

Berechnung umweltrelevanter thermodynamischer Stoffeigenschaften aus der
Molekülstruktur – Parametrisierung des quantenchemischen COSMO-RS-Modells für das
semiempirische AM1-Verfahren

Der Fakultät für Chemie und Mineralogie

der Universität Leipzig

eingereichte

DISSERTATION

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM

(Dr. rer. nat.)

vorgelegt von

Diplom Chemiker Björn Gunnar Loeprecht

geboren am 18.08.1971 in Donauwörth

Leipzig, den 1. 10. 2001

Danksagung

Die vorliegende Arbeit fasst Forschungsergebnisse, aus der Zeit von März 1998 bis Mai 2001 am Umweltforschungszentrum Leipzig-Halle zusammen.

Ich danke Herrn Prof. Dr. Schüürmann und Herrn Prof. Dr. Reinhold für die wissenschaftliche Betreuung und Dr. Andreas Klamt für klärenden Gespräche über die COSMO-RS-Theorie.

Bei den Mitarbeitern und „Ehemaligen“ der Arbeitsgruppe „Theoretische Methoden in der Umweltchemie“ der Sektion Chemische Ökotoxikologie Ralf-Uwe Ebert, Tobias Kind, Dr. Ralf Kühne, und Uwe Dietze, möchte ich mich für informative Gespräche über physikochemische Stoffeigenschaften und statistische Methoden bedanken.

Ralf-Uwe Ebert gilt besonderer Dank für die Unterstützung bei der Zusammenstellung und Bewertung experimenteller Stoffdaten.

Herrn Dr. Paschke danke ich für die Bestimmung der experimentellen Fischöl/Wasser-Verteilungskoeffizienten.

Besonderer Dank gilt meiner lieben Freundin Annett Hoyer, und all meinen Freunden mit denen ich viel Spaß in Leipzig hatte.

Dank auch an meine Eltern und Großeltern.

Für „Schnurre“

Inhaltsverzeichnis

1	EINFÜHRUNG UND MOTIVATION	1-1
2	THEORETISCHE GRUNDLAGEN	2-1
2.1	DAS KONTINUUMSSOLVENSMODELL COSMO	2-1
2.1.1	<i>Numerische Berechnung der COSMO-Solvatationsenthalpie</i>	<i>2-3</i>
2.1.2	<i>Gradienten</i>	<i>2-6</i>
2.2	COSMO-RS	2-8
2.2.1	<i>Grundprinzip.....</i>	<i>2-8</i>
2.2.2	<i>Typische Kontaktsegmente und Misfitenergie</i>	<i>2-10</i>
2.2.3	<i>Ladungsdichten und \mathbf{s}-Profil.....</i>	<i>2-12</i>
2.2.4	<i>Freie Energie und chemisches Potenzial</i>	<i>2-14</i>
2.2.5	<i>Erweiterung der COSMO-RS-Theorie.....</i>	<i>2-17</i>
2.2.6	<i>Chemisches Potenzial in der Gasphase</i>	<i>2-19</i>
2.3	PHYSIKOCHEMISCHE STOFFEIGENSCHAFTEN	2-20
2.3.1	<i>Die Henry-Konstante.....</i>	<i>2-20</i>
2.3.2	<i>Der Dampfdruck</i>	<i>2-23</i>
2.3.3	<i>Der Oktanol/Wasser-Verteilungskoeffizient</i>	<i>2-24</i>
2.3.4	<i>Die Wasserlöslichkeit.....</i>	<i>2-26</i>
3	COSMO-RS/AM1-PARAMETEROPTIMIERUNG	3-1
3.1	DER MOLEKÜLDATENSATZ	3-3
3.1.1	<i>Der Trainingsdatensatz.....</i>	<i>3-5</i>
3.1.2	<i>Der Vorhersagedatensatz.....</i>	<i>3-7</i>
3.2	OPTIMIERUNGEN UND PARAMETER	3-7
3.3	OPTIMIERUNGEN MIT DEM CHO-TEILDATENSATZ	3-13
3.4	CHONFCL-OPTIMIERUNGEN.....	3-20
3.4.1	<i>Lokale Grenzorbitale.....</i>	<i>3-22</i>
3.4.2	<i>Core-Ausgleichsladungen</i>	<i>3-28</i>
3.5	METHODENVERGLEICHE.....	3-40
3.5.1	<i>COSMO-RS/DFT.....</i>	<i>3-41</i>
3.5.2	<i>Weitere Kontinuumsmodelle</i>	<i>3-42</i>
3.5.3	<i>Strukturinterpolierende Methoden</i>	<i>3-43</i>
3.5.4	<i>Vergleichende Statistik.....</i>	<i>3-44</i>
3.5.4.1	<i>Henry-Konstanten.....</i>	<i>3-45</i>
3.5.4.2	<i>Dampfdruck.....</i>	<i>3-50</i>

3.5.4.3	Oktanol/Wasser-Verteilungskoeffizient.....	3-52
4	COSMO-RS-ANWENDUNGEN.....	4-1
4.1	BERECHNUNG TEMPERATURABHÄNGIGER HENRY-KONSTANTEN VON 21 AROMATISCHEN KOHLENWASSERSTOFFEN MIT COSMO-RS/AM1	4-1
4.2	TEMPERATURABHÄNGIGKEIT VON S_w , P_v UND H^s AM BEISPIEL VON VIER CHEMIKALIEN	4-9
4.3	MODELLIERUNG DES FISCHÖL/WASSER-VERTEILUNGSKOEFFIZIENTEN	4-17
4.4	ABSCHÄTZUNG DES BLUT/HIRNSCHRANKEN-ÜBERGANGSKOEFFIZIENTEN $\log P_{BB}$	4-21
5	ZUSAMMENFASSUNG UND AUSBLICK.....	5-1
6	LITERATURVERZEICHNIS.....	6-1
A	ANHANG.....	A-1
A.1	MOLEKÜLDATENSÄTZE.....	A-1
	<i>Tabelle A.1 Trainingsdatensatz.....</i>	<i>A-1</i>
	<i>Tabelle A.2 Vorhersagedatensatz.....</i>	<i>A-16</i>
A.2	REFERENZEN DER EXPERIMENTELLEN STOFFEIGENSCHAFTEN	A-21
A.3	FISCHÖL/WASSER- VERTEILUNGSKOEFFIZIENT-MESSUNGEN	A-26
A.4	TABELLEN ZU DEN TEMPERATURABHÄNGIGEN BERECHNUNGEN IN KAPITEL 4.2.....	A-28

Abkürzungs- und Symbolverzeichnis:

Abkürzung oder Symbol	Wert Einheit	Bedeutung
A	Å ²	Moleküloberfläche
AM1		Austin Model 1
AMSOL		Semiempirisches Softwarepaket
c _{hb}		Wasserstoffbrückenparameter (skalierend)
CLOGP		Software zur Berechnung des 1-Oktanol/Wasser-Verteilungskoeffizienten
COSMO		Conductor-like screening model
COSMO-RS		Conductor-like screening model for real solvents
c _S	mol/L	Konzentration eines Stoffes im Lösungsmittel S
DFT		Dichtefunktionaltheorie
E	kJ/mol	Energie
F	kJ	freie Energie
f _{corr}		elektrostatischer Korrekturfaktor (COSMO-RS-Parameter)
f _{pol}		Polarisationsfaktor (COSMO-RS-Parameter)
H	Pa L/mol	Henry-Konstante
H'		Henry-Konstante
HF		Hartree Fock
k	1.3806 * 10 ⁻²³ J/K	Boltzmannkonstante
K		Gleichgewichtskonstante
K _{OW}		Oktanol/Wasser-Verteilungskoeffizient
LSER		Linear Solvation Energy Relationship
MAF		Mittlerer absoluter Fehler
MG	g/mol	Molgewicht
MOPAC		Molecular Orbital Package (Semiempirisches Softwarepaket)
MST		Kontinuum Solvationsmodelle nach Miertus Scrocco und Tomasi
N		Teilchenzahl
n _{ra}		Anzahl der Ringatome eines Moleküls
p(σ)		Sigmaprofil
PCM		Polarizable Continuum Model
PM3		(MNDO-PM3) Modified Neglect of Diatomic Overlap, Parametric Method Number 3
P _v	bar	Dampfdruck
q _μ	e	Ladung eines Segments auf der SAS
QSAR		Quantitative Struktur-Wirkungs-Beziehungen (Quantitativ Structure Activity Relationships)
QSPR		Quantitative Struktur-Eigenschafts-Beziehungen (Quantitativ Structure Property Relationships)

Abkürzungsverzeichnis

r	Å	Radius
R	8.31441 J/K mol	Gaskonstante
r^2		quadratischer Korrelationskoeffizient
S	kJ/mol K	Entropie
SAS	Å ²	Solvent Accessible Surface (Einem Lösungsmolekül zugängliche Moleküloberfläche)
S_μ	Å ²	Fläche eines Segments der SAS
SD		Standardabweichung (Standard Deviation)
SMx		Kontinuum Solvationsmodelle von Cramer und Truhlar
SOR		Fehlerquadratsumme (Sum of residuals)
S_w	mol/L	Wasserlöslichkeit
T	K	Temperatur
U	kJ	innere Energie
V	L	Volumen
x_i		Molenbruch
Z		Zustandssumme
a		Selbstenergiekonstante (COSMO-RS-Parameter)
g_k		Elementspezifischer Dispersionsparameter (COSMO-RS-Parameter)
d		Kroneckersymbol
ΔG_{hydr}	kJ/mol	freie Hydratationsenthalpie
e		Dielektrizitätskonstante
h		Entropieparameter (COSMO-RS-Parameter)
m	kcal/mol	chemisches Potenzial
r	g/L	Dichte
s_{hb}	e/Å ²	Wasserstoffbrückenparameter (Ladungsdichte-Schwellwert) (COSMO-RS-Parameter)
s_μ	e/Å ²	Ladungsdichte eines COSMO-RS Segments
t		linearer Korrekturparameter für Hydroxygruppen (COSMO-RS-Parameter)
w		Ringparameter (COSMO-RS-Parameter)

1 Einführung und Motivation

In den letzten dreißig Jahren hat sich in der industrialisierten Welt und insbesondere auch in Europa ein Umweltbewusstsein herausgebildet, und Begriffe wie Ökologie, Ökosysteme und Ökotoxikologie wurden geprägt und mit Inhalt gefüllt. Die Einsicht, dass die Lebensweise der menschlichen Erdbevölkerung großen Einfluss auf die Pflanzen- und Tierwelt hat, und dass deren Veränderungen auf die Menschen zurückwirken, führte zur wissenschaftlichen Untersuchung dieser Phänomene und zur Suche nach Verhaltenschemata für soziale Gemeinschaften, die einen schonenden Umgang mit der Umwelt ermöglichen. In diesem Zusammenhang wurde mehr und mehr untersucht, in welchem Ausmaß vom Menschen produzierte Stoffe auf die Umwelt einwirken und wie sie diese beeinflussen.

Speziell die Entfaltung der chemischen Großindustrie in der Zweiten Hälfte des vorletzten Jahrhunderts führte zu einem immensen Ausstoß an chemischen Substanzen, die in der Natur gar nicht oder nur in erheblich geringeren Konzentrationen vorlagen. Nach Auskunft der Organization of Economic Cooperation and Development (OECD) sind ca. 70000 Chemikalien – die meisten davon organisch – im täglichen Gebrauch (1992) [1], und es werden mehr¹. Obwohl eine Vielzahl dieser Substanzen nicht als umweltgefährdend eingeschätzt wird, gibt es doch zahlreiche Chemikalien mit hohem Gefährdungspotenzial, welche in großen Mengen in die Umwelt gelangen: Lösungsmittel, Farbstoffe, Zusatzstoffe in Kunststoffen und Textilien, Düngemittel, Herbizide, Insektizide, Fungizide und Rückstände der fossilen Energiegewinnung und Verbrennung. Nach Informationen des Umweltbundesamtes [2] wurden 1999 in Deutschland 100 neue Stoffe angemeldet, von denen 76 einer Risikobewertung unterzogen wurden. Die Ergebnisse lassen aufmerken: Für sieben Stoffe sind weitere Informationen und Prüfnachweise bei Erreichen der Vermarktungsmenge von 10 t/a notwendig, wohingegen für fünf Stoffe sofortige Prüfnachweise und zusätzliche Informationen oder sogar Risikominderung gefordert wurde.

In der Europäischen Union wurden 1999 insgesamt 427 Chemikalien angemeldet. Eine umfassende Risikobewertung erscheint unerlässlich, wenn Umweltgefährdungen präventiv ausgeschlossen werden sollen. Die chemische Industrie bemängelt aber die hohen Kosten der

¹ Im aktuellen EINECS-Register (European INventory of Existing Commercial Substances) sind bereits 100195 Stoffe gelistet

Risikobewertung [3]. Eine pflichtmäßige Prüfung von 2700 Substanzen seit 1982 kostete die Industrie mehr als eine Milliarde Mark. Dies sei unverhältnismäßig, da diese Substanzen nur einen Marktanteil von 0.01 Prozent haben. Durch hieraus resultierende Wettbewerbs- und Innovationsnachteile steigere sich der finanzielle Schaden bis in mehrstellige Milliardenbeträge. Die Notwendigkeit von präventivem Umweltschutz und die Forderung nach Kostenverringerung legitimieren Forschungen in der Risikobewertung von Xenobiotika. Der Wissenschaftszweig der *chemischen Ökotoxikologie* beschäftigt sich deshalb mit der Erforschung der Auswirkungen von Chemikalien auf Flora, Fauna, ganze Ökosysteme und nicht zuletzt auf den Menschen selbst.

Die biochemischen Wirkweisen chemischer Substanzen sind so vielfältig wie die unterschiedlichen strukturellen Merkmale der Stoffe. Werden Organismen durch Chemikalien geschädigt, ergeben sich neben Fragestellungen der biochemischen Wirkungen und der analytischen Frage nach der gefährdenden oder toxischen Menge auch die Fragen wie, warum, und in welchen Mengen die Chemikalie zum Organismus gelangte, und warum sie von diesem aufgenommen wurde. Zur Klärung dieser Fragen muss man wissen, wie sich die Substanz in der Umwelt verteilt hat, wo die Substanz angereichert wird, und wie schnell sie durch biochemische, photochemische oder thermochemische Reaktionsmechanismen abgebaut wird. Es ist also von großem Interesse, das *Verteilungsbestreben* und den *Transport* von Substanzen zwischen den Kompartimenten der Umwelt zu verstehen. In den letzten Jahren wurden große Anstrengungen unternommen, der *Gleichgewichtsthermodynamik* entlehnte *makroskopische physikalische Stoffeigenschaften*, welche das Verteilungsbestreben von Chemikalien beschreiben, direkt *aus der chemischen Struktur* vorherzusagen.

Wasser spielt eine herausragende Rolle in unserer Umwelt, ist es doch allgegenwärtig: fluid in Flüssen, Seen und Meeren, gasförmig in der Luft, kondensiert als Nebel, in fester Form als Eis, und als Bestandteil des Bodens und aller Lebensformen. Aus diesem Grunde spielt es auch in dieser Arbeit eine bevorzugte Rolle. An zwei von drei Verteilungskoeffizienten, mit deren Hilfe das COSMO-RS-Verfahren (COSMO-RS = Conductor-like screening model for realistic solvents) zur Berechnung physikochemischer Stoffkonstanten angepasst wird, ist Wasser als Lösungsmittel beteiligt: bei der Henry-Konstante und beim Oktanol/Wasser-Verteilungskoeffizient. Die Henry-Konstante H [4-718] gibt das Bestreben einer Chemikalie an, sich zwischen Luft und Wasser zu verteilen, und wird deshalb beispielsweise zur Beschreibung der Verdunstung aus Oberflächengewässern verwendet. Ebenso lassen sich mit

der Henry-Konstante Vorhersagen machen, in wieweit Chemikalien durch Regen in der unteren Troposphäre ausgewaschen werden. Die Henry-Konstante in ihrer logarithmischen Form hängt linear von der freien Hydratationsenthalpie ΔG_{hydr} ab (siehe Gleichung 66 Seite 2-21), und ist eine Funktion der Temperatur. Deshalb ist der oft benützte 25°C-Wert nicht für alle Jahreszeiten oder Erdregionen die richtige Wahl.

Die Verdampfung von Chemikalien aus trockenen Medien hängt stark von der Temperatur und von der Verdampfungswärme ab, und deshalb ist auch der Dampfdruck von großem Interesse für den Umweltchemiker. Hohe Dampfdrücke P_V korrespondieren mit hohen Transportraten und mit einer höheren Aufnahme der Substanz durch terrestrische Pflanzen. Neben diesen in dieser Arbeit speziell behandelten gibt es natürlich zahlreiche weitere umweltrelevante Stoffeigenschaften: Wasserlöslichkeitsdaten werden ebenfalls zur Beschreibung von Transportvorgängen herangezogen. Je höher die Wasserlöslichkeit eines Stoffes ist, desto größer ist auch sein Dispersionsbestreben, und desto kleiner ist beispielsweise die Ablehnung gegenüber abiotischem und biotischem Abbau. All diese physikalischen Eigenschaften zusammen werden nach Mackay [8] benützt, um die Verteilung einer Substanz in der Umwelt vorherzusagen.

