



Waldemar Klingspohn (Autor)
Optimistische Verzerrung in der Chemieinformatik



<https://cuvillier.de/de/shop/publications/8031>

Copyright:
Cuvillier Verlag, Inhaberin Annette Jentsch-Cuvillier, Nonnenstieg 8, 37075 Göttingen,
Germany
Telefon: +49 (0)551 54724-0, E-Mail: info@cuvillier.de, Website: <https://cuvillier.de>



Teil I

Einleitung





Einleitung

1.1 Der Beginn moderner Wirkstoffentwicklung

„Wenn wir also im Sinne der modernen Pharmakologie diese Aufgabe erweitern wollen, so heißt das nichts anderes, als daß wir „zielen“ lernen, und zwar „durch chemische Variationen“ zielen lernen“^[1]!

Diese auf Paul Ehrlich (1854–1915) zurückgehende Aussage legte den Grundstein für die heutige medizinische Chemie und Chemotherapie. Ehrlich postulierte damit den Grundgedanken, dass chemische Substanzen durch diverse strukturelle Variationen in einer Weise so abgeändert werden können, dass sie selektiv an ihre Zielstrukturen binden und einen pharmakologischen Effekt im Organismus ausüben. Solche Substanzen wurden von Ehrlich „Zauberkugeln“^[2] genannt und waren Grundlage für seine Theorie, die u. a. als Vorlage für die heutige Wirkstoffentwicklung und für das Prinzip der quantitativen Struktur-Aktivitäts-Beziehung (Abk. QSAR, engl. *quantitative structure-activity relationship*) verstanden werden kann.^[3–5]

Die Suche nach einer solchen Zauberkugel begann für Ehrlich im Jahre 1891 mit der Untersuchung des Methylenblaus als Wirkstoff gegen die Malaria, welches sich jedoch gegen den damaligen Standard Chinin nicht behaupten konnte. Er untersuchte daraufhin Trypanrot als mögliches Therapeutikum gegen den Erreger *Trypanosoma equinum*, welcher jedoch Resistenzen gegen das Trypanrot entwickelte.

Im Folgenden richtete Ehrlich seine Aufmerksamkeit auf organische Arsenverbindungen, insbesondere auf die von Pierre Jaques Antoine Béchamp (1816–1914) im Jahre 1863 synthetisierte Arsanilsäure (Atoxy¹®, Abb. 1.1a), die 1905 Einsatz bei der Afrikanischen Trypanosomiasis (Schlafkrankheit) fand. Jedoch zeigten sich bei langfristiger Therapie toxische Effekte der Arsanilsäure. Mit dem Be-

streben, eine erfolgreiche Zauberkegel zu finden, führte Ehrlich zusammen mit seinem Kollegen und Chemiker Alfred Bertheim (1879–1914) diverse strukturelle Modifikationen an der Arsanilsäure durch mit dem Ziel vor Augen, die therapeutische Wirksamkeit der Arsanilsäure einerseits zu erhöhen und andererseits ihre toxische Wirkung zu erniedrigen: „*Es hat sich hierbei gezeigt, daß je nach den verschiedenen Eingriffen und Umformungen der Arsanilsäure die Verbindung nach Belieben entgiftet oder giftiger gemacht werden kann*“^[1].

Diese Beobachtung, dass strukturelle Modifikationen eines Wirkstoffs einen Einfluss auf dessen pharmakologische Wirksamkeit haben, zeigte erste Anzeichen des Verständnisses über die heutige Allgemeingültigkeit der Struktur-Aktivitäts-Beziehung (Abk. SAR, engl. *structure-activity relationship*) in der medizinischen Chemie.

In ihrer Serie von Modifikationen der Arsanilsäure stießen Ehrlich und Bertheim schließlich auf die 306. Modifikation (Compound No. 306), namentlich Arsacetin (Abb. 1.1b). Doch auch bei dieser Modifikation traten toxische Effekte auf. Erst die 418. Modifikation, das Arsenophenylglycin (Spirarsyl[®], Abb. 1.1c), versprach den erhofften Erfolg.

