosten resultieren (Tong et al, 2005).

Zur Erhöhung der Akzeptanz von (Q)SAR-Systemen im regulatorischen Bereich hat die OECD international harmonisierte Prinzipien für die Validierung von (Q)SARs entwickelt (OECD, 2004). Demnach sollte ein (Q)SAR-Modell sich auf einen definierten Endpunkt beziehen, einen eindeutigen Algorithmus und eine definierte Anwendungsdomäne beinhalten, sowie geeignete Maßgaben für statistische Verlässlichkeit, Robustheit und Vorhersagegenauigkeit und, wenn möglich, auch eine mechanistische Interpretation umfassen.

Derzeit sind einige kommerzielle (Q)SAR-Systeme zur Vorhersage toxikologischer Eigenschaften verfügbar. Es stellt sich daher die Frage, in wie weit diese Systeme im regulatorischen Umfeld verwendet werden können.

Verfügbare Modelle zur Toxizitätsmodellierung

Prädiktive computergestützte (Q)SAR-Systeme zur Vorhersage der Toxizität basieren auf der Analyse der Beziehung zwischen der chemischen (Sub-)Struktur einer Substanz oder Substanzgruppe und der toxikologischen Wirkung der Stoffe, die diese (Sub-)Struktur enthalten. Seit den 1980er Jahren werden (Q)SAR-Systeme für toxische Endpunkte entwickelt. Je nach dem, wie die Struktur-Wirkungsbeziehung entwickelt wird, werden zwei verschiedene Arten von Systemen unterschieden: sogenannte **Experten-systeme** und **statistisch-korrelative Systeme**.

Bei der Entwicklung von Expertensystemen erstellen Experten generalisierte Beziehungen zwischen einer (Sub-)Struktur und deren biologischer Wirkung und übertragen diese in eine computerlesbare Form. Die dazugehörigen

Informationen zur Toxizität werden in einer Datenbank gespeichert, die mit den Strukturinformationen verbunden ist. Die Vorhersage neuer chemischer Substanzen erfolgt über die qualitative Identifizierung jener Substrukturen, die mit einer Toxizität verbunden sind (Alerts oder Biophores). Expertensysteme sind daher klassische SAR-Systeme und generieren ja/nein-Antworten.

Statistisch-korrelative Systeme hingegen generieren quantitative Vorhersagen (QSAR). Dazu erfolgt die Analyse der chemischen Struktur anhand von Deskriptoren, d.h. numerischen Repräsentationen der chemischen Struktur. Deskriptoren können von atomaren oder molekularen Eigenschaften der Substanzen abgeleitet werden, wie

- physikalisch-chemische Eigenschaften (z.B. Verteilungskoeffizient Oktanol/Wasser),
- topologische Eigenschaften (z.B. Elektronegativitäten) oder
- Oberflächeneigenschaften (z.B. Polarität).

Diese Deskriptoren werden mit der Toxizität der Substanz korreliert und in einen mathematischen Algorithmus umgewandelt. Dieser Algorithmus wird zur Berechnung der quantitativen Effekte neuer Moleküle verwendet.

Außerdem können die (Q)SAR-Systeme hinsichtlich der Anzahl der vorhergesagten Endpunkte in ‚spezielle‘ und ‚universelle‘ Systeme unterteilt werden. ‚Spezielle Systeme‘ sind solche, die nur einen toxikologischen Endpunkt vorhersagen, z.B. die Expertensysteme OncoLogic der US Umweltbehörde EPA (Kanzerogenität) und Desicion Support System (DSS) des Bundesinstitutes für Risikobewertung (Haut- und Schleimhautreizung). ‚Universelle Systeme‘ dagegen umfassen eine Vielzahl von toxikologischen Endpunkten von akut bis chronisch, von lokal bis systemisch. Die bekanntesten Systeme sind die Expertensysteme DEREK und HazardExpert und die (Q)SAR-Systeme TOPKAT und MultiCASE. Weitere Informationen zu den Systemen (Endpunkte, Systemvoraussetzungen, Kosten) sind im ECETOC Technical Report on (Q)SARs enthalten (ECETOC, 2003).