Die direkte Vorhersage solcher Eigenschaften aus der chemischen Struktur (quantitative structure-property relationships, QSPRs) hat nun schon Tradition seit den ersten Veröffentlichungen von Leo und Hansch [9,10]. Eine Vielzahl von Rechenmodellen, vor allem rein empirische Methoden wie die Gruppenbeitragsmethoden (Inkrementmethoden) wurden seitdem zur Berechnung des Oktanol/Wasser-Verteilungskoeffizienten [11,12] und der Henry-Konstante entwickelt. In den letzten Jahren wurden auch neuronale Netzwerke [13] oder anspruchsvolle quantenchemische Modelle [14] zur Berechnung makroskopischer Eigenschaften von Molekülen im thermodynamischen Gleichgewicht herangezogen. Mit der Entwicklung der quantenchemischen Kontinuumsmodelle und deren Implementierung in verschiedene Programmpakete [15] eröffnen sich neue Möglichkeiten des theoretischen Studiums physikalischer Eigenschaften von fluiden Phasen [16-18].

Warum COSMO-RS/AMI?

Obwohl mit den hochparametrisierten Gruppenbeitragsmethoden eine Vielzahl an Spezialverfahren zur Berechnung verschiedener Stoffeigenschaften existieren, bietet der COSMO-RS-Ansatz dennoch einige grundlegende Vorzüge. Im Gegensatz zu den Inkrementverfahren, bei denen die Moleküle in einzelne Unterfragmente zerlegt werden deren zugeordnete Beiträge dann additiv die Stoffeigenschaft ergeben, berücksichtigt COSMO-RS die dreidimensionale

Struktur des Stoffes. So können feine Unterschiede zwischen Strukturisomeren oder gar Stereoisomeren aufgelöst werden. Die breite Anwendbarkeit des COSMO-RS-Modells (beliebig zusammengesetzte fluide Systeme sind berechenbar) ist ein weiteres Plus.

An dieser Stelle wird auf einen Widerspruch aufmerksam gemacht: *Thermodynamische Stoffeigenschaften* können strenggenommen nicht aus *einer einzelnen 3D-Molekülstruktur* berechnet werden, da sich makroskopisch messbare Zustandsgrößen immer als Mittelwert eines statistischen Ensembles von Molekülen in verschiedenen energetischen Zuständen (also auch in verschiedenen Konformationen) ergeben. COSMO-RS kann diesem Umstand Rechnung tragen, denn es können Boltzmann-gewichtete Konformeren- oder Stereoisomeren-Mischungen behandelt werden.

Ist COSMO-RS für ein Set von chemischen Elementen parametrisiert, kann es *alle* denkbaren Strukturkombinationen berechnen. Dies ist wohl der größte prinzipielle Vorteil gegenüber den zweidimensionalen Gruppenbeitragsmethoden. Ein weiterer grundlegender Vorteil ist die implizit berücksichtigte Temperaturabhängigkeit der modellierten chemischen Potentiale. So ergeben sich die makroskopischen Eigenschaften fluider Systeme zwanglos als Funktion der Temperatur. Einzig Verteilungsgrößen, die mit dem chemischen Gasphasenpotenzial berechnet werden, benötigen eine zusätzliche Anpassung an experimentelle Daten. Zwar gibt es für die Berechnung der Henry-Konstante bei Raumtemperatur verschiedene Gruppenbeitragsverfahren [19,20] und Kontinuumsmodelle, doch COSMO-RS ist das einzige Verfahren, mit dem die Henry-Konstante direkt temperaturabhängig berechnet werden kann. Durch die physikalisch fundierten Grundlagen von COSMO-RS ist es beispielsweise möglich, Lösungsvorgänge energetisch genauer zu studieren. Abschätzungen der Anteile elektrostatischer Stabilisierung oder dispersiver Energieanteile der Solvatationsenthalpien sind möglich.

Bisher existiert noch *keine* systematisch ermittelte COSMO-RS-Parametrisierung auf der Grundlage eines semiempirischen Hamiltonoperators. Die Motivation für die Ermittlung eines semiempirischen COSMO-RS-Parametersatzes ist die hohe Recheneffizienz der semiempirischen Methoden. Im Gegensatz zu den vorhandenen DFT-Varianten können dann auch sehr große Moleküle behandelt werden. COSMO-RS unterliegt keinen prinzipiellen Beschränkungen, was die Größe der zu untersuchenden Moleküle angeht. Ein Fernziel ist sicher die Behandlung von Biomolekülen mit anspruchsvollen funktionellen Gruppen, denn für eben diese versagen auch die bekannten Gruppenbeitragsmethoden aufgrund fehlender Parametrisierungen. Mit COSMO-RS/AM1 könnten so schon vor der Synthese einer Substanz deren Umweltverträglichkeit (*ecological design*) und deren pharmakologisches Potenzial (*virtual*

drug design) abgeschätzt werden. Durch die relativ einfach und schnell durchzuführenden Berechnungen ist COSMO-RS/AM1 eine Methode, die sich hervorragend in (*in silico*) *high throughput* Prozeduren integrieren lässt. Hierbei sollte die große Variabilität der COSMO-RS-Methode voll zur Geltung kommen. Muss nämlich für jede Stoffeigenschaft bei anderen Methoden für jedes einzelne Molekül eine neue Rechnung durchgeführt werden, so ist bei COSMO-RS nur eine quantenchemische Rechnung pro Molekül notwendig, aus der die verschiedenen Stoffeigenschaften schnell und effizient berechnet werden können.

In der vorliegenden Arbeit wird eine COSMO-RS-Grundparametrisierung basierend auf drei umweltrelevanten Stoffeigenschaften für das AM1-Verfahren vorgestellt. Dieser Parametersatz kann in späterer Zeit als Grundlage für einen vergrößerten Satz von Elementen dienen, oder eine Basis für eine Vielzahl von Spezialparametrisierungen bilden. Natürlich leidet eine semiempirische COSMO-RS-Parametrisierung auch an den Unzulänglichkeiten des semiempirischen Modells selbst. So kann eine Parametrisierung auf semiempirischen Niveau wohl nicht die Qualität von COSMO-RS/DFT erreichen, da ja schon die zugrundeliegende theoretische Beschreibung der Moleküle aufgrund der semiempirischen Näherungen ungenauer ist. Deshalb wird schon in der vorliegenden Arbeit versucht, zusätzliche Veränderungen und Parameter einzuführen, um die Vorhersagequalität des COSMO-RS/AM1-Modells zu verbessern. Die Suche nach Eingriffsmöglichkeiten und deren programmtechnische Umsetzung bildet den Hauptteil des Forschungsprojekts, welches in der vorliegenden Arbeit zusammengefasst wird.

Um die Methodik der Optimierung und die dabei eingeführten programmtechnischen Veränderungen zu verstehen, ist es unerlässlich, in die COSMO-RS-Theorie einzuführen und alle relevanten Gleichungen abzuleiten. Dies soll im nächsten Kapitel geschehen.

2 Theoretische Grundlagen

2.1 Das Kontinuumsmodell COSMO

Das COSMO-Verfahren [21] (COSMO = **CO**nductor-like **S**creening **MO**del) zur Modellierung von Lösungsmittelleffekten ist ein Verfahren aus der Klasse der Kontinuumsmodelle. Aufgrund der Existenz einer Vielzahl verschiedener Kontinuumsansätze soll im Folgenden nur die COSMO-Methode näher erklärt werden. Zur weiteren Einführung in die Materie wird auf Übersichtsartikel [15,22-25] verwiesen. Bei allen Kontinuumsansätzen (*CSM Continuum Solvation Models*) wird ein gelöstes Molekül als ein in ein dielektrisches Kontinuum mit der Dielektrizitätszahl ϵ eingebettetes Molekül betrachtet. Die Grenzfläche zwischen gelöstem Molekül und Kontinuum bildet die Solvent Accessible Surface (SAS). Im Gegensatz zu älteren Ansätzen wird die Geometrie des um das Molekül gebildeten Käfigs nicht durch eine Kugel oder ein Ellipsoid angenähert, sondern explizit nach der Molekülgeometrie geformt. Zur Konstruktion dieser Hohlräume (Käfige) wird um die außenliegenden Atome des betrachteten Moleküls eine Oberfläche unter Verwendung von elementspezifischen van der Waals-ähnlichen Radien modelliert.

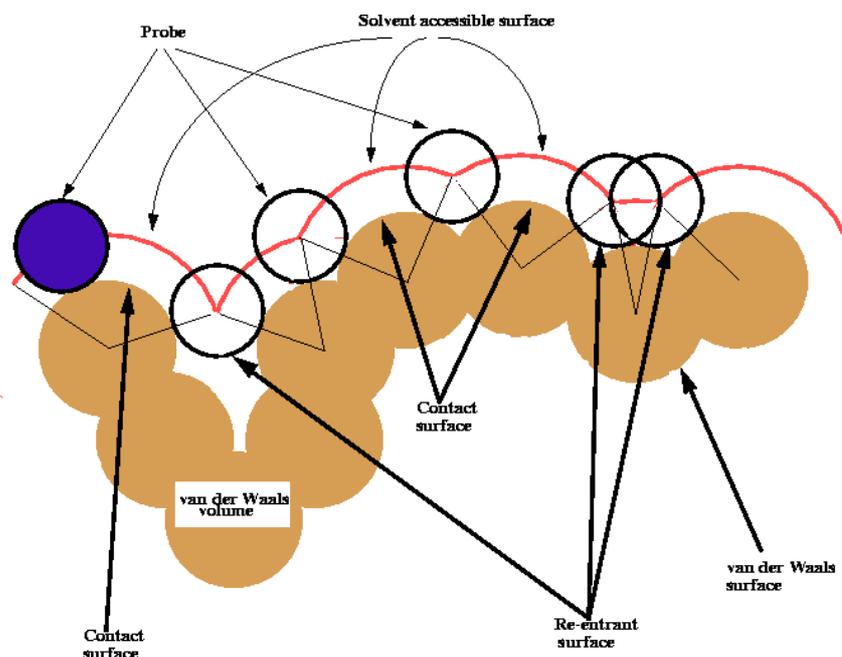


Abbildung 1: Konstruktion der Solvent Accessible Surface (SAS). Auf der mit van der Waals-Radien konstruierten Oberfläche wird eine Kugel (violett) abgerollt, um die SAS (rote Kurve) zu formen [26].

Um von der so gewonnenen Moleküloberfläche zu einem vom Lösungsmittel nicht zugänglichen Hohlraum zu gelangen, wird auf der Oberfläche noch eine Probekugel mit Radius r_s abgerollt (siehe Abbildung 1). Die elektronischen Ladungen der Atomkerne und Elektronen des Moleküls im Hohlraum polarisieren das Kontinuum außerhalb des Käfigs, welches seinerseits auf die Ladungsverteilung zurückwirkt. In der Literatur wird die elektrostatische Antwort des polarisierten Dielektrikums oft als Reaktionsfeld (reaction field) bezeichnet. Für die Dielektrizitätszahl innerhalb des Käfigs wird ein Wert von eins angenommen, während außerhalb des Käfigs das postulierte Lösungsmittel die Dielektrizitätszahl bestimmt. Es gilt also:

$$\varepsilon(\mathbf{r}) = 1 \quad \mathbf{r} \in V_{in} \quad \text{Gleichung 1}$$

$$\varepsilon(\mathbf{r}) = \varepsilon \quad \mathbf{r} \in V_{out} \quad \text{Gleichung 2}$$

Bei manchen Kontinuumsmodellen wird ε für Regionen außerhalb des Käfigs auch als Funktion des Orts betrachtet. Zusätzlich wird bei den Kontinuumsmodellen angenommen, dass sich die gesamte Ladung des gelösten Moleküls innerhalb des Käfigs befindet. Dies ist bei Molekülen im allgemeinen nicht der Fall und der resultierende Ladungsfehler muss korrigiert werden [27-29]. Die zentrale Gleichung für die elektrostatischen Kontinuumsmodelle ist die Poissonsche Gleichung:

$$-\frac{4\pi\mathbf{s}(r)}{e} = \nabla^2\Phi(r) \quad \text{Gleichung 3}$$

Sie beschreibt die Beziehung zwischen der Ladungsdichte \mathbf{s} am Punkt r und der Divergenz des Gradientenfeldes des elektrostatischen Potentials $\Phi(r)$. Diese Gleichung gilt es, bei Kontinuumsansätzen zu lösen. Je nach Wahl der Randbedingungen gibt es unterschiedliche Lösungsansätze für die Poisson-Gleichung. Im Falle sphärischer Molekülkäfige mit konstanter Dielektrizitätszahl außerhalb des Käfigs lässt sich das elektrostatische Potenzial nach einer Multipolentwicklung mit Hilfe der bekannten Legendrepolynome analytisch bestimmen.

Eine Berechnung der dielektrischen Abschirmenergie für Moleküle in beliebig geformten Hohlräumen – formal entspricht dies der Suche nach einer geeigneten Greenschen Funktion [30] – ist analytisch nicht mehr möglich. Eine elegante Lösungsmöglichkeit bieten die Abschirmladungsmethoden. Hier werden die von einer Ladungsverteilung induzierten Abschirmladungen auf der Kontaktfläche zwischen Dielektrikum und Molekül (also auf der SAS) berechnet. Abbildung 2 zeigt schematisch ein Molekül in einem Dielektrikum mit

den Abschirmladungen auf der SAS. Wie erwähnt, muss die Berechnung numerisch erfolgen. Hierzu wird die SAS in eine große Anzahl von M Segmenten S_μ mit konstanten Ladungsdichten σ_μ unterteilt. Das weit verbreitete PCM-Modell (PCM = Polarizable Continuum Model) nach Miertus, Scrocco und Tomasi (MST-Modell) ist ein Vertreter dieser Klasse. Dieses Verfahren wird auch als ASC-Modell (ASC = Apparent Surface Charge) bezeichnet. Auch COSMO ist ein Abschirmladungsmodell, bei dem die SAS in kleine Flächenstücke eingeteilt wird.

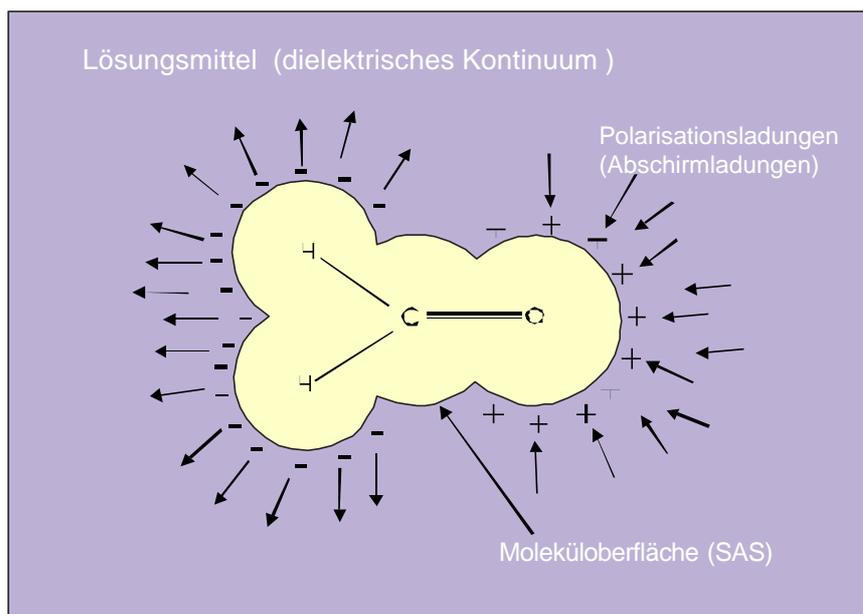


Abbildung 2: Abschirmladungen auf der dem Lösungsmittel zugängliche Oberfläche (SAS).

Beim COSMO-Ansatz wird nun der Formalismus der dielektrischen Abschirmenergie (dielectric screening energy) in elektrischen Leitern (d.h. $\epsilon = \infty$) angewendet. Grund hierfür ist die starke Vereinfachung der Mathematik und die daraus resultierende Einsparung von Rechenzeit und Speicherkapazität bei den Berechnungen. Die Abschirmenergie skaliert für eine gegebene Geometrie mit $f = (\epsilon - 1)/(\epsilon + x)$ mit $0 \leq x \leq 2$ und die Energien und Ladungen, die im COSMO-Bild (für $\epsilon = \infty$) erhalten werden, werden mit diesem Faktor f (mit $\epsilon = \epsilon_{\text{Lösungsmittel}}$ und $x = 0.5$) im SCF-Teil korrigiert.