1907 wurde das Arsenophenylglycin am Menschen getestet und bewies sich als sicher und effektiv gegen Trypanosomen. Allerdings zeigte sich bei einem kleinen Patientenkollektiv eine starke Hypersensitivität. Nichtsdestotrotz zeigten sich erste Indizien zur Bestätigung seiner Theorie. Nach vielen weiteren Modifikationen wurde 1907 letztlich die 606. Substanz (Compound No. 606) als potenzieller Wirkstoff gegen Trypanosomen synthetisiert, das Arsphenamin (Salvarsan[®], Abb. 1.1d), dessen Wirksamkeit gegen Trypanosomen jedoch nicht belegt werden konnte und zunächst in Vergessenheit geriet. Erst Jahre später sollte Arsphenamin wiederentdeckt werden und Ehrlichs Theorie bestätigen. Denn schon im Jahre 1905 entdeckten Fritz Schaudinn (1871–1906) und Erich Hoffmann (1868–1959) die Spirochäten der Art *Treponema pallidum* als den Erreger der Syphilis.^[6,7] Von Hoffmann auf die Verwandtschaft zwischen *Treponema pallidum* und Trypanosomen hingewiesen, testete Ehrlich seine Arsenverbindungen als mögliche Wirkstoffe gegen die Syphilis. Tatsächlich stellte sich das Arsenophenylglycin als pharmakologisch wirksam heraus.

Der Durchbruch bei der Behandlung der Syphilis gelang aber erst im Jahre 1909, als Ehrlichs Kollege Sahachiro Hata (1873–1938) alle bis dato von Bertheim synthetisierten Arsenverbindungen an mit Syphilis infizierten Kaninchen testen sollte und dabei auf Arsphenamin stieß, welches eine herausragende Wirksamkeit gegen die Syphilis zeigte. Im Jahre 1910 machte Paul Ehrlich die Entdeckung publik.^[8] Endlich war die Zauberkegel gefunden.

Fortan wurde Arsphenamin von der breiten Masse auch als „606“ oder „Ehrlich-Hata 606“ bezeichnet und stieß auf große Bewunderung, die sich in zahlreichen Schlagzeilen in diversen Zeitungen und Fachzeitschriften widerspiegelte. Ehrlichs Erfolg mündete sogar in einer Verfilmung aus dem Jahre 1940 mit dem Titel „Dr. Ehrlich's Magic Bullet“.^[9] Es ergaben sich bei der Therapie der Syphilis mit Arsphenamin jedoch einige problematische Aspekte. Zum einen war die Therapie langwierig, da mehrere Injektionen des Wirkstoffs für einen therapeutischen Erfolg erfolgen mussten. Zum anderen war eine Kombinationstherapie mit Quecksilber und Bismut für die vollständige Eindämmung des Erregers nötig. Ein weiteres Problem stellte die galenische Formulierung des Arsphenamins dar. Aufgrund der Unlöslichkeit in Wasser wurde Arsphenamin als Hydrochlorid eingesetzt und musste vor jeder Injektion mit einer basischen Lösung verdünnt werden. Ehrlich erkannte all diese Aspekte und suchte nach einer weiteren und zugleich verbesserten Modifikation. Zu Ehrlichs 60. Geburtstag im Jahre 1914 wurde die 914. Modifikation entdeckt (Compound No. 914), das Neosarsphenamin (Neosalvarsan[®], Abb. 1.1e). Mit dem wasserlöslichen und weniger oxidationsempfindlichen Neosarsphenamin konnte letztlich die Therapiedauer verkürzt und sein Vorgänger ersetzt werden.

1915 verstarb Paul Ehrlich, aber auch nach seinem Tod gingen die Forschungen an Arsenverbindungen weiter und so wurde das Oxidationsprodukt des Arsphenamins als der wirksame Metabolit erkannt und 1930 das Oxophenarsin (Mapharsen[®], Abb. 1.1f) entwickelt. Das Oxophenarsin behauptete sich im Hinblick auf seine Löslichkeit und Stabilität und galt bis zum Einsatz des Penicillins als das Standard-Therapeutikum bei der Syphilis.