Verwendung von (Q)SAR-Systemen bei der regulatorischen Stoffbewertung

Die derzeitige regulatorische Verwendung von (Q)SARs unterscheidet sich beträchtlich zwischen verschiedenen Ländern und sogar zwischen verschiedenen Behörden innerhalb eines Landes. Das resultiert zum Teil aus den unterschiedlichen regulatorischen Rahmenbedingungen, die verschiedene Anforderungen bedingen (Walker et al, 2002; Cronin et al, 2003). Eine weite Verbreitung hat die Verwendung von Analogiebetrachtungen („read across“) bei der Bewertung von Chemikalien gefunden. So wird z. B. in

den USA und China bei der Anmeldung neuer Chemikalien Analogiebetrachtungen zu strukturell ähnlichen Substanzen im jeweiligen nationalen Chemikalienrecht verankert und können zu einem eingeschränkten (öko-)toxikologischen Prüfprogramm und somit zur Reduktion von Tierversuchen führen. In der EU und Südkorea werden Analogiebetrachtungen nur in Einzelfällen bei der Notifizierung von Neustoffen akzeptiert (Walker et al, 2002).

In der EU sind außerdem Gruppenbetrachtungen von homologen Substanzen bei der Einstufung und Kennzeichnung von Stoffen gemäß Gefahrstoffverordnung in der Vergangenheit in Einzelfällen erfolgreich angewandt worden, z.B. bei anorganischen Schwermetallsalzen. Im Gegensatz dazu wird im Technical Guidance Document on Risk Assessment der EU (ECB, 2003) für toxikologische Endpunkte der Einsatz von Struktur-Wirkungsbeziehungen bei der Risikobewertung im wesentlichen auf „expert judgement“ beschränkt, welches dann als „zusätzlicher Hinweis“ gewertet wird.

Deutlich weiter gehen die amerikanische Food and Drug Administration (US FDA) und die dänische Umweltbehörde EPA bei dem Einsatz von (Q)SAR-Systemen. Das Center for Drug Evaluation and Research (CDER) und die Informatics and Computational Safety Analysis Staff (ICSAS) der US FDA ermitteln mit diesen Systemen Vorhersagen zu toxischen und adversen klinischen Effekten bei der Beurteilung von Inhaltsstoffen und Verunreinigungen von Arznei- und Nahrungsmittelzusatzstoffen (Matthews et al, 2004a, Matthews et al, 2004b). Darüber hinaus hat die FDA anhand von vertraulichen und öffentlichen Daten eigene (Q)SAR-Modelle für die kommerziellen Systeme MultiCASE und MDL entwickelt. Bisher stehen bereits Modelle für folgende Endpunkte zur Verfügung,

- Gentoxizität (Ames-, Mouse-Lymphoma-, und in vitro-Chromosomenaberrationstest)
- Kanzerogenität (Ratte und Maus)
- Reproduktionstoxizität (Entwicklungstoxizität)
- Maximal tolerierte Dosis
- Hepatotoxizität.

Diese Modelle werden von der Behörde bei der Bewertung von neuen Stoffen verwendet und sind bei den jeweiligen Anbietern erhältlich. Ziel der FDA ist die Entwicklung einer kompletten Batterie von Vorhersageprogrammen für alle wesentlichen toxikologischen Studien, die von der FDA im Rahmen von Zulassungsverfahren für Arzneimittel oder Nahrungsmittelzusatzstoffen empfohlen werden (Contrera et al, 2004; Matthews et al, 2006).

Zur Abschätzung der Sicherheit von Substanzen, die mit Lebensmitteln in Kontakt kommen erhebt das SAR-Team des Office of Food Additive Safety (OFAS) der US FDA im Rahmen der Food Contact Notification (FCN) Daten zur Kanzerogenität mit kommerziellen und selbstentwickelten Systemen und Datenbanken, wie z.B. OncoLogic, MultiCASE, TOXSYS, Leadscope (Bailey et al., 2005).