2.1.1 Numerische Berechnung der COSMO-Solvatationsenthalpie

Zur numerischen Bestimmung der elektrostatischen Potentiale und des elektrostatischen Anteils der Solvatationsenergie wird die SAS in kleine Flächenstücke unterteilt. Die

Segmente befinden sich an den Orten t_m und jede Segmentfläche $|S_m|$ trägt die Ladung $q_m = |S_m| \mathbf{s}_m$. Dann ist:

$$b_{im} = \frac{1}{|S_m| 4\pi \epsilon_0} \int_{S_m} \|r - r_i\|^{-1} d^2 r \quad \text{Gleichung 4}$$

die elektrostatische Wechselwirkung einer Einheitsladung im Molekül am Ort r_i mit einer Standardladung des Segments S_m auf der Grenzfläche. Die Wechselwirkung zweier Ladungen auf den Käfigoberflächensegmenten $|S_m|$ und $|S_n|$ ist allgemein gegeben durch:

$$a_{mn} = \frac{1}{|S_m| |S_n| 4\pi \epsilon_0} \int_{S_m} \int_{S_n} \|r - r'\|^{-1} d^2 r' d^2 r \quad \text{Gleichung 5}$$

Für $m \neq n$ gilt näherungsweise:

$$a_{mn} = (4\pi \epsilon_0 \|t_m - t_n\|)^{-1} \quad \text{Gleichung 6}$$

Hier wird also die Segmentflächenladungsdichte durch eine im Segmentmittelpunkt befindliche Punktladung ersetzt.

Für die Eigenenergie eines Segments ist eine genauere Betrachtung notwendig. Wird eine *Einheitsladung* auf einer Kugel mit Radius R als homogen verteilt angenommen, so ist die elektrostatische Gesamtenergie gleich:

$$E_{tot} = \frac{1}{2R} \quad \text{Gleichung 7}$$

Bei Unterteilung der Oberfläche in M äquivalente gleich große Segmente mit der Fläche $|S|$:

$$|S| = \frac{4\pi R^2}{M} \quad \text{Gleichung 8}$$

kann die Gesamtenergie auch als Summe der Coulombenergien der Ladungen (M^{-1}) auf verschiedenen Segmenten addiert zur Summe der Selbstenergien bestimmt werden. Die Beziehung lautet dann:

$$\frac{1}{2R} = \frac{M}{2} \sum_{n=2}^M \frac{M^{-2}}{\|t_1 - t_n\|} + M M^{-2} \frac{a_{diag}}{2} \quad \text{Gleichung 9}$$

Wird Gleichung 9 nach dem gesuchten Selbstenergieterm aufgelöst, erhält man:

$$a_{diag} = \sqrt{4\pi} |S_m|^{-\frac{1}{2}} \sqrt{M} \left(M - \sum_{n=2}^M \frac{R}{\|t_1 - t_n\|} \right) \quad \text{Gleichung 10}$$

Für Segmentierungen von $M = 4$ bis $M = 3242$ nimmt das Produkt aus Klammerausdruck und $M^{-1/2}$ näherungsweise den Wert 1.07 an, und man erhält die genäherte Endformel für die Selbstenergie inklusive dem elektrostatischen Faktor $(4\pi\epsilon_0)^{-1}$:

$$a_{mm} = \frac{3.8}{4\pi\epsilon_0} |S_m|^{-\frac{1}{2}} \quad \text{Gleichung 11}$$

Mit den Beziehungen b_{im} , a_{mm} und a_{mm} für die Wechselwirkungsenergien der verschiedenen Ladungen lässt sich nun ein Ausdruck für die Gesamtenergie E des Systems formulieren. In Vektor- und Matrixschreibweise mit dem Vektor \mathbf{Q} für die N Ursprungsladungen Q_i des Moleküls im Käfig und dem Vektor \mathbf{q} für die M Ladungen q_j auf der Oberfläche erhält man:

$$E(\mathbf{q}) = \frac{1}{2} \mathbf{Q}^T \mathbf{C} \mathbf{Q} + \mathbf{Q}^T \mathbf{B} \mathbf{q} + \frac{1}{2} \mathbf{q}^T \mathbf{A} \mathbf{q} \quad \text{Gleichung 12}$$

Hier ist \mathbf{C} die Coulombmatrix mit den Matrixelementen:

$$c_{ij} = \|r_i - r_j\|^{-\frac{1}{2}} \quad \text{Gleichung 13}$$

$$c_{ii} = 0 \quad \text{Gleichung 14}$$

Die Terme der rechten Seite von Gleichung 12 repräsentieren von links nach rechts:

- die Wechselwirkungsenergien der Ursprungsladungen untereinander
- die Wechselwirkungsenergien der Ursprungsladungen mit den Ladungen der SAS
- die Wechselwirkungen zwischen Ladungen auf der SAS

Die bisher unbekannte Größe in Gleichung 12 ist der Vektor \mathbf{q} mit den Abschirmladungen. Eine optimale Abschirmladungsverteilung \mathbf{q}^* sollte die Gesamtenergie minimieren und deshalb kann \mathbf{q}^* aus folgender Extremwertbedingung gewonnen werden:

$$\nabla_{\mathbf{q}} E(\mathbf{q})|_{\mathbf{q}^*} = \mathbf{B}^T \mathbf{Q} + \mathbf{A} \mathbf{q}^* = 0 \quad \text{Gleichung 15}$$

Gleichung 15 kann umgeformt werden und man erhält \mathbf{q}^* als:

$$\mathbf{q}^* = -\mathbf{A}^{-1} \mathbf{B}^T \mathbf{Q} \quad \text{Gleichung 16}$$

Setzt man Gleichung 16 in Gleichung 12 ein, ergibt sich die elektrostatische Gesamtenergie zu:

$$E(\mathbf{s}^*) = \frac{1}{2} \mathbf{Q}^T (\mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T) \mathbf{Q} \quad \text{Gleichung 17}$$

Wird die Coulombwechselwirkung der Ladungen des gelösten Moleküls (dies entspricht der Gasphasenenergie) subtrahiert, kann die Solvatationsenthalpie Δ^X angegeben werden:

$$\Delta^X = -\frac{1}{2} \mathbf{Q}^T (\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T) \mathbf{Q} \quad \text{Gleichung 18}$$

Alternativ können die idealen Abschirmladungen auch aus der Beziehung des verschwindenden elektrostatischen Potentials auf der Oberfläche idealer Leiter ermittelt werden [31]. Dies führt ebenfalls zu Gleichung 16.

Nachdem dargelegt wurde, nach welchem mathematischen Formalismus bei der COSMO-Theorie für eine vorgegebene Molekülstruktur die Gesamtenergie ermittelt wird, sollen im Folgenden die für Geometrieoptimierungen notwendigen Gradienten der Energie bezüglich der Kernkoordinaten vorgestellt werden.

2.1.2 Gradienten

Unter der Voraussetzung, dass die einzelnen Segmente der SAS explizit nur jeweils einem Atom zugeordnet sind, lässt sich der Gradient der Energie bezüglich der Kernpositionen R_α angeben:

$$\begin{aligned} \nabla_{R_a} \Delta^X &= -\frac{1}{2} \nabla_{R_a} (\mathbf{Q}^T \mathbf{B}^T \mathbf{A}^{-1} \mathbf{B} \mathbf{Q}) \\ &= -\mathbf{Q}^T \mathbf{B}^T \mathbf{A}^{-1} (\nabla_{R_a} \mathbf{B}) \mathbf{Q} + \frac{1}{2} \mathbf{Q}^T \mathbf{B}^T \mathbf{A}^{-1} (\nabla_{R_a} \mathbf{A}) \mathbf{A}^{-1} \mathbf{B} \mathbf{Q} \\ &= -q^* (\nabla_{R_a} \mathbf{B}) \mathbf{Q} + \frac{1}{2} \mathbf{q}^* (\nabla_{R_a} \mathbf{A}) \mathbf{q}^* \end{aligned} \quad \text{Gleichung 19}$$

Eine physikalische Deutung weist dem ersten Summanden die Änderung der Coulombwechselwirkung der Ladungen im Käfig mit den Abschirmladungen zu, während der zweite Term die Änderung der Abschirmladungswechselwirkung beschreibt. Die Definitionen der einzelnen Matrixelemente lauten:

$$\nabla_{R_a} b_{im} \approx \nabla_{R_a} \|t_m - r_i\|^{-1} \approx \frac{t_m - r_i}{\|t_m - r_i\|^3} (\mathbf{d}_{ia} - \mathbf{d}_{ma}) \quad \text{Gleichung 20}$$

$$\nabla_{R_a} a_{mm} \approx 0 \quad \text{Gleichung 21}$$

$$\nabla_{R_\alpha} a_{m\alpha} \approx \nabla_{R_\alpha} \|t_m - t_n\|^{-1} \approx \frac{t_m - t_n}{\|t_m - t_n\|^3} (\mathbf{d}_{na} - \mathbf{d}_{ma}) \quad \text{Gleichung 22}$$

Hierbei veranschaulichen die Kroneckersymbole δ , dass der Gradient eines Matrixelements nur dann verschwindet, wenn die Ladung am Punkt r_i nicht am Atom α zentriert ist.

Betrachtet man anstelle von Punktladungen die Ladungsdichten des Moleküls, so lassen sich die Matrixelemente von \mathbf{B} auch mit Hilfe der Molekülorbitale darstellen:

$$b_{kl,m} = \frac{1}{|S_m|} \iint_{S_m} \frac{\mathbf{j}_k(r) \mathbf{j}_l(r)}{\|r' - r\|} d^3 r d^2 r' \approx \int \frac{\mathbf{j}_k(r) \mathbf{j}_l(r)}{\|t_m - r\|} \quad \text{Gleichung 23}$$

Bei der in dieser Arbeit verwendeten MOPAC-Implementierung wird die Ladungsverteilung des Moleküls mit atomaren Multipolen repräsentiert. Für eine genauere Beschreibung wird auf die COSMO-Referenz [21] verwiesen.

Mit der COSMO-Methode ist es also möglich, Energien und Geometrien beliebiger Moleküle in verschiedenen Lösungsmitteln zu berechnen. Unterschiedliche Lösungsmittel werden jedoch ausschließlich durch die jeweils zugehörige Dielektrizitätskonstante ϵ festgelegt. Die mit $f(\epsilon)$ korrigierte COSMO-Solvationsenergie lautet:

$$\Delta^x = -\frac{\epsilon - 1}{2\epsilon + 1} \mathbf{Q}^T (\mathbf{B} \mathbf{A}^{-1} \mathbf{B}^T) \mathbf{Q} \quad \text{Gleichung 24}$$

Zwar ist die COSMO-Theorie in sich geschlossen und kann Solvationsprozesse qualitativ richtig beschreiben, aber trotzdem sind vor allem zwei Defizite offensichtlich: Erstens berücksichtigt das Modell keinerlei strukturelle mikroskopische Eigenschaften der Lösungsmittel, und zweitens können mit COSMO ausschließlich elektrostatische Anteile der Solvationsenthalpie quantifiziert werden. Diese Betrachtung ist ungenügend, ist es doch hinlänglich bekannt, dass neben den elektrostatischen auch nicht-elektrostatische Energiebeiträge beim Lösungsvorgang eine Rolle spielen. Zu nennen sind hier beispielsweise Dispersions-, Polarisations-, Kavitations- oder Wasserstoffbrückenbindungsenergien. Deshalb ist der oft durchgeführte direkte Vergleich von COSMO mit anderen quantenchemischen Solvationsmodellen nur wenig sinnvoll, da diese meistens nicht-elektrostatische Energiebeiträge berücksichtigen. Sowohl beim PCM-MST-Modell als auch bei den etablierten SMx-Modellen von Cramer und Truhlar werden die nicht-elektrostatischen Anteile der freien Lösungsenthalpie nach unterschiedlichen Verfahren quantifiziert.

Für eine verbesserte Beschreibung beliebiger fluider Phasen wurde in die COSMO-Theorie eine mikroskopische Betrachtung des Lösungsmittels eingeführt, und die Theorie wurde um

nicht-elektrostatische Anteile erweitert. Mittels statistischer thermodynamischer Methoden können dann auch komplexe fluide Systeme mit der neuen COSMO-RS-Theorie quantitativ beschrieben werden.

2.2 COSMO-RS

Physikalisch betrachtet ist es unklar, warum sich Lösungsmittelleffekte im COSMO-Modell nur durch die makroskopische Dielektrizitätskonstante ϵ beschreiben lassen, treten doch bei der elektrostatischen Wechselwirkung zwischen Molekülen und Ionen sehr starke elektrostatische Felder auf, wohingegen die Theorie der elektrischen Abschirmung strenggenommen nur für homogen polarisierbare makroskopische Medien entwickelt wurde. Deshalb sollte diese Theorie eigentlich keine Gültigkeit für die elektrostatischen Wechselwirkungen auf atomarer Ebene haben, denn Moleküle werden keinesfalls homogen polarisiert und wichtiger noch: Die molekulare Reorientierungspolarisation ist nicht linear. Wie kann nun die COSMO-Theorie, innerhalb derer (wie in jeder anderen Kontinuumstheorie) das Lösungsmittel eben nur makroskopisch behandelt wird, um eine mikroskopische Sichtweise ergänzt werden? Eine relativ junge und vielseitige Möglichkeit bietet der COSMO-RS-Ansatz [32-34] (COSMO-RS = **C**onductor like **S**creening **M**odel for **R**eal **S**olvents), welcher sich der COSMO-Theorie bedient, und diese mit Aspekten der statistischen Thermodynamik weiterentwickelt. Die Grundprinzipien der COSMO-RS-Theorie soll das folgende Gedankenexperiment verdeutlichen.

2.2.1 Grundprinzip

Im COSMO-Bild betrachteten wir ein Lösungsmittel als homogenes Kontinuum, welches nur durch *eine* spezifische Elektrizitätskonstante beschrieben wird - ganz gleich welche mikroskopische Zusammensetzung des Lösungsmittels im Realfall vorliegt. Diese Vereinfachung wird nun zurückgenommen, indem auch die Lösungsmittelmoleküle mikroskopisch behandelt werden. Betrachten wir ein Ensemble von vier kubischen Lösungsmittelmolekülen S und einem gelösten Molekül X, welche in ein ideal elektrisch leitendes Medium eingebettet sind und dadurch den Zustand idealer Abschirmung erreicht haben. Die Abschirmung kann wie in der COSMO-Theorie beschrieben durch Abschirmladungsdichten auf den Moleküloberflächen veranschaulicht werden (Abbildung 3-A).

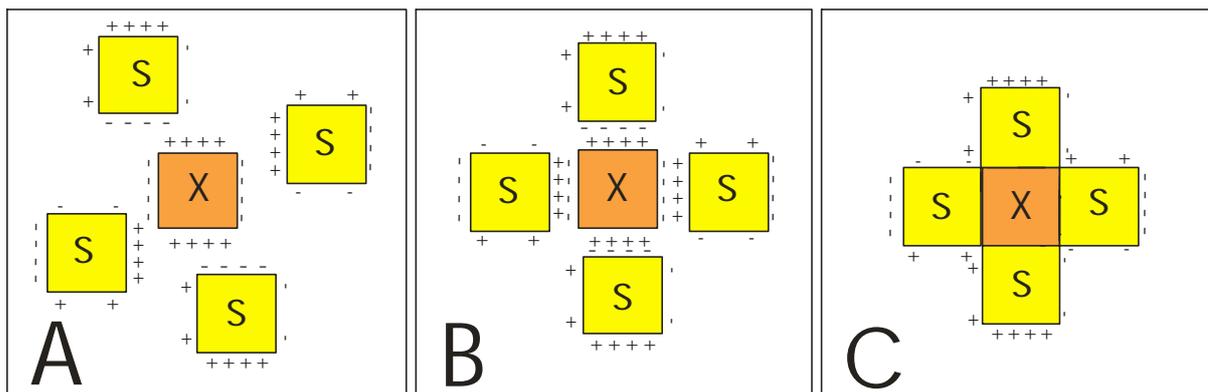


Abbildung 3 A bis C: Idealisierte kubische Moleküle in einem perfekt leitenden Medium

Es wird nun angenommen, dass die Oberflächenladung auf den kubischen Moleküloberflächen nahezu konstant ist und dass die Lösungsmittelmoleküle dem gelösten Molekül genau entgegengesetzte Oberflächenladungen präsentieren können. Durch die perfekte Abschirmung können die Moleküle ohne Energieänderung verschoben werden, und können nun so arrangiert werden, dass sich die gegensätzlichen Oberflächenladungen zu Null kompensieren (Abbildung 3-B). In diesem Fall macht es energetisch keinen Unterschied, ob die perfekt leitenden Flächen zwischen den Molekülen entfernt werden oder nicht. Schließlich erhalten wir ein gelöstes Molekül X, welches von vier Lösungsmittelmolekülen ideal abgeschirmt wird (siehe Abbildung 3 C).

Hieraus folgt:

Kann das Lösungsmittel einem Molekül für dessen gesamte Oberfläche die passenden gegensätzlichen Oberflächenladungen präsentieren, dann kann es dieses genauso gut wie ein perfekter Leiter abschirmen.