Paradoxaerweise wurde Oxophenarsin bereits von Ehrlich und Berthelm als die 599. Modifikation synthetisiert, jedoch aufgrund seiner vermuteten Toxizität nicht weiter untersucht.^[10,11] 1949 wurde das Arsenderivat Melarsoprol (Arsobal[®], Abb. 1.1g) zur Behandlung der Schlafkrankheit vorgestellt^[12], welches bis heute von der WHO (World Health Organization) empfohlen wird.^[13] Die Entwicklung von der Arsanilsäure bis hin zum Arsphenamin^[4,5,11,14], welches bis dato das erste synthetisch hergestellte Antibiotikum war und Ehrlichs Zauberkugeln populär machte, zeigt in eindrucksvoller Weise die schon früh erkannte Abhängigkeit der biologischen Aktivität einer Substanz von dessen Struktur (SAR) und die damit einhergehende erste rationale Entwicklung eines synthetischen Wirkstoffs. Heutzutage gilt Paul Ehrlich als einer der Begründer der modernen Chemotherapie und Arzneistoffentwicklung. Ehrlichs Herangehensweise, die Synthese mit biologischen Testungen zu kombinieren, stellt heute das Grundprinzip für das Design und für die Evaluierung neuer Wirkstoffkandidaten in der medizinischen Chemie dar.^[3]

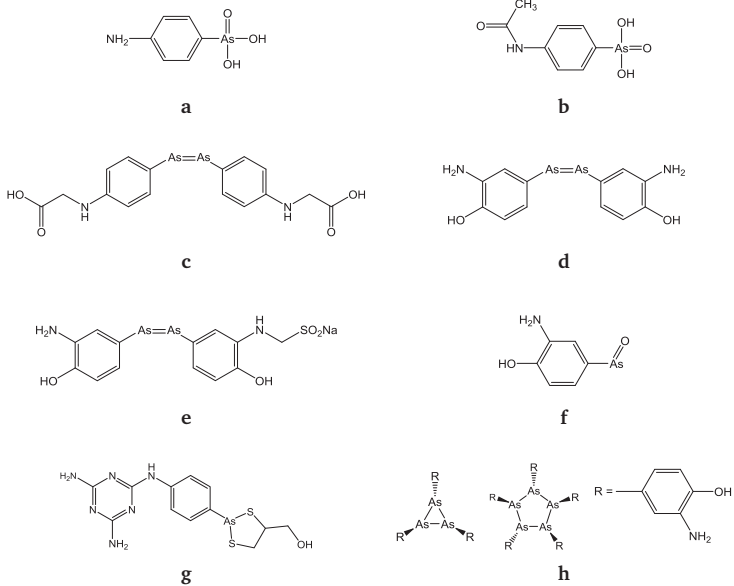


Abb. 1.1: Ehrlich Strukturen. **a** Atoxyl **b** Arsacetin **c** Arsenophenylglycin **d** Arspenamin **e** Nearsphenamin **f** Oxophenarsin **g** Melarsoprol **h** Revidierte Struktur von Arspenamin als Gemisch aus einem Drei- und Fünfring von einbindigen Arsenatomen.^[15]

1.2 Analyse der QSAR

Die QSAR beruht auf der Annahme, dass die strukturellen und physikochemischen Eigenschaften einer Substanz, z. B. eines Wirkstoffkandidaten, im funktionellen Zusammenhang mit dessen biologischer Aktivität stehen:

$$\text{biologische Aktivität} = f(\text{Eigenschaften}). \quad (1.1)$$

Das Prinzip der QSAR hat seinen Ursprung in der Toxikologie. Bereits 1848 beobachtete Blake^[16,17] eine potenzielle Beziehung zwischen der chemischen Konstitution anorganischer Salze und ihrer physiologischen Wirkung. Im Jahre 1863 erkannte Cros^[18] eine Beziehung zwischen der Toxizität primärer, aliphatischer Alkohole und ihrer Wasserlöslichkeit. Einen funktionellen Zusammenhang und die erste allgemeine Formulierung einer QSAR stellten erst Crum-Brown und Fraser^[19,20] im Jahre 1868/69 auf. Ihre Untersuchungen an Alkaloiden führte sie zu der Annahme, dass die physiologische Aktivität Φ eine Funktion der chemischen Struktur C ist:^[21]

$$\Phi = f(C). \quad (1.2)$$

1893 entdeckte Richet^[22], dass die Toxizität organischer Verbindungen in einem inversen Verhältnis zu ihrer Wasserlöslichkeit steht. Damit untermauerte Richet Crum-Browns und Frasers Annahme, erweiterte aber auch deren funktionellen Zusammenhang auf:^[21]