Die dänische Umweltbehörde hat mit den (Q)SAR-Systemen MultiCASE und TOPKAT Vorhersagen für ca. 20.000 Chemikalien aus dem EINECS-Verzeichnis ermittelt, daraus Einstufungs- und Kennzeichnungsvorschläge nach dem europäischen Gefahrstoffrecht abgeleitet und im Internet veröffentlicht (Tyle et al, 2002).

In Deutschland hat das Bundesinstitut für Risikobewertung (BfR) ein Decision Support System zur Vorhersage von lokalen Effekten auf Haut- und Schleimhaut (reizende/ätzende Wirkung) entwickelt. Die Vorhersagen von diesem System sollen für die Einstufung und Kennzeichnung der Substanzen nach den Kriterien der Gefahrstoffverordnung geeignet sein (Gerner & Schlede, 2002). Das europäische Chemikalienbüro hat dieses System analysiert, bewertet und Vorschläge zur Weiterentwicklung abgeleitet (Rorije & Hulzeboos, 2005; Tsakovska et al, 2005).

Vorhersagegenauigkeit der (Q)SAR-Systeme

Voraussetzung für einen breiteren Einsatz von (Q)SAR-Systeme für regulatorische Zwecke ist, dass die Vorhersagen valide und verlässlich sind. Die Vorhersagesysteme basieren auf Daten aus experimentellen Prüfungen. Somit ist die Modellentwicklung abhängig von der Verfügbarkeit einer ausreichend großen Datenbank mit qualitativ hochwertigen Daten, denn die Vorhersagesysteme können nur so gut sein, wie die zur Entwicklung verwendeten Daten. Weiterhin sollten die Datenbanken eine möglichst große Vielzahl an verschiedenen chemischen Strukturen und Funktionalitäten abdecken, um ‚universell‘ einsetzbare Systeme zu erhalten. (Sub-)Strukturen oder Moleküleigenschaften, die nicht ausreichend im sogenannten Training Set der Modelle vorhanden sind, können nicht verlässlich vorhergesagt werden.

(Q)SAR-Modelle können für jeden beliebigen toxikologischen Endpunkt entwickelt werden. Für toxische Wirkungen, die auf einen bekannten Mechanismus zurückgeführt werden können, sind die (Q)SAR-Modelle oft erfolversprechend. Je spezifischer und genauer die Mechanismen beschrieben werden können, die dem modellierten Effekt zugrunde liegen, desto genauer ist das resultierende (Q)SAR-Modell. Wenn der Effekt jedoch von einer Vielzahl unterschiedlicher Mechanismen ausgelöst werden kann, so kann die statistische Korrelation ohne Berücksichtigung der Mechanismen zu fehlerhaften Beziehungen und zu nicht verlässlichen (Q)SAR-Modellen führen (Simon-Hettich et al, 2006).

Zur Zeit liefern die wichtigsten kommerziellen Vorhersagesysteme (z.B. DEREK, MultiCASE, TOPKAT) recht gute Vorhersagen für den Endpunkt Mutagenität mit einer Vorhersagegenauigkeit von >70% (White et al, 2003; Hayashi et al, 2005; Hayashi et al, 2006). Auch für lokale Effekte (Haut- und Schleimhautreizung), die stark von physikalisch-chemischen Eigenschaften der Substanzen beeinflusst werden, ist eine erfolgreiche Modellierung möglich (Gerner & Schlede, 2002). Besonders schwierig ist die Modell-