In der realen Lösung kann man in keiner Weise von kubischen Molekülen sprechen und es wird auch nicht immer eine ideale Oberflächenladungskompensation möglich sein.

Abbildung 4 (Seite 2-10) veranschaulicht zweidimensional, wie im COSMO-RS-Bild eine Flüssigkeit behandelt wird: als Ensemble von Molekülen, welche in sich berührenden Hohlräumen untergebracht sind. Die Größe der Hohlräume wird durch elementspezifische van der Waals-ähnliche Radien festgelegt. Die infinitesimal dicken Berührungsflächen sollen, wie schon erwähnt, aus perfekt leitenden Materialien bestehen. Jedes Molekül der betrachteten Flüssigkeit befindet sich also in der ideal abgeschirmten Situation und kann somit *einzel*n mit einer COSMO-Rechnung behandelt werden. Da die Abschirmladungen einer Molekülsorte im perfekt leitenden Medium immer gleich sind, muss also nur eine COSMO-Rechnung mit $\epsilon = \infty$ für jede Molekülart (eigentlich für jedes Konformer) der Flüssigkeit durchgeführt werden.

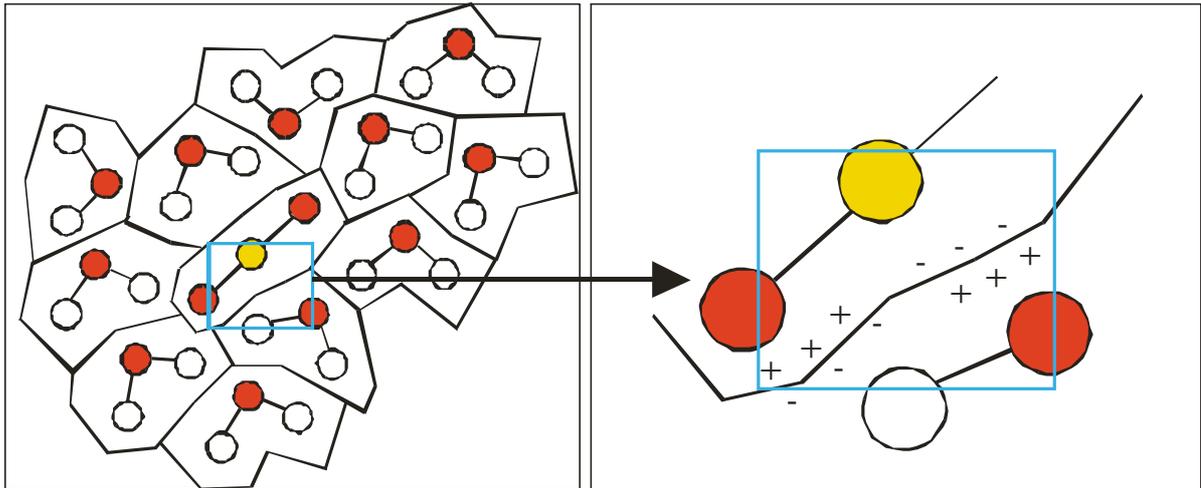


Abbildung 4: Ausschnitt aus einer Flüssigkeit, wie sie im COSMO-RS-Bild behandelt wird.

Nun muss die drastische Vereinfachung des idealen Leiters wieder ausgeglichen werden. Mit der Annahme, dass sich die Abschirmladungen und die Polarisierung des Moleküls im ideal abgeschirmten Zustand nicht ändern, ist es möglich, die Kontaktfläche in kleine Flächenstücke zu unterteilen.

2.2.2 Typische Kontaktsegmente und Misfitenergie

Diese Flächenstücke sollen die Gesamtladung s_{res} tragen, welche sich additiv aus den Ladungen sich berührender Oberflächensegmente zusammensetzen:

$$\mathbf{s}_{res} = \mathbf{s}_1 + \mathbf{s}_2$$

Gleichung 25

Die optimale Größe der Fläche a_{eff} der Segmente ist a priori unbekannt, ergibt sich aber später zwanglos bei der Optimierung.

Zwei vom Betrag gleich große Ladungsdichten gegensätzlich geladener Segmente kompensieren sich zu Null. In einer realen Lösung existieren allerdings nicht nur die Kombinationen exakt gegensätzlicher Ladungen, sondern auch eine große Anzahl von nicht idealen Paarungen. Es gilt die Energie (im weiteren als *Misfitenergie* bezeichnet) dieser nicht idealen Paarungen von Segmenten mit $s_{res} \neq 0$ zu quantifizieren.

Zwei Segmente mit den Oberflächenladungsdichten s_1 und s_2 bilden zusammen ein Segmentpaar mit dem effektiven Radius R_{eff} . Hierbei wird angenommen, dass die Ladung auf dem Segment homogen verteilt ist. Im COSMO Kapitel wurde eine Näherungsformel für die zweifache Selbstenergie (siehe Gleichung 11 Seite 2-5) eingeführt. Diese Formel wird nun für

die Misfitenergie verwendet und man erhält mit den Oberflächenladungsdichten und dem elektrostatischen Faktor $(4\pi\epsilon_0)^{-1}$ die Misfitenergie E_{misfit} :

$$E_{misfit} = \frac{1}{2} \mathbf{a} (\mathbf{s}_1 + \mathbf{s}_2)^2 \quad \text{Gleichung 26}$$

mit \mathbf{a} gleich:

$$\mathbf{a} = 1.2 \frac{\mathbf{p}^{\frac{5}{2}} R_{eff}^3}{4\pi\epsilon_0} \quad \text{Gleichung 27}$$

Die nicht verschwindenden Oberflächenladungen \mathbf{s}_{res} müssen mit jeweils gegensätzlichen Kompensationsladungen, die auf den jeweiligen Oberflächen positioniert werden, ausgeglichen werden, um wieder den Fall der idealen Abschirmung im elektrischen Leiter zu erhalten. Sind alle Ladungsdichten \mathbf{s}_{res} kompensiert, so ergibt sich die Gesamtenergie des Systems additiv aus:

- der Energie des ideal abgeschirmten Systems
- den Wechselwirkungsenergien der Kompensationsladungen mit dem ideal abgeschirmten System
- den Wechselwirkungsenergien der Kompensationsladungen
- der Summe der Eigenenergien der Kompensationsladungen

Im Folgenden soll angenommen werden, dass die Oberflächenladungen \mathbf{s}_{res} voneinander unabhängig und unkorreliert sind. Hieraus folgt ein Verschwinden der Wechselwirkungsenergien der Kompensationsladungen aufgrund der Vorzeichenwechsel der einzelnen Summanden. Per Definition ist auch das elektrostatische Potenzial des ideal abgeschirmten Systems gleich Null. Dieser Zustand wurde ja durch das Hinzufügen der Kompensationsladungen erreicht, und deshalb müssen auch die Wechselwirkungsanteile der Kompensationsladungen und dem ideal abgeschirmten System verschwinden.

Die Gesamtenergie eines realen fluiden Systems setzt sich also aus der Summe der idealen Abschirmenergien und der Summe der Eigenenergien der Kompensationsladungsdichten (also der Summe der Misfitenergien) der Komponente X in der Flüssigkeit zusammen:

$$E_{condensed}^{tot} = \sum_X E_{ideal}^X + \frac{\mathbf{a}'}{2} \sum_n (\mathbf{s}_{n1} + \mathbf{s}_{n2})^2 \quad \text{Gleichung 28}$$

Der Selbstenergieterm α' enthält einen Polarisationsfaktor f_{pol} , der die Antwort der Moleküle auf die veränderte elektrostatische Situation, welche durch Entfernung der Kompensationsladungen entsteht, berücksichtigen soll. Ein Molekül wird durch den Übergang in einen

günstigeren Polarisationszustand die Misfitenergie verringern. Für ein makroskopisches Kontinuum gilt die Beziehung:

$$f_{pol} = 1 - \frac{(n^2 - 1)}{(n^2 - \frac{1}{2})} \quad \text{Gleichung 29}$$

setzt man den Brechungsindex $n^2 = 2$, so ist $f_{pol} = 0.66$. Die Variable wird im weiteren Verlauf mit in die Misfit-Variable \mathbf{a} integriert:

$$\mathbf{a}' = \mathbf{a} \cdot f_{pol} \quad \text{Gleichung 30}$$

2.2.3 Ladungsdichten und σ -Profil

Bis zu diesem Stadium sind alle bisherigen Größen aus einer COSMO-Rechnung (mit $\epsilon = \infty$) erhältlich, abgesehen von a_{eff} und f_{pol} . Diese werden innerhalb der COSMO-RS-Theorie als zu optimierende Variablen betrachtet. Die im Misfitenergieterm verwendeten Ladungsdichten sind allerdings nicht die COSMO-Abschirmladungsdichten (\mathbf{s}^*), sondern werden nach folgendem Algorithmus über eine Kreisfläche mit Radius r_{av} gemittelt:

$$\mathbf{s}_n = \frac{\sum_m \mathbf{s}_m^* \frac{r_m^2 r_{av}^2}{r_m^2 + r_{av}^2} \exp\left\{-\frac{d_{mm}^2}{r_m^2 + r_{av}^2}\right\}}{\sum_m \frac{r_m^2 r_{av}^2}{r_m^2 + r_{av}^2} \exp\left\{-\frac{d_{mm}^2}{r_m^2 + r_{av}^2}\right\}} \quad \text{Gleichung 31}$$

Hier ist d_{mm} der Abstand der Segmente m und n , und r_m ist der mittlere Radius des Segments. Diese Prozedur wird durch die Forderung nach konstanten Ladungsdichten auf den effektiven Kontaktflächen notwendig.

Entkoppelte Segmente

An dieser Stelle wird darauf hingewiesen, dass innerhalb des COSMO-RS-Ansatzes alle möglichen Segmentpaare gebildet werden. Dies beinhaltet auch die Bildung von Paaren, die in der realen Flüssigkeit aufgrund sterischer Gegebenheiten nicht vorkommen können. Die COSMO-RS-Theorie wird hierdurch stark vereinfacht, deshalb birgt diese Näherung auch einen a priori nicht bekannten Fehler in sich.

***s*-Profil**

Ein für ein Molekül (eigentlich für ein bestimmtes Konformer eines Moleküls) charakteristisches Merkmal ist das σ -Profil $p^X(\mathbf{s})$. Hierbei handelt es sich um eine Häufigkeitsverteilung der mittleren Flächenladungsdichten. Diese werden entsprechend ihrer Fläche gaußverteilt zum \mathbf{s} -Profil aufaddiert. Ist die betrachtete Flüssigkeit eine Mischung mehrerer Komponenten, ergibt sich das normierte \mathbf{s} -Profil $p_S(\mathbf{s})$ dieser Mischung als gewichtete Summe der \mathbf{s} -Profile der einzelnen Komponenten analog ihrer Molenbrüche x_i (siehe nächste Gleichung).

$$p_S(\mathbf{s}) = \frac{\sum_i x_i p^{X_i}(\mathbf{s})}{\sum_i x_i}$$

Gleichung 32

Nachfolgende Abbildung zeigt beispielhaft die \mathbf{s} -Profile dreier Substanzen.

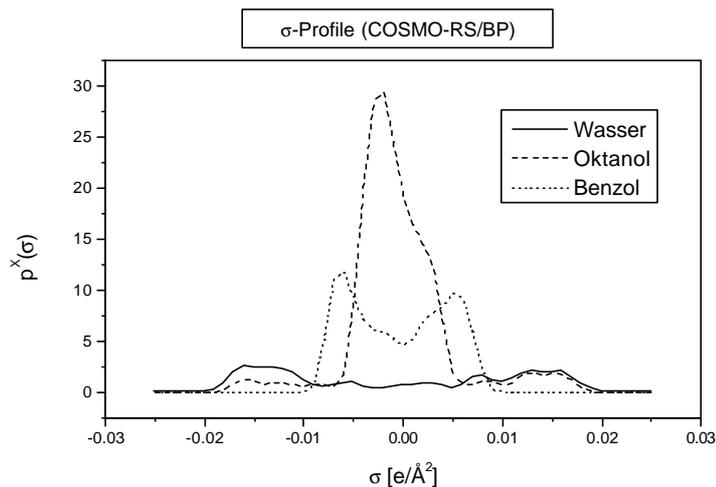


Abbildung 5: \mathbf{s} -Profile (= Häufigkeitsverteilung der Abschirmflächenladungsdichten auf der SAS) von Wasser, Oktanol und Benzol.

Nun soll das chemische Potenzial eines Stoffes in einer fluiden Phase quantitativ bestimmt werden. Hierzu wird mit Hilfe der COSMO-Abschirmladungsdichten, daraus berechneter Misfitenergien und mit Hilfe des \mathbf{s} -Profils eines Ensembles von Oberflächenladungsdichten ein Ausdruck für die freie Energie des fluiden Systems abgeleitet.

2.2.4 Freie Energie und chemisches Potenzial

Im Folgenden soll ein Ausdruck für das chemische Potenzial \mathbf{m} eines Systems ermittelt werden.

Hierzu wird die freie Energie F_s des Systems verwendet. Eine Beziehung zwischen F_s und dem chemischen Potenzial \mathbf{m} lässt sich aus den Zustandsgleichungen der Thermodynamik ableiten. So ist die freie Energie per Definition:

$$F = U - TS \quad \text{Gleichung 33}$$

mit der inneren Energie U , der Entropie S und der Temperatur T . Für eine infinitesimale Änderung der freien Energie gilt:

$$dF = dU - TdS - SdT \quad \text{Gleichung 34}$$

mit der Beziehung für die Änderung der inneren Energie, abhängig von der Teilchenzahl N :

$$dU = TdS - pdV + \mathbf{m}dN \quad \text{Gleichung 35}$$

erhält man:

$$dF = -SdT - pdV + \mathbf{m}dN \quad \text{Gleichung 36}$$

Auf der andern Seite kann auch das totale Differential der freien Energie formuliert werden, da bekannt ist, dass F eine Zustandsfunktion von T , V und N ist. Also:

$$dF = \left(\frac{\partial F}{\partial T} \right)_{V,N} dT + \left(\frac{\partial F}{\partial V} \right)_{T,N} dV + \left(\frac{\partial F}{\partial N} \right)_{V,T} dN \quad \text{Gleichung 37}$$

Ein Koeffizientenvergleich von Gleichung 36 und Gleichung 37 liefert die gesuchte Beziehung für das chemische Potenzial:

$$\mathbf{m} = \left(\frac{\partial F}{\partial N} \right)_{T,V} \quad \text{Gleichung 38}$$

Der statistischen Thermodynamik entspringt eine weitere grundlegende Gleichung für die freie Energie F_s eines Ensembles (in diesem Fall eines Ensembles entkoppelter Segmente):

$$F_s = -kT \ln Z_s \quad \text{Gleichung 39}$$

Hier ist Z_s die Zustandssumme des Ensembles. Wird Gleichung 39 in Gleichung 38 eingesetzt, so erhält man für das chemische Potenzial bei gegebenen Ladungsdichten \mathbf{s} :

$$\mathbf{m}(\mathbf{s}) = -kT \frac{d \ln Z_s}{dN(\mathbf{s})} \quad \text{Gleichung 40}$$

Im Falle einer großen Anzahl von Teilchen (Segmenten) ist der Logarithmus der Zustandssumme eines um zwei Teilchen i und j verkleinerten Systems gleich dem Logarithmus der ursprünglichen Zustandssumme addiert zu den chemischen Potenzialen der Segmente i und j , denn die Änderung der logarithmischen Zustandssumme lautet:

$$\frac{d \ln Z}{dN} = -\frac{\mathbf{m}}{kT} \quad \text{Gleichung 41}$$

Das negative Vorzeichen veranschaulicht per Definition eine Vergrößerung des Systems, so dass für das verkleinerte System gilt:

$$\ln Z_{ij} = \ln Z_s + \frac{\mathbf{m}(\mathbf{s}_i)}{kT} + \frac{\mathbf{m}(\mathbf{s}_j)}{kT} \quad \text{Gleichung 42}$$

oder:

$$Z_{ij} = Z_s e^{\frac{\mathbf{m}(\mathbf{s}_i) + \mathbf{m}(\mathbf{s}_j)}{kT}} \quad \text{Gleichung 43}$$

Die Entartung E_N eines Ensembles mit N Segmenten, welche paarweise kombiniert werden ist gleich:

$$E_N = 2^{\frac{N}{2}} \left(\frac{N}{2} \right)! \quad \text{Gleichung 44}$$

Die Gesamtanzahl aller möglichen Permutationen von N Segmenten ist $N!$. Da es für die gebildeten Paare aber keine Rolle spielt, an welchem Ort diese gepaart werden, stehen für ein Paar (i,j) nur $N/2$ Plätze zur Verfügung, für ein weiteres Paar nur noch $(N/2)-1$ usw. Für das gesamte Ensemble ist die Entartung also $(N/2)!$. Darüber hinaus muss noch beachtet werden, dass die Paarungen (i,j) und (j,i) die gleichen Misfitenergien liefern, und dies entspricht einem Entartungsfaktor von zwei. Wiederum für alle $N/2$ Paare erhält man $2^{N/2}$ entartete Zustände. Werden die beiden Beiträge multipliziert, erhält man den in Gleichung 44 angegebenen Entartungsgrad. Bei dem um zwei Segmente verkleinerten System ist die Entartung E_{N-2} um einen Faktor N kleiner.