$$\Delta\Phi = f(\Delta C). \quad (1.3)$$

Hierbei steht Δ für eine Differenz bzw. Änderung. Das bedeutet, dass eine Änderung in der chemischen Struktur mit einer Änderung der biologischen Aktivität einhergeht. Die Wende zum 20. Jahrhundert brachte zwei unabhängig voneinander verlaufende Beobachtungen von Meyer^[23] und Overton^[24] hervor, nämlich einen linearen Zusammenhang zwischen der Lipophilie, ausgedrückt als Öl-Wasser-Verteilungskoeffizient, und narkotischer sowie toxischer Wirkung von Narkotika. Den Grundstein für die moderne QSAR-Analyse legten aber erst Hansch und Fujita^[25] im Jahre 1964. Sie machten den physiologischen Effekt einer Substanz von zwei Stufen abhängig.^[3,25-27] So sind der Transport (Pharmakokinetik) und die Bindung der Substanz an die Zielstruktur (Pharma-

kodynamik) die entscheidenden Schritte für einen physiologischen Effekt. Mit der sog. Hansch-Analyse wurde ein lineares mathematisches Modell aufgestellt, in das lipophile (π)^[28], elektrostatische (σ)^[29] und sterische (E_S)^[30] Parameter einfließen, um eine QSAR zu beschreiben:^[3]

$$\log \frac{1}{C} = k_1(\pi) + k_2(\sigma) + k_3(E_S) + k_4. \quad (1.4)$$

Hierbei steht C für die erforderliche Konzentration für einen bestimmten physiologischen Effekt, während k_1 , k_2 und k_3 den Einfluss des jeweiligen Parameters angeben (Regressionskoeffizienten) und k_4 eine Konstante ist. Dieses Modell beruht auf der Annahme, dass der physiologische Effekt hauptsächlich von den drei genannten Parametern, welche die physikochemischen und strukturellen Eigenschaften einer Substanz repräsentieren, beeinflusst wird.

So bestimmt die Lipophilie (π) die Aufnahmefähigkeit einer Substanz durch die Zellmembran in die Zelle. Die elektrostatischen Eigenschaften (σ) einer Substanz spielen eine wichtige Rolle bei der Bindung an die Zielstruktur. Größe und Konformation (E_S) einer Substanz bestimmen u. a. dessen Bindungsfähigkeit an die Zielstruktur.

Die Hansch-Analyse stellte damit einen mathematischen Zusammenhang zwischen der biologischen Aktivität einer Substanz und dessen Eigenschaften dar. Fortan stellte sie eine Methode dar, die biologische Aktivität von Substanzen vorherzusagen.

Eine QSAR-Analyse dient letztlich der Vorhersage von abhängigen Variablen experimentell ungetesteter Substanzen und umfasst im wesentlichen Sinne drei Komponenten:^[31–33]

1. Die zu modellierende abhängige Variable y .
2. Die in Zahlenwerte codierten physikochemischen Eigenschaften der Trainingssubstanzen, die sog. Moleküldeskriptoren (unabhängige Variablen), X (vgl. Kapitel 4).
3. Das mathematische Modell zum Aufstellen eines funktionellen Zusammenhangs, der Algorithmus (vgl. Kapitel 7).

Handelt es sich bei der zu modellierenden abhängigen Variable um eine kategorische, so bezeichnet man das aufgestellte Modell als Klassifikationsmodell (vgl. Kapitel 6). Beispiele für eine kategorische Variable sind die biologische Aktivität und die metabolische Stabilität einer Substanz, die kategorisiert werden können in aktiv/inaktiv bzw. stabil/instabil. Diese kategorischen Variablen

stellen keine Werte im eigentlichen Sinne dar und werden i. Allg. als diskrete, numerische Variablen codiert. So kann die biologische Aktivität als „0“ (aktiv) oder „1“ (inaktiv) wiedergegeben werden.

Soll dagegen eine experimentelle Messgröße oder ein physikochemischer Parameter einer Substanz modelliert werden, so bezeichnet man das aufgestellte Modell als Regressionsmodell. Hierbei handelt es sich bei der zu modellierenden abhängigen Variable um eine kontinuierliche und numerische Variable. Um die vorangehenden Beispiele aufzugreifen, ist die mittlere inhibitorische Konzentration (Abk. IC_{50} , engl. *half maximal inhibitory concentration*) eine Messgröße für die biologische Aktivität und die Bioverfügbarkeit (z. B. Konzentration) eine Messgröße für die metabolische Stabilität und das Permeationsvermögen einer Substanz.