entwicklung für komplexe Endpunkte wie akute Toxizität, chronische Toxizität, Organtoxizität, Kanzerogenität oder Reproduktionstoxizität, da die in diesen Studien erfassten Effekte von einer Vielzahl von Variablen und Mechanismen abhängen: Aufnahme, Verteilung, Metabolismus und Ausscheidung (ADME) der Substanz in vivo, spezies-spezifische Effekte, Reproduzierbarkeit und Variabilität der experimentellen Ergebnisse sowie nicht standardisierte Beurteilungskriterien für die beobachteten Effekte erschweren die Modellentwicklung (Gottmann et al., 2001). Für spezifische systemische Effekte wie die Peroxisomenproliferation in der Leber, deren Entstehungsmechanismus gut untersucht und verstanden ist, können jedoch erfolgreich (Q)SAR-Modelle entwickelt werden (Simon-Hettich et al., 2006). Darüber hat sich auch die Entwicklung von (Q)SARs für bestimmte Substanzklassen mit ähnlichen Eigenschaften (kongenere Substanzen) als vielversprechend erwiesen (Richard & Benigni, 2001).

Je nach beabsichtigter Verwendung der Vorhersagen kann die benötigte Vorhersagegenauigkeit unterschiedlich sein. Für das Screening einer Vielzahl von Substanzen z.B. zur Auswahl von Leitsubstanzen in der Arzneimittelforschung oder zur Priorisierung von Substanzen für eine toxikologische Evaluierung ist ein geringerer Grad an Vorhersagegenauigkeit akzeptabel. Sollten die Vorhersagen aber für Einstufungen und Kennzeichnungen im Sinne des Chemikaliengesetzes oder für eine umfassende Risikobeschreibung verwendet werden und somit mit weiteren regulatorischen Maßnahmen verbunden sein, ist ein hohes Maß an Vorhersagegenauigkeit erforderlich.

Zusammenfassung

Computergestützte Vorhersagesysteme stellen bereits heute eine sinnvolle Ergänzung des toxikologischen Repertoires für die Bewertung von Substanzen dar. Vor allem beim Screening und bei der Priorisierung von Substanzen können diese Systeme eine wichtige Hilfestellung leisten. Da die kommerziell verfügbaren Systeme derzeit noch nicht den OECD-Validitätskriterien entsprechen, ist eine weitergehende Verwendung im regulatorischen Bereich bisher noch eingeschränkt. Unerlässlich für die Bewertung der Vorhersageergebnisse ist eine ausreichende Expertise im Umgang mit diesen Systemen. Aufgrund der relativ hohen Kosten sind die kommerziellen Vorhersagesysteme jedoch nicht allen Interessengruppen (Wissenschaft, Behörden, Industrie) zugänglich (ECETOC, 2003).

Im Rahmen des neuen europäischen Chemikalienrechts REACH werden neben in vivo- und in vitro-Testsystemen (Q)SAR-Systeme (in silico) ein fester Bestandteil der toxikologischen Prüfstrategie werden. Eine Verwendung dieser Systeme in einer ‚integrated testing strategy‘ zur Identifizierung von möglichen toxischen Eigenschaften ist wissenschaftlich möglich und aus Kostengründen auch erforderlich. Mit Hilfe der in silico-Ergebnisse können

substanzspezifische toxikologische Prüfprogramme entwickelt werden, die zu enormen Einsparungen bei der Prüfung von Stoffen im Rahmen von REACH führen. Trotz allem müssen aber auch die Grenzen der Verwendung von (Q)SAR-Systemen bei der Vorhersage von systemisch toxischen Effekten oder bei vollkommen neuen chemischen Strukturen, die nicht im Training Set des Systems vorhanden sind, beachtet werden.