$p(0,i)$ sei nun die Wahrscheinlichkeit ein bestimmtes Paar $(0,i)$ von Segmenten im Originalensemble anzutreffen. Diese ist dann gleich dem Produkt aus:

- der Vergrößerung der Entartung,
- der Boltzmann-gewichteten Misfitenergie des Paares $(0,i)$,

- und der Zustandssumme des um die Segmente i und j verkleinerten Systems

also:

$$p(0, i) = Ne^{-\frac{a'(s_0+s_i)^2}{2kT}} Z_s e^{-\frac{m(s_0)+m(s_i)}{kT}} \quad \text{Gleichung 45}$$

Die Überlegung, dass im ursprünglichen Ensemble das Segment 0 immer mit einem der $N-1$ übrigen Segmente gepaart vorliegt, führt zur Erkenntnis, dass sich die Zustandssumme der Originalgesamtheit als Summe der Wahrscheinlichkeiten ein Segment 0 in einem Paar $(0, i)$ zu finden, ergibt:

$$Z_s = \sum_{i=1}^{N-1} p(0, i) = \sum_{i=1}^{N-1} Ne^{-\frac{a'(s_0+s_i)^2}{2kT}} Z_s e^{-\frac{m(s_0)+m(s_i)}{kT}} \quad \text{Gleichung 46}$$

Aus dieser Gleichung kann nun die Zustandssumme gekürzt werden. Logarithmiert man anschließend und bringt das chemische Potenzial $m(s_0)$ auf die linke Seite, so erhält man:

$$m(s_0) = -kT \ln \left[N \sum_{i=1}^{N-1} e^{-\frac{a'(s_0+s_i)^2}{2kT} + \frac{m(s_i)}{kT}} \right] \quad \text{Gleichung 47}$$

Jetzt soll die Summation nicht mehr von eins bis $N-1$, sondern (in erster Näherung für eine große Anzahl von Segmenten) von eins bis N erfolgen. Dann kann die Summe ohne Probleme durch das früher eingeführte \mathbf{s} -Profil multipliziert mit N ersetzt werden. Werden \mathbf{s}_0 und \mathbf{s}_1 noch in \mathbf{s} und \mathbf{s}' umbenannt, gilt:

$$m(\mathbf{s}) = -kT \ln \left[N^2 \int p_s(\mathbf{s}') e^{-\frac{a'(\mathbf{s}+\mathbf{s}')^2}{2kT} + \frac{m_s(\mathbf{s}')}{kT}} d\mathbf{s}' \right] \quad \text{Gleichung 48}$$

Wird anstelle des von der Teilchenzahl N abhängigen Potentials ein normalisiertes Potenzial $m'_s(\mathbf{s})$ eines Ensembles mit einem mol Segmenten eingeführt:

$$m'_s(\mathbf{s}) = m_s(\mathbf{s}) + kT \ln N_{mol} \quad \text{Gleichung 49}$$

so ergibt sich mit Gleichung 48 ein Ausdruck für das chemische Potenzial eines Segmentes in einem mol des Segmentensembles.

$$m'_s(\mathbf{s}) = -kT \ln \left[\int p'_s(\mathbf{s}') e^{-\frac{a'(\mathbf{s}+\mathbf{s}')^2}{2kT} + \frac{m'_s(\mathbf{s}')}{kT}} d\mathbf{s}' \right] \quad \text{Gleichung 50}$$

In dieser Gleichung ist $p'_s(\mathbf{s}')$ das auf die Oberfläche normierte \mathbf{s} -Profil. Gleichung 50 muss iterativ gelöst werden, da das Potenzial der einzelnen Segmente augenscheinlich von den

Potenzialen der anderen abhängt. Hierzu geht man von einem Segmentpotenzial gleich Null aus und iteriert bis zur Selbstkonsistenz. Die Funktion $\mathbf{m}'_s(\mathbf{s})$ wird in der COSMO-RS-Nomenklatur \mathbf{s} -Potenzial genannt. Wird das chemische Potenzial eines Segments pro Flächenstück in Einheiten von kT angegeben: $\mathbf{m}_s(\mathbf{s}) = \mathbf{b}^{-1} \mathbf{m}'_s(\mathbf{s})$ mit $\mathbf{b} = kT / a_{\text{eff}}$ und der Misfitenergieterm durch den allgemeinen Ausdruck $E(\mathbf{s}, \mathbf{s}')$ ersetzt, gilt:

$$\mathbf{m}_s(\mathbf{s}) = \mathbf{m}'_s(\mathbf{s}) \frac{a_{\text{eff}}}{kT} = -\ln \left[\int d\mathbf{s}' p'_s(\mathbf{s}') e^{-E(\mathbf{s}, \mathbf{s}') + \mathbf{m}_s(\mathbf{s}')} \right] \quad \text{Gleichung 51}$$

Von wirklichem Interesse ist aber nicht das chemische Potenzial eines Segments, sondern das chemische Potenzial des Moleküls X im Lösungsmittel S. Dieses lässt sich nun zwanglos mit Hilfe des \mathbf{s} -Potenzials des Lösungsmittels und dem \mathbf{s} -Profil des gelösten Moleküls formulieren:

$$\mathbf{m}_s^{*X} = \int d\mathbf{s} p^X(\mathbf{s}) \mathbf{m}'_s(\mathbf{s}) - I kT \ln A^S = \mathbf{b} \mathbf{m}_s^X - I kT \ln A^S \quad \text{Gleichung 52}$$

Das Produkt im Integral veranschaulicht die Menge der freien Energie $\mathbf{m}'(\mathbf{s})$, die $p^X(\mathbf{s})$ Segmente mit der Flächenladungsdichte \mathbf{s} beim Lösen aufzubringen haben. Der entscheidende Vorteil dieser Gleichung ist, dass das chemische Potenzial nicht nur in reinen Lösungsmitteln, sondern auch in beliebigen Mischungen berechnet werden kann. Neu eingeführt wurde hier der kombinatorische Faktor $I kT \ln A^S$ ($A^S =$ Moleküloberfläche des Lösungsmittels S), der den Einfluss der Lösungsmittelmolekülgröße auf das chemische Potenzial quantifiziert. I ist ein zu optimierender COSMO-RS-spezifischer Parameter.

Das \mathbf{s} -Potenzial eines Stoffes in einer frei wählbaren Umgebung erweist sich als außerordentlich nützlich, ist es doch der Schlüssel zu einer Reihe thermodynamischer Eigenschaften von Flüssigkeiten.

2.2.5 Erweiterung der COSMO-RS-Theorie

In einer weiteren Entwicklungsstufe wurde ein Wasserstoffbrückenbindungsterm in die COSMO-RS-Theorie eingeführt. Zwar werden elektrostatische Anteile der Wasserstoffbrückenbindungsenergie durch die Misfitenergien sehr wohl berücksichtigt, die gegenseitige Durchdringung der Elektronendichten von Donor und Akzeptor wird dagegen nicht wiedergegeben. Deshalb wurde der Energieoperator, der bisher allein aus der Summe der Misfitenergien bestand, um einen Wasserstoffbrückenbindungsterm erweitert. Mit diesem

werden zwei neue zu optimierende Parameter \mathbf{s}_{hb} und c_{hb} eingeführt. Mit deren Hilfe ist es möglich, das Wasserstoffbrückenbindungsphänomen besser zu beschreiben. Je stärker polar die Bindungspartner, desto stärker die Brückenbindung. Unterhalb einer Schranke ist keine Bindungsstabilisierung zu beobachten. Durch folgende Gleichung werden beide Anforderungen erfüllt:

$$E_{hb}(\mathbf{s}, \mathbf{s}') = c_{hb} \max[0, \mathbf{s}_{acc} - \mathbf{s}_{hb}] \cdot \min[0, \mathbf{s}_{don} + \mathbf{s}_{hb}] \quad \text{Gleichung 53}$$

Ein Wasserstoffbrückenbindungsbeitrag zweier Segmente kommt also nur dann zustande, wenn beide Segmente gegensätzlich geladen sind und wenn die Beträge ihrer Ladungsdichten die Schranke \mathbf{s}_{hb} übersteigen. Sind diese Bedingungen erfüllt, ergibt sich der Energiebeitrag aus dem Produkt der Ladungsdichten mit dem Parameter c_{hb} .

Bis jetzt sind die gemittelten Oberflächenladungsdichten die einzigen Parameter, mit deren Hilfe die chemischen Potentiale berechnet werden. Innerhalb der COSMO-RS-Theorie besteht zwischen der dielektrischen Abschirmenergie E_{diel}^X :

$$E_{diel}^X = \frac{1}{2} \sum_n \Phi_n^X s_n \mathbf{s}_n \quad \text{Gleichung 54}$$

(mit dem chemischen Potenzial Φ , der Fläche s und der Ladungsdichte \mathbf{s} des Segments v) und dem Energiegewinn durch die Abschirmung ($\Delta^X = E_{gas} - E_{COSMO}$) ein streng linearer Zusammenhang ($\Delta^X = 0.8 E_{diel}^X$). Allerdings zeigen sich systematische Abweichungen. So wird Δ^X bei Molekülen mit großen polaren Regionen auf der SAS (z. B. bei Carbonylen oder bei Nitrilen) überschätzt. Grund hierfür sind Korrelationseffekte der Ladungsdichten auf den relativ großen Oberflächen des sp^2 -hybridisierten Sauerstoffs bzw. des sp -hybridisierten Stickstoffs. Um dieser starken Korrelation der Ladungsdichten zwischen Segmenten auf der SAS abzuweichen, werden neue Ladungsdichten \mathbf{s}_n° eingeführt. Diese werden wiederum nach Gleichung 31 (Seite 2-12) durch Mittelung über eine Fläche mit doppeltem Radius ($2*r_{av}$), aus den COSMO-Abschirmladungsdichten berechnet. Die neuen Abschirmladungsdichten sind nach wie vor mit den Abschirmladungsdichten \mathbf{s}_n korreliert, aber letztendlich lassen sich durch Orthogonalisierung der Dichten nach:

$$\mathbf{s}_n^\perp = \mathbf{s}_n^\circ - 0.81\mathbf{s}_n \quad \text{Gleichung 55}$$

neue Deskriptoren \mathbf{s}^\perp berechnen, die die lokale Korrelation eines Segments quantifizieren. Werden diese Ladungsdichten in die Misfitenergie integriert, führt dies zu:

$$E_{\text{misfit}}((\mathbf{s}, \mathbf{s}^\perp), (\mathbf{s}', \mathbf{s}'^\perp)) = \frac{\mathbf{a}'}{2} (\mathbf{s}, \mathbf{s}^\perp) [(\mathbf{s} + \mathbf{s}') + f_{\text{corr}}(\mathbf{s}^\perp + \mathbf{s}'^\perp)] \quad \text{Gleichung 56}$$

Durch die Erweiterung der Variablen muss das Integral in Gleichung 51 (Seite 2-17) durch eine gewichtete Summe ersetzt werden. Es gilt dann:

$$\mathbf{m}_s(d) = -\ln \left[W^{-1} \sum_i x_i \sum_{n \in X_i} s_n^i e^{-b^{-1}E(d, d_n^i) + \mathbf{m}_s(d_n^i)} \right] \quad \text{Gleichung 57}$$

Hier sind s^i und d^i Fläche und Deskriptor des Segments n der i -ten Komponente des Systems, x_i ist der Molenbruch und W ist der Normierungsfaktor der seinerseits als

$$W = -\sum_i x_i \sum_{n \in X_i} s_n^i \quad \text{Gleichung 58}$$

definiert ist.

Mit Hilfe dieses neuen Ausdrucks für das \mathbf{s} -Potenzial kann nun nach Gleichung 52 (Seite 2-17) das chemische Potenzial einer Komponente in einer fluiden Phase berechnet werden.

2.2.6 Chemisches Potenzial in der Gasphase

Zur Beschreibung von physikochemischen Verteilungsgrößen zwischen der Gasphase und einer fluiden Phase ist es notwendig, das Gasphasenpotenzial des Stoffes zu bestimmen. Gleichung 59 definiert das chemische Potenzial einer Substanz in der Gasphase relativ zum ideal abgeschirmten Zustand bei einem Partialdruck von einem bar.

$$\mathbf{m}_{\text{gas}}^X = -\Delta^X - \sum_k \mathbf{g}_k A_k^X - \mathbf{w} n_{ra}^X - \mathbf{h} RT \quad \text{Gleichung 59}$$

Der zweite Term auf der rechten Seite repräsentiert den Dispersionsenergiegewinn beim Transfer eines Moleküls von der Gasphase in die Lösung. Dieser ist von elementspezifischen Parametern \mathbf{g}_k abhängig, die mit den jeweiligen Oberflächenanteilen A_k multipliziert werden. Der dritte Term in Gleichung 59 ist ein Korrekturfaktor für Ringverbindungen. Der zu optimierende Parameter \mathbf{w} wird mit der Anzahl der Ringatome n_{ra} der Substanz multipliziert. Der letzte Term mit dem zu optimierenden Parameter \mathbf{h} berücksichtigt die Entropie des Moleküls in der Gasphase und sorgt daneben für die richtige Beziehung zum Referenzzustand der Gasphase.

Δ^X ist die ideale Abschirmenergie, dies ist die Differenz aus Gasphasen- und COSMO-Energie (bei $\epsilon = \infty$) plus der Differenz der dielektrischen Energien des ideal abgeschirmten

Systems (E_{diel}^X) und dem mit gemittelten Abschirmladungsdichten (vgl. Gleichung 31 Seite 2-12) berechneten ideal abgeschirmten Zustand (E_{diel}^X):

$$\Delta^X = E_{gas}^X - E_{COSMO}^X + 0.4 \underbrace{\sum_k s_n \Phi_n^X s_n}_{2E_{diel}^X} - 0.4 \underbrace{\sum_k s_n \Phi_n^X s_n^*}_{2E_{diel}^X} \quad \text{Gleichung 60}$$

Mit dem Gasphasenpotenzial ist die COSMO-RS-Theorie nun hinreichend erklärt. Im nächsten Abschnitt sollen die Definitionen der im experimentellen Teil verwendeten physikochemischen Verteilungsgrößen erläutert werden.

2.3 Physikochemische Stoffeigenschaften

Die Parametrisierung der COSMO-RS/AM1-Methode erfolgt durch Anpassung an experimentell gemessene Daten. Aus der Vielzahl von möglichen Zielgrößen wurden drei Stoffeigenschaften ausgewählt: die Henry-Konstante, der Oktanol/Wasser-Verteilungskoeffizient und der Dampfdruck. Nachfolgend wird beschrieben, wie man aus den mit COSMO-RS ermittelten Potenzialen die verschiedenen Verteilungsgrößen berechnen kann. Besonderes Augenmerk wird hierbei auf die Einheiten und Referenzsysteme gelegt, da hier oftmals Ungenauigkeiten in der Literatur gefunden werden.

Prinzipiell lassen sich Gleichgewichtskonstanten K mit den zugehörigen freien Gleichgewichtsenthalpien ΔG berechnen:

$$\Delta G = -RT \ln K \quad \text{Gleichung 61}$$

Im den nächsten Kapiteln werden mit dieser Gleichung die verwendeten Zielgrößen abgeleitet.