Das Aufstellen eines funktionellen Zusammenhangs mithilfe eines mathematischen Modells wird auch als Training bezeichnet. Dieses Training impliziert, dass die abhängigen Variablen der betrachteten Substanzen (Trainingssubstanzen oder -objekte) bereits bekannt sind, d. h. experimentell ermittelt wurden. Da das Ziel einer QSAR-Analyse die Vorhersage der abhängigen Variablen von experimentell noch nicht untersuchten Substanzen (Testsubstanzen oder -objekte) ist, beruht die Vorhersage von Testsubstanzen allein auf deren molekularen Deskriptoren, die im Kapitel 4 näher erläutert werden.

Aufgrund ihres prädiktiven Charakters ist die QSAR-Analyse heutzutage von zentraler Bedeutung in der Arzneimittelforschung und -entwicklung, zumal die Entwicklung eines Arzneimittels einen langwierigen und vor allem kostenintensiven Prozess darstellt. So dauert es von der Suche nach einem potenziellen Arzneistoffkandidaten über dessen Synthese bis hin zur Zulassung des Arzneimittels über zehn Jahre. Laut einer aktuellen Studie von DiMasi^[34] aus dem Jahre 2016 belaufen sich die Kosten eines pharmazeutischen Unternehmens für ein zugelassenes Arzneimittel auf etwa 2.6 Milliarden US Dollar. Folglich ist ein pharmazeutisches Unternehmen dahin gehend bestrebt, möglichst effizient nach potenziell neuen Wirkstoffkandidaten zu suchen (engl. *screening*) und diese auf den Markt zu bringen. Und genau hier setzt die QSAR-Analyse an.

Aus einem Pool von mehreren Tausend potenziellen Wirkstoffkandidaten gilt es, solche mithilfe einer QSAR-Analyse/eines QSAR-Modells auszusortieren, die *in vitro* den gewünschten physiologischen Effekt aufweisen bzw. nicht aufweisen.^[3,35] Dabei unterliegt der QSAR, direkt oder indirekt, ein Ähnlichkeits-Prinzip. Haben Substanzen ähnliche strukturelle und/oder physikochemische Eigenschaften, so sollte auch ihr physiologischer Effekt ähnlich sein.^[31]

Bei dem schon frühen Aussortieren mithilfe eines QSAR-Modells lassen sich zwei Aspekte festhalten, die einerseits für einen Wirkstoffkandidaten die ersten



großen Hürden bei der Arzneimittelentwicklung darstellen, andererseits aber auch für den Aufschwung der QSAR-Analyse verantwortlich sind.

Zum einen ist es die Toxikologie und zum anderen das sog. ADME-Prinzip (Absorption-Distribution-Metabolismus-Exkretion). Bei der Toxizitätsbestimmung einer Substanz werden mithilfe der QSAR-Analyse wichtige Endpunkte wie die akute Toxizität, die Mutagenität und die Karzinogenität betrachtet bzw. vorhergesagt.^[36] Somit bietet die QSAR-Analyse den entscheidenden Vorteil, dass durch ihre Anwendung auf die Durchführung umstrittener Tierversuche zur Toxizitätsbestimmung bei Vorliegen verlässlicher QSAR-Modelle verzichtet werden kann.

Auf der anderen Seite muss ein neues Arzneimittel bei seiner Zulassung ein ausreichend gutes pharmakokinetisches Verhalten aufweisen. Dieses umfasst die Aufnahme (Absorption) des Wirkstoffs in die Blutbahn, die Verteilung (Distribution) im Gewebe, den Um- und Abbauprozess (Metabolismus) und die Ausscheidung (Exkretion). Mithilfe einer QSAR-Analyse kann *in vitro* schon früh bei der Suche nach einem Wirkstoffkandidaten dessen ADME-Verhalten vorhergesagt und der Kandidat ggf. aussortiert werden. Hierbei dient u. a. das Schema der „5er-Regel“ (engl. *rule of five*) von Lipinski^[37] als Ausschlusskriterium.