Literatur

- [1] Bailey AB, Chanderbhan R, Collazo-Braier N, Cheeseman MA, Twaroski ML (2005) The use of structure-activity relationship analysis in the food contact notification program. *Reg Toxicol Pharmacol* 42: 225-235
- [2] Contrera JF, Matthews EJ, Kruhlak NL, Benz RD (2004) Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. *Reg Toxicol Pharmacol* 40: 185-206
- [3] Cronin MTD, Jaworska JS, Walker JD, Comber MHI, Watts CD, Worth AP (2003) Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environ Health Perspect* 111(10): 1391-1401
- [4] ECB (2003) Technical guidance Document on Risk Assessment, 2nd ed. European Chemicals Bureau, Ispra (<http://ecb.jrc.it/tgdoc>)
- [5] ECETOC (2003) (Q)SARs: Evaluation of the commercially available software for human health endpoints with respect to chemical management applications. Technical Report No. 89. European Centre for Ecotoxicology and Toxicology of Chemicals. Brussels
- [6] EU (2004) <http://europa.eu.int/comm/enterprise/reach>
- [7] Gerner I, Schlede E (2002) Introduction of in vitro data into local irritation/corrosion testing strategies by means of SAR considerations: assessment of chemicals. *Toxicol Lett* 127: 169-175
- [8] Gottmann E, Kramer S, Pfahringer B, Helma C (2001) Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. *Environ Health Perspect* 109: 509-514
- [9] Hayashi M, Kamata E, Hirose A, Takahashi M, Morita T, Ema M (2005) In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals. *Mutat Res* 588: 129-135
- [10] Hayashi M, Kamata E, Hirose A, Takahashi M, Morita T, Ema M (2006) Erratum to "In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals" [*Mutat Res* 588: 129-135]. *Mutat Res* 603: 196
- [11] Matthews EJ, Kruhlak NL, benz RD, Contrera JF (2004a) Assessment of the health effects in humans: I. QSAR estimation of the maximum recommended

therapeutic dose (MRTD) and no effect level (NOEL) of organic chemicals based on clinical trial data. *Curr Drug Discov Technol* 1(1): 61-76

- [12] Matthews EJ, Kruhlak NL, Weaver JL, Benz RD, Contrera JF (2004b) Assessment of the health effects in humans: II. Construction of an adverse effects database for QSAR modeling. *Curr Drug Discov Technol* 1(4): 243-254
- [13] Matthews EJ, Kruhlak NL, Cimino MC, Benz RD, Contrera JF (2006) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. *Reg Toxicol Pharmacol* 44: 97-100
- [14] OECD (2004) The report for the expert group on (quantitative) structure-activity relationships [(Q)SARs] on the principles for the validation of (Q)SARs. OECD Series on Testing and Assessment No. 49. Organisation for the Economic Co-operation and Development. Paris. ENV/JM/MONO(2004)24
- [15] Richard AM, Benigni A (2001) AI and SAR approaches for predicting chemical carcinogenicity: survey and status report. *SAR QSAR Environ Res* 13: 1-19
- [16] Rorije E, Hulzeboos E (2005) Evaluation of (Q)SARs for the Prediction of Skin Irritation/Corrosion Potential. National Institute of Public Health and Environment (RIVM), Bilthoven, NL.
- [17] Simon-Hettich B, Rothfuss A, Steger-Hartmann T (2006) Use of computer-assisted prediction of toxic effects of chemical substances. *Toxicology* 224: 156-162
- [18] Tong W, Hong H, Xie Q, Shi L, Fang H, Perkins R (2005) Assessing QSAR limitations – a regulatory perspective. *Current Computer-Aided Drug Design* 1(2): 195-205
- [19] Tsakovska I, Netzeva T, Worth A (2005) Evaluation of (Q)SARs for Eye Irritation/Corrosion Potential. European Chemicals Bureau, Ispa, I (EUR 21897 EN)
- [20] Tyle H, Larsen HS, Wedebye EB, Sijm D, Krog TP, Niemela J (2002) Identification of potential PBTs and vPvBs by use of QSARs. Danish EPA SHC/TS 2-3/029
- [21] Walker JD, Carlsen L, Hulzebos E, Simon-Hettich B (2002) Global government application of analogues, SARs and QSARs to predict aquatic toxicity, chemical or physical properties, environmental fate parameters and health effects of organic chemicals. *SAR QSAR Environ Res* 13(6): 607-616
- [22] White AC, Mueller RA, Gallavan RH, Aaron S, Wilson AG (2003) A multiple in silico program approach for the prediction of mutagenicity from chemical structure. *Mutat Res* 539: 77-89

Korrespondenzadresse:

Dr. Brigitte Simon-Hettich
Institut für Toxikologie, Merck KGaA
Frankfurter Straße 250
64293 Darmstadt