2.3.1 Die Henry-Konstante

Die Henry-Konstante beschreibt das Verteilungsgleichgewicht eines Stoffes bei konstantem Druck und konstanter Temperatur zwischen der wässrigen Phase und der Gasphase. In der dimensionslosen Form ist die Henry-Konstante H' gleich dem Reziprokwert der Konzentrationen des Stoffes in der Gasphase c_g und der aquatischen Phase c_w :

$$K_{g \rightarrow w} = \frac{c_w}{c_g} = H' \quad \text{Gleichung 62}$$

Nach Gleichung 61 ergibt sich für die Änderung der freien Enthalpie des Phasenübergangs:

$$\Delta G_{g \rightarrow w} = -RT \ln K_{g \rightarrow w} = RT \ln H' \quad \text{Gleichung 63}$$

$\Delta G_{g \rightarrow w}$ ist die freie Hydratationsenthalpie und wird im Folgenden als ΔG_{hydr} bezeichnet. Neben der dimensionslosen Henry-Konstante (die in der Literatur auch oft als K_{aw} bezeichnet wird) existieren noch weitere Formulierungen der Henry-Konstante. Oft wird die dimensionsbehaftete Henry-Konstante H angegeben:

$$H = H' \cdot RT = \frac{P_v}{S_w} \quad \text{Gleichung 64}$$

H ist als Quotient aus Dampfdruck P_v und Wasserlöslichkeit S_w definiert. (Physikalisch richtig sollte anstelle des Dampfdrucks der Partialdruck P_c geschrieben werden, für Feststoffe und Flüssigkeiten gilt jedoch $P_v \approx P_c$). Die dimensionsbehaftete Henry-Konstante wird oft in Einheiten von $[\text{Pa L (Mol)}^{-1}]$ angegeben. Nachfolgende Gleichung veranschaulicht die Beziehung zwischen der freien Hydratationsenthalpie und der Henry-Konstante:

$$\Delta G_{\text{hydr}} = RT \ln H - RT \ln RT \quad \text{Gleichung 65}$$

Die logarithmische Henry-Konstante kann auch als Funktion von ΔG_{hydr} ausgedrückt werden:

$$\log H = \frac{\Delta G_{\text{hydr}}}{2.3RT} + \log RT \quad \text{Gleichung 66}$$

Die freie Hydratationsenthalpie lässt sich direkt aus den chemischen Potenzialen m_x des Stoffes in den Phasen X berechnen:

$$\Delta G_{\text{hydr}} = m_{\text{H}_2\text{O}} - m_{\text{gas}} \quad \text{Gleichung 67}$$

Wird ΔG_{hydr} nun aus den COSMO-RS-Potenzialen berechnet, müssen die unterschiedlichen Bezugszustände beachtet werden. Das Potenzial in der Gasphase (Gleichung 59 Seite 2-19) bezieht sich auf den Referenzzustand 1 bar, und das chemische Potenzial in der flüssigen Phase (Gleichung 52 Seite 2-17) bezieht sich auf 1 mol der betrachteten Spezies. Es ist also notwendig, bei der Angabe von freien Energien immer den jeweiligen Bezugzustand anzugeben. Hieraus folgt, dass auch die Gleichgewichtskonstanten je nach Referenzzustand unterschiedliche Werte annehmen.

Um das chemische Potenzial einer Substanz in der Gasphase vom Bezugszustand bar nach mol/L umzurechnen, muss wie folgt vorgegangen werden. Grundlage für die Umrechnung bildet die Gibbs-Duhemsche Gleichung:

$$\Delta G = -SdT + Vdp + \sum_i m_i dn_i \quad \text{Gleichung 68}$$

Bei konstanter Temperatur und konstanter Stoffmenge gilt für die Änderung der freien Enthalpie:

$$\Delta G = Vdp \quad \text{Gleichung 69}$$

Unter Annahme der Gültigkeit des idealen Gasgesetzes ergibt sich das molare Volumen eines Gases bei 1 bar zu 22.46 L/mol (bei Raumtemperatur). Dieses Volumen soll nun formal auf 1L/mol komprimiert werden. Mit dem allgemeinen Gasgesetz:

$$p = \frac{1}{V} RT \quad \text{Gleichung 70}$$

und der Änderung des Druckes in Abhängigkeit vom Volumen:

$$dp = -\frac{RT}{V^2} dV \quad \text{Gleichung 71}$$

kann ein Ausdruck für Vdp angegeben werden:

$$Vdp = -\frac{RT}{V} dV = -RTd(\ln V) \quad \text{Gleichung 72}$$

Setzt man die linke Seite von Gleichung 69 in Gleichung 72 ein und integriert von $V_M = 22.46$ L/mol bis $V_M = 1.0$ L/mol, ergibt sich:

$$\Delta G(\text{bar} \rightarrow \text{molL}^{-1}) = -RT(\ln 1 - \ln 22.46) = RT \ln(22.46) \quad \text{Gleichung 73}$$

Analog zur Umrechnung für die Gasphase kann für das chemische Potenzial in der flüssigen Phase verfahren werden. Entspricht der Gasphasenreferenzzustand 1 bar 22.46 L/mol, so ist nun das chemische Potenzial von 1 mol Substanz bezogen auf 1 mol Lösungsmittel umzurechnen auf 1 mol des Stoffes pro Liter flüssigen Mediums. Da bei unendlichen Verdünnungen gerechnet wird, wird das Volumen des gelösten Stoffes vernachlässigt, und das molare Volumen der Lösung ist:

$$\frac{MG_{H_2O}}{d_{H_2O}} = \frac{18 \text{ g mol}^{-1}}{1000 \text{ g L}^{-1}} = 0.018 \text{ L mol}^{-1} \quad \text{Gleichung 74}$$

Formal muss eine Expansion von $V_M = 0.018 \text{ Lmol}^{-1}$ auf $V_M = 1.0 \text{ Lmol}^{-1}$ durchgeführt werden. Analog zu Gleichung 73 ist ΔG :

$$\Delta G^{H_2O} (\text{mol} \rightarrow \text{molL}^{-1}) = -RT(\ln 1 - \ln 0.018) = RT \ln 0.018 \quad \text{Gleichung 75}$$

Die freie Hydratationsenthalpie wird aus den COSMO-RS-Potenzialen und den Korrekturenthalpien ΔG^X berechnet:

$$\Delta G_{hydr} = \underbrace{m_{H_2O}^{COSMO-RS} - m_{gas}^{COSMO-RS}}_{\Delta G_{hydr}^{COSMO-RS}} - \underbrace{(\Delta G^{gas} + \Delta G^{H_2O})}_{4.26 \text{ kcal/mol}} \quad \text{Gleichung 76}$$

Bei Raumtemperatur müssen -4.26 kcal/mol zu $\Delta G_{hydr}^{COSMO-RS}$ addiert werden, um den üblichen Referenzzuständen gerecht zu werden. Für die mit ΔG_{hydr} korrespondierende logarithmische Henry-Konstante ($\log H'$) gilt also:

$$\log H' = \frac{m_{H_2O} - m_{gas}}{2.3RT} - \log(RT) - \log(0.018) \quad \text{Gleichung 77}$$

Die zweite Stoffeigenschaft, mit Hilfe derer das COSMO-RS/AM1-Verfahren angepasst wird, ist der Dampfdruck.

2.3.2 Der Dampfdruck

Hat sich für einen Reinstoff ein Gleichgewicht zwischen der Gasphase und der flüssigen (oder festen) Phase eingestellt, so ist die Gleichgewichtskonstante $K_{c@g}$ definiert als:

$$K_{c \rightarrow g} = \frac{c_g}{c_c} \quad \text{Gleichung 78}$$

mit den Konzentrationen des Stoffes in der Gasphase c_g und in der fluiden Phase c_c . Die freie Enthalpie des Phasenübergangs ist wiederum gegeben durch:

$$\Delta G_{c \rightarrow g} = -RT \ln K_{c \rightarrow g} = -RT(\ln c_g - \ln c_c) \quad \text{Gleichung 79}$$

Mit der Definitionsgleichung für den Dampfdruck:

$$P_v = c_g RT \quad \text{Gleichung 80}$$

kann eine Beziehung zwischen dieser Größe und der Verdampfungsenthalpie formuliert werden:

$$\log P_v = \frac{\Delta G_{c \rightarrow g}}{2.3RT} + \log c_c + \log RT \quad \text{Gleichung 81}$$

$\Delta G_{c \rightarrow g}$ wird als Differenz der chemischen Potentiale des Stoffes in der Gasphase und des Stoffes in der reinen Komponente berechnet. Hier offenbart sich eine Schwierigkeit bei der Berechnung des Dampfdrucks. Da das COSMO-RS-Modell eine Theorie für flüssige Zustände ist, wird bei der Berechnung des chemischen Potentials eines Stoffes davon ausgegangen, dass es sich um eine Flüssigkeit handelt. Wird COSMO-RS zur Berechnung der Dampfdrücke von Feststoffen verwendet - dies ist formal ja ohne Problem möglich – sind die Resultate fehlerhaft. Eine COSMO-RS-Dampfdruckberechnung darf also nur für Substanzen erfolgen, die im zu untersuchenden Temperaturbereich als Reinstoff flüssig sind. Für Feststoffe müsste gegebenenfalls die von COSMO-RS nicht erfasste freie Schmelzenthalpie zur Differenz der chemischen Potentiale addiert werden.

Eine weitere Stoffeigenschaft, die zur Parametrisierung der COSMO-RS/AM1-Methode herangezogen wird und bei der die erwähnten Probleme mit Feststoffen nicht auftreten, ist der Oktanol/Wasser-Verteilungskoeffizient.

2.3.3 Der Oktanol/Wasser-Verteilungskoeffizient

Die Verteilung eines Stoffes zwischen einer wässrigen Phase und einer Oktanolphase kann wiederum mit den molaren Konzentrationen c_w und c_o des Stoffes in den beiden Phasen beschrieben werden.

$$\frac{c_w}{c_o} = K_{o \rightarrow w} = \frac{1}{K_{ow}} \quad \text{Gleichung 82}$$

Gleichung 82 zeigt, dass der Oktanol/Wasser-Verteilungskoeffizient K_{OW} definitionsgemäß der Gleichgewichtskonstante des Phasenübergangs eines Stoffes *von Wasser nach* Oktanol entspricht. Wie schon bei den vorangegangenen physikochemischen Größen, lässt sich ein Zusammenhang zwischen der freien Enthalpie des Phasenübergangs und dem Oktanol/Wasser-Verteilungskoeffizient formulieren:

$$\Delta G_{o \rightarrow w} = -RT \ln K_{o \rightarrow w} = RT \ln K_{ow} \quad \text{Gleichung 83}$$

Werden in die Gleichung wieder die COSMO-RS-Potentiale eingesetzt, ergibt sich die Gleichgewichtskonstante als Quotient der Molenbrüche:

$$\ln K_{o \rightarrow w}^x = \ln \left(\frac{x_o}{x_w} \right) = \frac{m_w^{\text{COSMO-RS}} - m_o^{\text{COSMO-RS}}}{RT} \quad \text{Gleichung 84}$$

Nach einfacher Umformung und mit der Definition für den Molenbruch:

$$x_a = c_a \cdot v_a \quad \text{Gleichung 85}$$

mit c_a : Konzentration des Stoffs in a

und v_a : molares Volumen

ergibt Gleichung 84:

$$\log K_{ow}^c = \log \left(\frac{c_o}{c_w} \right) = \log \left[\frac{v_w}{v_o} e^{\frac{m_w^{\text{COSMO-RS}} - m_o^{\text{COSMO-RS}}}{RT}} \right] \quad \text{Gleichung 86}$$

Der Quotient der molaren Volumina kann mit den Molgewichten angenähert werden. Im Falle sehr unterschiedlicher Dichten r sollte der Quotient allerdings mit den Molgewichten MG und den Dichten nach:

$$\frac{v_w}{v_o} = \frac{r_o MG_w}{MG_o r_w} \quad \text{Gleichung 87}$$

berechnet werden.

Im Realfall muss beim Oktanol/Wasser-Verteilungskoeffizient beachtet werden, dass Wasser eine nicht verschwindende Löslichkeit in Oktanol aufweist. Experimentelle Bestimmungen des K_{OW} erfolgen meist durch Methoden, bei denen der zu untersuchende Stoff in ein Zweiphasensystem aus Wasser und Oktanol eingebracht wird. Nach guter Durchmischung und anschließender Separation ist aber ein gewisser Anteil Wasser in der Oktanolphase gelöst. Experimentell ermittelt wird also der Verteilungskoeffizient eines Stoffes zwischen der Wasserphase und der mit Wasser gesättigten Oktanolphase. Natürlich beeinflusst die Löslichkeit des Stoffes wiederum die Löslichkeit von Wasser in Oktanol. Dieser Effekt wird allerdings vernachlässigt und die Berechnung des Oktanol/Wasser-Verteilungskoeffizienten erfolgt zwischen der reinen Wasserphase und einer Oktanolphase, die aus 74 mol % Oktanol und 26 mol % Wasser besteht [35]. Dieses Verhältnis entspricht dem realen Wassergehalt der Oktanolphase des Zweikomponentensystems bei Raumtemperatur (25 °C) im Gleichgewicht. Prinzipiell können auch beliebige weitere Gleichgewichtskonstanten mit COSMO-RS berechnet werden. Als Beispiel sei hier noch die Wasserlöslichkeit eines Stoffes erwähnt.

2.3.4 Die Wasserlöslichkeit

Wird eine Substanz C in eine wässrige Phase eingebracht, kann nach Gleichgewichtseinstellung wieder eine Gleichgewichtskonstante K als Quotient der Konzentrationen des Stoffes formuliert werden:

$$\frac{c_w}{c_c} = K_{c \rightarrow w} \quad \text{Gleichung 88}$$

Mit der Beziehung zwischen der freien Enthalpie und der Gleichgewichtskonstante:

$$\Delta G_{c \rightarrow w} = -RT \ln K_{c \rightarrow w} = RT(\ln c_w - \ln c_c) \quad \text{Gleichung 89}$$

kann ein Ausdruck für die Wasserlöslichkeit S_w (diese ist gleich der Konzentration in der wässrigen Phase c_w) angegeben werden:

$$\log \frac{S_w}{c_c} = \frac{\Delta G_{c \rightarrow w}}{2.3RT} \quad \text{Gleichung 90}$$

Nach dem Logarithmus der Wasserlöslichkeit aufgelöst erhält man:

$$\log S_w = \frac{\Delta G_{c \rightarrow w}}{2.3RT} + \log c_c \quad \text{Gleichung 91}$$

Die mit dem Lösungsvorgang assoziierte Änderung der freien Lösungsenthalpie $\Delta G_{c \rightarrow w}$ wird nun wiederum durch die Differenz der chemischen Potentiale ausgedrückt:

$$\log S_w = -\frac{\mathbf{m}_{H_2O} - \mathbf{m}_c}{2.3RT} + \log c_c \quad \text{Gleichung 92}$$

Analog zu den im Kapitel 2.3.1 für die Henry-Konstante berücksichtigten speziellen COSMO-RS-Referenzzuständen müssen auch jetzt freie Enthalpiebeiträge addiert werden, um vom Bezugssystem [mol/mol] auf [mol/L] umzurechnen. Nachfolgende Gleichung zeigt die Beziehung zwischen der Wasserlöslichkeit in mol/L und den mit COSMO-RS bestimmten Potentialen plus Enthalpiebeiträgen ΔG^X (mit $X = H_2O$ oder $X = C$) zur Einheitenkonversion.

$$\log \frac{S_w}{c_c} = -\frac{1}{2.3RT} (\mathbf{m}_{H_2O}^{COSMO-RS} + \Delta G^{H_2O} - \mathbf{m}_C^{COSMO-RS} - \Delta G^C) \quad \text{Gleichung 93}$$

Mit den Korrekturen für die Referenzzustandsumrechnung der einzelnen Komponenten:

$$\Delta G^X (mol \rightarrow molL^{-1}) = RT \ln \left(\frac{MG^X}{1000 r^X} \right) \quad \text{Gleichung 94}$$

(mit dem Molgewicht $MG^{H_2O} = 18$ g/mol und der Dichte r^{H_2O} von Wasser) erhält man schließlich:

$$\log \frac{S_w}{c_c} = \frac{\mathbf{m}_c^{COSMO-RS} - \mathbf{m}_{H_2O}^{COSMO-RS}}{2.3RT} - \log 0.018 + \log \frac{MG^c}{1000 r^c} \quad \text{Gleichung 95}$$

Die Konzentration des Stoffes in der Reinstoffphase ist:

$$c_c = \frac{1000 r^c}{MG^c} \quad \text{Gleichung 96}$$

Setzt man Gleichung 96 in Gleichung 95 ein, ergibt sich der gewünschte Zusammenhang zwischen der Wasserlöslichkeit [mol/L] und den COSMO-RS-Potenzialen:

$$\log S_w = \frac{\mathbf{m}_c - \mathbf{m}_{H_2O}}{2.3RT} - \log 0.018 \quad \text{Gleichung 97}$$

Nun sind alle zum Verständnis der durchgeführten Berechnungen notwendigen theoretischen Grundlagen hinreichend erklärt. Im nächsten Kapitel werden die Optimierung der COSMO-RS-Parameter auf AM1-Niveau, die berechneten physikochemischen Stoffeigenschaften und die Moleküldatensätze vorgestellt und mit alternativen Rechnungsmodellen verglichen.

3 COSMO-RS/AM1-Parameteroptimierung

Alle nachfolgend behandelten COSMO-RS-Optimierungsarbeiten wurden mit dem Programmpaket MOPAC [36] mit dem semiempirischen AM1-Verfahren (AM1 = Austin Model 1) durchgeführt. Da es sich bei der AM1-Methode um ein bekanntes semiempirisches Standardverfahren auf Basis der MNDO-Näherung (MNDO = modified neglect of differential overlap) handelt, werden die theoretischen Hintergründe an dieser Stelle nicht erläutert, sondern auf die Literatur verwiesen [37,38]. Unter den semiempirischen Verfahren kann auch die PM3-Methode [39] ähnlich gute Resultate erzielen. Obwohl AM1 größere Fehler als PM3 bei den Bildungsenthalpien aufweist [40], wurde die Entscheidung zu Gunsten des AM1-Verfahrens hauptsächlich wegen der besseren Beschreibung des Stickstoffs gefällt. Im Gegensatz zum AM1-Verfahren zeigt PM3 teilweise fehlerhafte Ladungen auf Stickstoffatomen [40]. Darüberhinaus gibt es Hinweise [41], dass Wasserstoffbrückenbindungsenergien mit AM1 ebenfalls besser vorhergesagt werden.

Noch vor den eigentlichen Optimierungsrechnungen stand die Frage, an welche physikochemischen Observablen das COSMO-RS/AM1-Modell angepasst werden sollte, und wie groß der Datensatz zu wählen ist. Mehrere Aspekte wurden bei diesen Überlegungen berücksichtigt:

- Umweltchemische Relevanz der Stoffeigenschaften
- Verfügbarkeit experimenteller Daten
- Möglichst hohe Vielfalt der physikalischen Aggregatzustände
- Beachtung der Rechenzeit-Ressourcen

Ausgewählt wurden schließlich: Henry-Konstanten (bzw. freie Hydratationsenthalpien), Oktanol/Wasser-Verteilungskoeffizienten und Dampfdrücke. Alle Größen sind umweltchemisch betrachtet ausgesprochen interessant, denn alle drei Größen sind grundlegende Variablen in Expositionsmodellen [8,42]. Die Verfügbarkeit experimenteller Daten bei 25°C ist erfreulich hoch. Wasser ist an zwei der Verteilungsgrößen als Lösungsmittel beteiligt (vergl. Kapitel 2.3) und wird somit entsprechend seiner Vorrangstellung in der Umwelt bei den Optimierungen gebührend berücksichtigt. Neben der umweltchemischen Relevanz ist der Dampfdruck auch modelltheoretisch gesehen sehr wichtig. Der Dampfdruck, der als Verteilung eines Stoffes zwischen Gasphase und *Reinstoff* definiert ist, bringt die große Variabilität an Lösungsmitteln und trägt damit entscheidend zur breiten Anwendbarkeit des

COSMO-RS-Modells bei. Alternativ oder zusätzlich hätten auch Wasserlöslichkeiten hinzugenommen werden können, doch wegen der Verwendung von aus Dampfdruck und Wasserlöslichkeit berechneten Henry-Konstanten wären dann weitere redundante Informationen enthalten.

Das Konzept der Parameteroptimierung gliedert sich in zwei aufeinanderfolgende Schritte. Im ersten Schritt wurde die Parametrisierung an einem Teildatensatz, bestehend aus CHO-Verbindungen, durchgeführt. Der Grund für diese Vorgehensweise ist vor allem die Verkleinerung des Parameterraums und somit die Einsparung von Rechenzeit. Als Nebeneffekt erhält man einen für reine CHO-Verbindungen maßgeschneiderten Parametersatz. In einem zweiten Schritt wurden die gewonnen optimalen Kohlenstoff-, Wasserstoff- und Sauerstoffradien festgelegt und COSMO/AM1-Rechnungen für den erweiterten Datensatz mit Variation der Stickstoff-, Fluor- und Chlorradien durchgeführt. Methodisch gleicht das Vorgehen der Parametrisierung des zugrundeliegenden AM1-Verfahrens, welches ebenfalls schrittweise parametrisiert wurde. Im Gegensatz dazu wurde das PM3-Verfahren global parametrisiert, was zu der vorher erwähnten Ungenauigkeit bei der Beschreibung der Stickstoffpartialladungen führte.

Im folgenden werden die Entwicklungsstadien der Anpassung in drei Kategorien eingeteilt:

- Optimierungen *nullter Ordnung* sind Optimierungen, bei denen die van der Waals-Radien der DFT-Parametrisierung verwendet, und *nur* die übrigen COSMO-RS/DFT analogen Parameter einer Optimierungsprozedur unterzogen wurden.
- Optimierungen *erster Ordnung* sind Anpassungen bei denen sowohl die COSMO-RS-Parameter als auch die COSMO-Radien optimiert wurden.
- Optimierungen *zweiter Ordnung* sind Optimierungen, bei denen programmtechnische Veränderungen vorgenommen und/oder zusätzliche Parameter eingeführt wurden.

Zur Validierung der Resultate wurden die Ergebnisse mit unterschiedlichen Alternativmethoden verglichen, mit denen die Zielgrößen ebenfalls ohne Zuhilfenahme experimenteller Zusatzinformationen direkt aus der chemischen Struktur berechnet werden können. Darüberhinaus wurden COSMO-RS-Rechnungen auf DFT-Niveau zum direkten Vergleich durchgeführt. Um zu überprüfen, inwieweit die Optimierungen zu Parametersätzen führten, die möglicherweise für den speziellen Trainingsdatensatz „maßgeschneidert“ sind, wurde der Moleküldatensatz in einen großen Trainingsdatensatz und einen kleineren Vorhersagedatensatz (Testdatensatz) aufgeteilt. Der Vorhersagedatensatz wurde von allen nachfolgend

beschriebenen Optimierungen ausgeschlossen und erst am Ende zur Validierung der Ergebnisse herangezogen. Die beiden Moleküldatensätze werden im nächsten Kapitel vorgestellt.

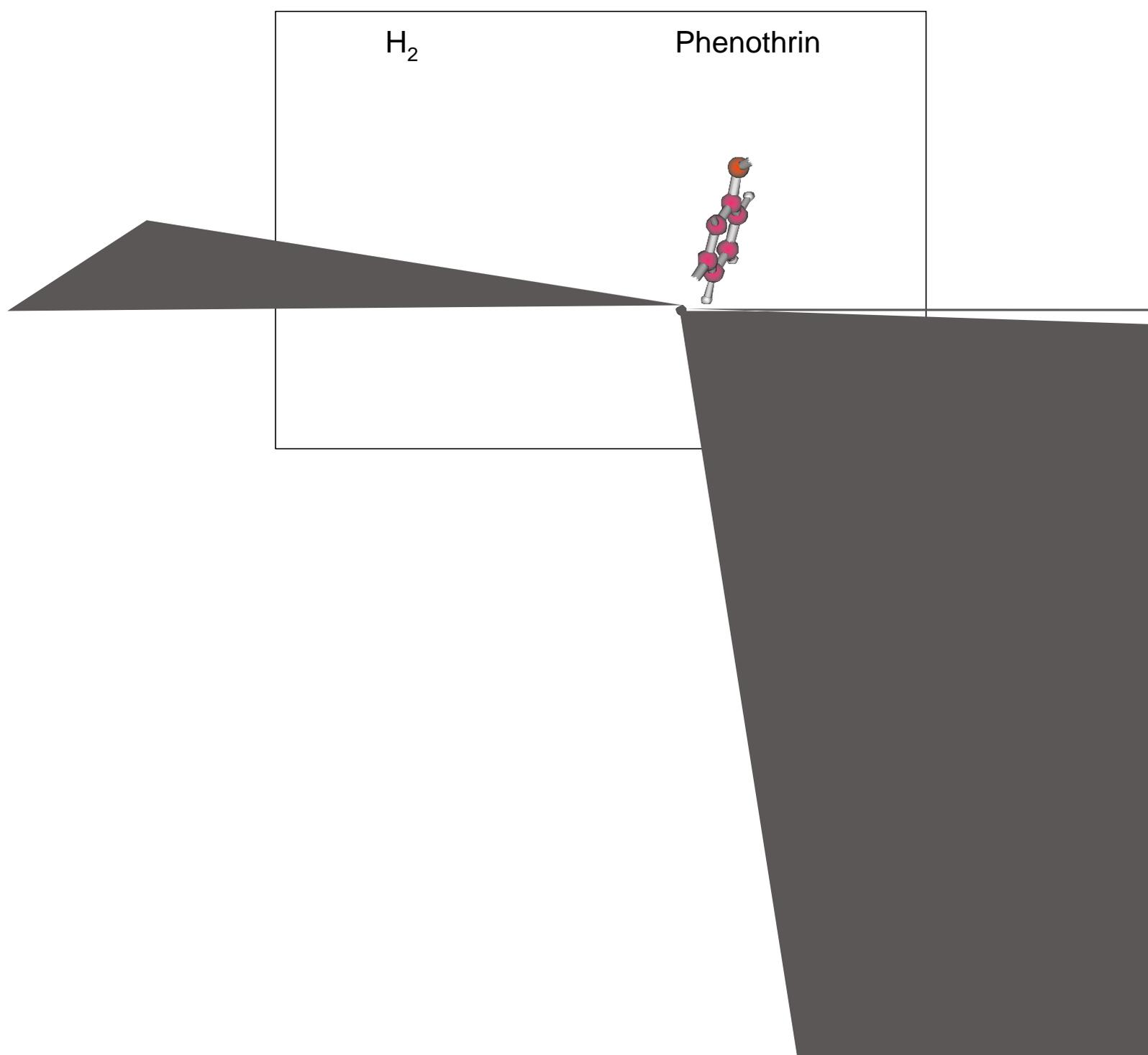
3.1 Der Moleküldatensatz

Die Auswahl des Moleküldatensatzes erfolgte nach der Verfügbarkeit experimenteller Daten bei Raumtemperatur (298.15 K). Da vor allem für Henry-Konstanten nur wenige zuverlässige Rechenmodelle (vor allem unter dem Aspekt der Temperaturabhängigkeit) existieren, wurden die Moleküle nach der Verfügbarkeit experimenteller Henry-Konstanten ausgewählt.

Zur Vergrößerung des CHO-Datensatzes wurden auch Henry-Konstanten nach Gleichung 64 (Seite 2-21) aus Dampfdruck und Wasserlöslichkeit berechnet. Mögliche Messfehler können sich hier allerdings multiplikativ auf die Henry-Konstante auswirken. Für den so gewonnenen Substanzdatensatz wurden im Anschluss noch experimentell ermittelte Oktanol/Wasser-Verteilungskoeffizienten der MedChem-Datenbank [43] entnommen. Experimentelle Dampfdrücke sind aus verschiedenen Quellen zusammengetragen (siehe Anhang A.2 und A.3). Anschließend wurden die Stoffe im Datensatz identifiziert, die bei 25 °C als Feststoffe vorliegen. Deren experimentelle Dampfdrücke wurden wegen der nicht erfassbaren Schmelzenthalpie (vgl. Kapitel 2.3.2 Seite 2-24) aus dem Datensatz entfernt. Der so gewonnene Gesamtdatensatz umfasst 952 Verbindungen mit 2113 physikochemischen Stoffeigenschaften. Eine weitere Reduktion des Datensatzes wurde zu diesem Zeitpunkt nicht vorgenommen. Allerdings wurden nach den ersten Optimierungen insgesamt 29 experimentelle Henry-Konstanten, die sich als eindeutige Ausreißer erwiesen, aus dem Datensatz entfernt.

Im Gegensatz dazu weist A. Klamt in seinem Artikel über die Parametrisierung von COSMO-RS/DFT [33] darauf hin, dass einige Moleküle mit unterschiedlichen Grundzustandskonformationen in Lösung und in der Gasphase aus dem Optimierungsdatensatz entfernt wurden, um eine möglichst robuste Grundparametrisierung zu erhalten. Diese Auswahl wurde in der vorliegenden Arbeit nicht getroffen, denn unter Beachtung der zukünftigen Anwendung der AM1-Parametrisierung für Wirkstoffe oder Proteine wurden auch komplizierte Strukturen im Datensatz belassen um eine möglichst große strukturelle Vielfalt - auch was die Molekülgröße und die Art der funktionellen Gruppen betrifft - zu erreichen. Natürlich steigt mit zunehmender Komplexität der Verbindungen auch die Komplexität des Lösungsvorgangs. Betrachtet man beispielsweise den Hydratationsvorgang einer Aminosäure, so wird diese

dissoziieren und dann in zwitterionischer Form vorliegen. Diesen Vorgängen wird bei den Optimierungen nicht Rechnung getragen. Mit COSMO-RS könnten solche Szenarien zwar prinzipiell simuliert werden, allerdings ist dies für eine Optimierungsprozedur zu komplex. Trotzdem wird beispielsweise auf dissoziierende Säuren im Datensatz nicht verzichtet, um eben die gewünschte Variabilität zu erreichen. Es besteht Grund zur Annahme, dass bei späteren genaueren Rechnungen einzelner Stoffe mit beispielsweise Boltzmann-gewichteten Konformerengemischen und unter Berücksichtigung eines gewissen Anteils dissoziierter Spezies bessere Ergebnisse erzielt werden, wenn die Funktionalität bei der Optimierung berücksichtigt wurde, und diese deshalb implizit in den Parametern berücksichtigt ist.



den globalen energetischen Minima erfolgte deshalb nach der chemischen Intuition und Erfahrung.

3.1.1 Der Trainingsdatensatz

Der Moleküldatensatz wurde wie bereits erwähnt in einen Trainingsdatensatz und einen kleineren Vorhersagedatensatz unterteilt. Die Auswahl erfolgte nach folgendem Verfahren: Zuerst wurde der Gesamtdatensatz in vier Hauptgruppen unterteilt.

- reine CH-Verbindungen
- nur CHO-Verbindungen
- nur CHON-Verbindungen
- CHONFCL-Verbindungen

Die Hauptgruppen wurden ihrerseits mit dem in der ChemProp-Software [93] enthaltenen Strukturanalyseprogramm nach Strukturgruppen geordnet. Aus diesen Untergruppen wurden mittels eines Zufallszahlenprogramms die Moleküle für den Vorhersagedatensatz ausgewählt. Der gesamte Trainingsdatensatz umfasst 749 Moleküle mit 120 Kohlenwasserstoffen, 234 CHO-Verbindungen, 160 Verbindungen mit Stickstoff (neben C, H und O) und schließlich noch einmal 234 Stoffe mit den vorangegangenen Elementen plus Chlor und Fluor. Für die ersten beiden Hauptgruppen wurden jeweils 33% der Substanzen für den Testdatensatz ausgewählt, wohingegen von den letzten beiden Gruppen nur jeweils 10% der Substanzen für den Vorhersagedatensatz ausgesucht wurden.

Tabelle 1 (nächste Seite) zeigt eine genaue hierarchische Einteilung des Trainingsdatensatzes am Beispiel der Substanzen mit verfügbarer experimenteller Henry-Konstante. Bei eingerückten Gruppen handelt es sich um Untergruppen der Vorangegangenen, d.h. die folgende Tabelle gliedert die Substanzen in sieben Hauptgruppen (**ohne Halogene, Halogene, N, O, N und O, anorganisch**) mit zahlreichen Untergruppen. Von der ersten Gruppe der Verbindungen ohne Halogenatome (118 Substanzen) sind beispielsweise 66 Verbindungen aromatisch und 52 nicht-aromatisch. Die nicht-aromatische Gruppe beinhaltet ihrerseits 35 Alkane, 24 Alkene und 7 Alkine.

Tabelle 1: Hierarchische Einteilung des Henry-Konstanten-Trainingsdatensatzes zur Veranschaulichung der strukturellen Vielfalt. Eingerückte Gruppen sind Untergruppen der vorangegangenen Gruppe.

Gruppen	#	Gruppen	#
ohne Halogene	118	mit O und N	115
nicht-aromatisch	66	ohne Halogene	14
ohne ungesättigten Alkyle	35	Amine	5
Doppelbindungen	24	Nitrile	2
Dreifachbindungen	7	ar-Azine	7
aromatisch	52	N-C=O Gruppen	14
ein aromatischer Ring	28	ohne Halogene	8
Biphenyle	2	Halogene	6
kondensierte Aromaten	22	NO-Gruppen	87
Halogene	169	ohne Halogene	76
F	13	disubstituierte Nitrosamine	8
nicht-aromatisch	11	(ohne O-Gruppen)	6
aromatisch	2	nicht aromatisch	6
Cl	143	Ether/Furane	2
nicht-aromatisch	43	Epoxide/Furane	2
aromatisch	100	Nitro	21
gemischte Halogene	13	(ohne O Gruppen)	15
Sauerstoff	241	nicht-aromatisch	5
keine Halogene	203	aromatisch	10
Alkohole	48	OH-Gruppen	4
Ether/Furane	34	aromatische Alkohole	4
Peroxid	0	Ether/Furane	2
Hydroperoxid	1	Ether	2
Carbonyle	44	N-NO ₂	2
Carboxyle	56	nicht-aromatisch	2
gemischte O-Gruppen	20	ohne ungesättigte Alkylgruppen	2
Halogene	38	Nitrat	38
Alkohole	10	(ohne O-Gruppen)	33
Ether/Furane	17	nicht-aromatisch	33
Carbonyle	1	OH-Gruppen	5
Carboxyle	9	p-Alkohole	2
gemischte O-Gruppen	1	s-Alkohole	3
Stickstoff	71	gemischte N-, NO-, NCO-Gruppen	7
keine Halogene	62	Halogene	11
Amine	34	Nitro	9
Nitrile	7	(ohne O-Gruppen)	8
Azole	1	Cl	8
ar-Azine	18	gemischte O-Gruppen	1
gemischte N-, NO-,NCO-Gruppen	2	gemischte N-, NO-, NCO-Gruppen	2

2

I-, NO-, NCO-tickstoff) Tj 189 0 51 Tc 4 0 TD 0.3 Tc 0 Tw (11) Tj -400.2 -13.8 TD -0.2729 Tc 0.1979 Tw (Amine) Tj 19 0 TD 0.3 Tc
g e n e

Tabelle A-1 im Anhang (Seite A-1 ff.) zeigt den kompletten Trainingsdatensatz inklusive CAS-Nummern, experimentellen Henry-Konstanten, Dampfdrücken und Oktanol/Wasser-Verteilungskoeffizienten mit den zugehörigen Referenzen. in Abbildung 7 sind die Wertebereiche der experimentellen Größen angegeben

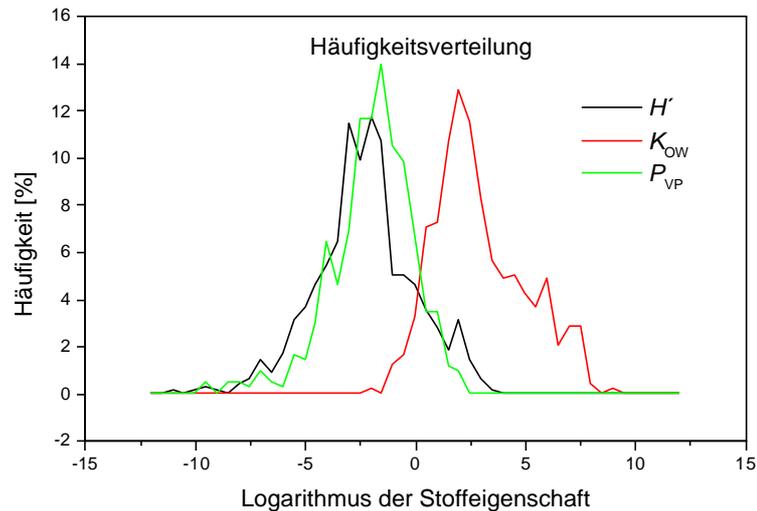


Abbildung 7: Wertebereiche und Häufigkeiten der experimentellen Daten im Trainingsdatensatz.

3.1.2 Der Vorhersagedatensatz

Der deutlich kleinere Vorhersagedatensatz (Tabelle A-2 im Anhang Seite A-16 ff.) umfasst insgesamt 203 Stoffe. Wegen der Verwendung berechneter Henry-Konstanten bei den Kohlenwasserstoffen und den CHO-Verbindungen herrschte ein Überfluss an Datenmaterial. Deshalb wurden ca. 30 % der verfügbaren Substanzen für den Vorhersagedatensatz verwendet. Da bei den stickstoff-, chlor- und fluorhaltigen Stoffen auf berechnete Henry-Konstanten verzichtet wurde, konnten nur 10 % der insgesamt verfügbaren Verbindungen für den Vorhersagedatensatz ausgewählt werden.

3.2 Optimierungen und Parameter

Am Beginn der Optimierungsarbeiten mussten die zu optimierenden Parameter ausgewählt, deren Wertebereiche bestimmt und die Startwerte festgelegt werden. Natürlich liegt es nahe von bekannten sinnvollen Werten, in diesem Falle also von den veröffentlichten COSMO-

RS/DFT-Parametern [33] auszugehen. Alle für DFT optimierten COSMO-RS-Parameter sind deshalb zur Übersicht in der nachfolgenden Tabelle angegeben:

Tabelle 2: Veröffentlichte COSMO-RS/DFT-Parameter [33] für DMOL-Rechnungen.

Parameter	Wert	Parameter	Wert
r_H	1.30 Å	γ_H	-0.041 kcal/(mol Å ²)
r_C	2.00 Å	γ_C	-0.037 kcal/(mol Å ²)
r_O	1.72 Å	γ_O	-0.042 kcal/(mol Å ²)
r_N	1.83 Å	γ_N	-0.027 kcal/(mol Å ²)
r_{av}	0.50 Å		
α'	1288 kcal/mol Å ² /e ²	a_{eff}	7.1 Å ²
f_{corr}	2.40	λ	0.14
c_{hb}	7400 kcal/(mol Å ²)/e ²	ω	-0.21 kcal/mol
σ_{hb}	0.0082 e/Å ²	η	-9.15

Die einzelnen Parameter sollen an dieser Stelle noch einmal beschrieben und ihre Funktion möglichst anschaulich erklärt werden.

r_X sind die elementspezifischen van der Waals-ähnlichen Radien [44-47] die schon vor der eigentlichen RS-Rechnung, nämlich schon bei den zugrundeliegenden MOPAC-COSMO/AM1-Rechnungen zur Konstruktion der SAS benötigt werden. Üblicherweise sind die in Kontinuumsansätzen verwendeten Radien ca. 20% größer als die von Bondi experimentell ermittelten. Auf der so gewonnen Oberfläche wird zur endgültigen Konstruktion der SAS noch eine Kugel mit Radius r (vergleiche Abbildung 1 Seite 2-1) abgerollt. Bei den Optimierungen wird dieser Radius gleich dem jeweils verwendeten van der Waals-Radius von Wasserstoff gesetzt. Die van der Waals-Radien werden nachfolgend auch als COSMO-Radien bezeichnet. Die Variation der COSMO-Radien beanspruchte den größten Teil der für die Optimierungsprozedur benötigten Rechenzeit, da für jede Radienkombination neue COSMO/AM1-Rechnungen für den gesamten Datensatz durchgeführt werden mussten. Die Prozedur der Radienkombination wurde deshalb automatisiert.

Nachfolgendes Diagramm zeigt schematisch den Ablauf des zur Radienvariation und Steuerung der MOPAC-Rechnungen produzierten Programms.

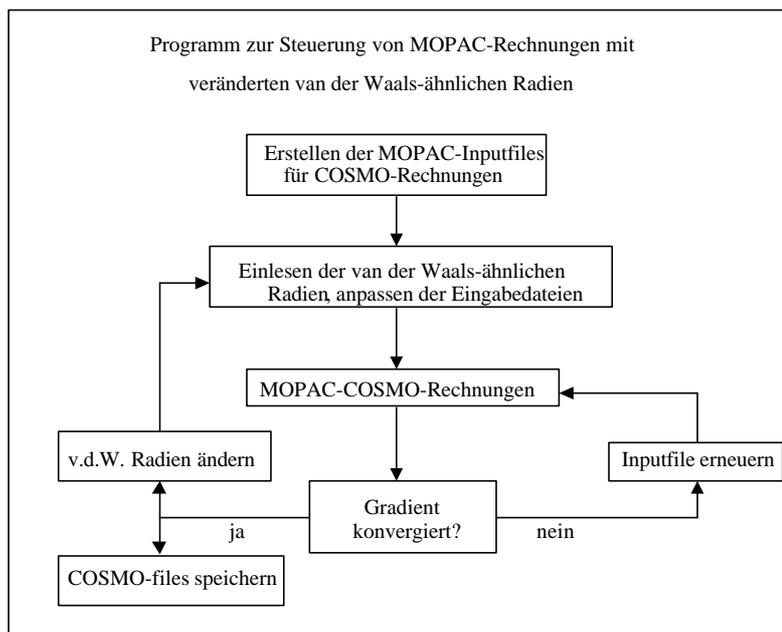


Abbildung 8: Flussdiagramm für automatisierte AM1-Rechnungen mit dem semiempirischen Programmpaket MOPAC.

Alle übrigen Parameter sind COSMO-RS-spezifisch und können anschließend für die einzelnen Radien-Kombinationen effizient ermittelt werden. Die COSMO-RS-Parameter gliedern sich ihrerseits in zwei Gruppen: Parameter, die schrittweise verändert werden müssen, und Parameter, von denen die Stoffeigenschaften linear abhängen. Letztere können für die Kombinationen der nicht-linearen Parameter durch Lösung eines überbestimmten linearen Gleichungssystems mit Hilfe einfacher Matrizenoperationen ermittelt werden.

Die elementspezifischen Dispersionsparameter g_x , die mit den Moleküloberflächenanteilen A_x multipliziert werden (siehe Gleichung 59 Seite 2-19), simulieren in der Summe die Stabilisierung des Moleküls durch Dispersionswechselwirkung beim Transfer des Moleküls von der Gasphase in die Lösung. Die Korrelation des Lösungsbestrebens unpolarer Stoffe mit der Molekülgröße ist im Augenblick Gegenstand theoretischer Untersuchungen [48,49]. Neben den Dispersionsparametern sind auch die übrigen Parameter des Gasphasenpotenzials: der Ringparameter w (zur Korrektur des Gasphasenpotenzials von Ringverbindungen) und der Entropieparameter h linearer Natur. Auch der kombinatorische Faktor I (siehe Gleichung 52 Seite 2-17) der den Einfluss der Molekülgröße des Lösungsmittels auf das chemische Potenzial quantifiziert und der Parameter $\beta (= kT / a_{eff})$ sind linearer Natur. Alle weiteren Parameter sind nicht linearer Natur: Anstelle des oben aufgeführten Parameters $a'(f_{pol})$ wird im Folgenden ein Parameter für die Misfitenergie c_{mf} angegeben.

$$c_{mf} = -\frac{\mathbf{a}'}{\mathbf{a}} a_{eff} \quad \text{Gleichung 98}$$

Werden zu den linearen COSMO-RS-Parametern die nicht-linearen adjustierbaren Größen: c_{mf} = Parameter zur Quantifizierung der Misfitenergie, r_{av} = Radius der Fläche über die die COSMO-Oberflächenladungsdichten gemittelt werden (vgl. Gleichung 31 Seite 2-12) und die beiden Wasserstoffbrückenbindungsparameter c_{hb} und s_{hb} (vgl. Gleichung 53 Seite 2-18) hinzugenommen, erhält man insgesamt acht COSMO-RS-spezifische Parameter plus je zwei elementspezifische Parameter (r_X und \mathbf{g}_X).

Eine Sonderstellung nimmt der zur Erfassung der lokalen Korrelation der Abschirmladungsdichten eingeführte Korrekturfaktor f_{corr} (siehe Gleichung 56 Seite 2-19) ein, da dieser nicht frei optimierbar sein muss, sondern aus den COSMO-Oberflächenladungsdichten durch bilineare Regression gewonnen werden kann:

$$\Delta^{iX} = -\frac{\mathbf{a}}{2} f_{pol}^{1/2} \left[\sum_{n \in X} s_v \mathbf{s}_n^2 + f_{corr} \sum_n s_v \mathbf{s}_n \mathbf{s}_n^\perp \right]^2 \quad \text{Gleichung 99}$$

Die Gleichung entspringt der Forderung, dass die einzigen Abschirmladungen eines in einem Ensemble unpolarer, nicht polarisierbarer Lösungsmittelmoleküle plazierten Moleküls ausschließlich die Abschirmladungen des gelösten Moleküls selbst sind. Die Energie dieses elektrostatisch eingefrorenen Moleküls im Vakuum relativ zum beschriebenen abgeschirmten Zustand ist:

$$E_{Vakuum}^{frozenX} = \frac{\mathbf{a}}{2} \sum_{n \in X} s_v \mathbf{s}_n^2 \cong -E_{diel}^{iX} = -f_{pol}^{1/2} \Delta^{iX} \quad \text{Gleichung 100}$$

Wird nun die lokale Korrelation der Abschirmladungsdichten berücksichtigt, dann ist die Summe in Gleichung 100 gleich:

$$\sum_{n \in X} s_v \mathbf{s}_n^2 \cong \sum_{n \in X} s_v \mathbf{s}_n (\mathbf{s}_n + f_{corr} \mathbf{s}_n^\perp) \quad \text{Gleichung 101}$$

und man erhält die in Gleichung 99 angegebene Beziehung.

Es zeigt sich, dass nur die Variation von r_{av} den Wert von f_{corr} nachhaltig beeinflusst. Bilineare Regression von Gleichung 99 liefert für die COSMO-Radienkombination $r_H = 1.25 \text{ \AA}$, $r_C = 2.00 \text{ \AA}$ und $r_O = 1.80 \text{ \AA}$ mit $r_{av} = 0.50 \text{ \AA}$ bzw. $r_{av} = 0.36 \text{ \AA}$ das folgende Bild:

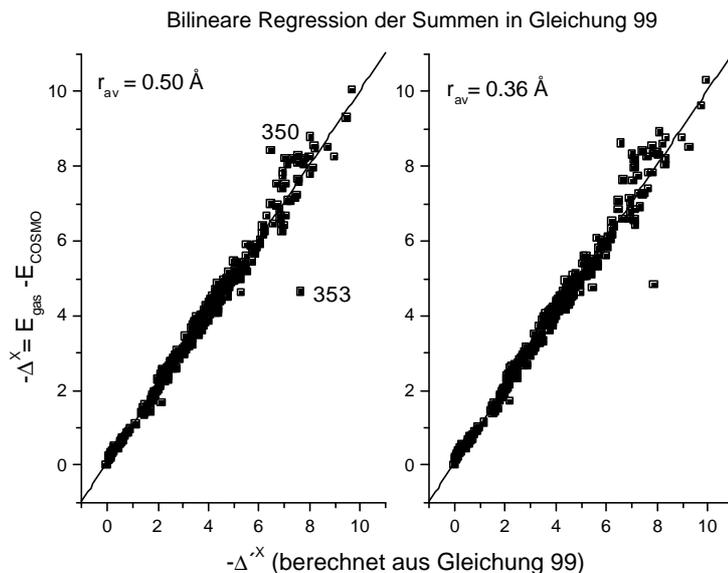


Abbildung 9: Bilineare Regression von Gleichung 99 zur Bestimmung von f_{corr} . Die Qualität der Korrelation ist für beide r_{av} -Radien gut, allerdings ergeben sich verschiedene f_{corr} -Werte.

Ausreißer sind die Substanzen T350 (Methoprene) und T353 (Pyruvic acid), aber dennoch korrelieren die beiden Summen in Gleichung 99 sehr gut mit Δ^X ($r^2 = 0.98$, $SD = 0.29$ mit $r_{av} = 0.50 \text{ \AA}$). f_{corr} ergibt sich aus dem Quotient der Steigungen der beiden Summen: $f_{corr} = 2.98$ (für $r_{av} = 0.50 \text{ \AA}$).

Für die gleiche van der Waals-Radienkombination aber mit einem r_{av} -Radius von 0.36 \AA ergibt sich: $f_{corr} = 1.54$. Die Korrelation ist in etwa von gleicher Qualität ($r^2 = 0.98$, $SD = 0.31$). Eine Vergrößerung der van der Waals-Radien ($r_H = 1.30 \text{ \AA}$, $r_O = 1.90 \text{ \AA}$ bei Beibehaltung von $r_C = 2.00 \text{ \AA}$) liefert: $f_{corr} = 2.74$ (für $r_{av} = 0.5 \text{ \AA}$) mit ähnlichem Regressionskoeffizient und gleicher Standardabweichung. Der f_{corr} -Faktor ändert sich bei geänderten van der Waals-Radien nur wenig (bei gleichem averaging-Radius). Wird nun die Regression wieder für $r_{av} = 0.36 \text{ \AA}$ durchgeführt, ermittelt man einen Korrekturfaktor f_{corr} von 1.63 .

f_{corr} hängt also in erster Linie weniger von den COSMO-Atomradien, als vom Mittelungsradius r_{av} ab.

Aus diesen Ergebnissen resultiert die Vorgehensweise bei der Optimierung: Für einen bestimmten Satz von van der Waals-Radien wird die Optimierungsroutine mehrmals mit verschiedenen Startwerten für die Variablen c_{mf} , s_{hb} , c_{hb} , f_{corr} und r_{av} gestartet. Besonderes Augenmerk wird bei den Ergebnissen auf r_{av} gelegt:

Ist der Radius r_{av} stark vom Startwert abgewichen, kann es sein, dass f_{corr} als zu optimierender Parameter sehr stark von dem Wert abgewichen ist, der sich bei einem Fit nach Gleichung 99 ergeben hätte. Es ist also in jedem Falle ratsam, für den neuen Radius f_{corr} noch einmal durch

eine Anpassung zu bestimmen. Anschließend ist die Optimierung mit dem nun festgesetzten f_{corr} -Wert noch einmal zu starten.

Diese Vorgehensweise wird für die Radienkombination $r_{\text{H}} = 1.25 \text{ \AA}$, $r_{\text{C}} = 2.00 \text{ \AA}$ und $r_{\text{O}} = 1.8 \text{ \AA}$ in Tabelle 3 gezeigt. Die Optimierungen eins bis fünf (mit Startwert $r_{\text{av}} = 0.5 \text{ \AA}$, alle anderen Parameter mit verschiedenen Startwerten) ergaben:

Tabelle 3: Endparameter und Fehlerquadratsummen (SOR) der Stoffeigenschaften für sechs verschiedene Startwertkombinationen der nicht-linearen COSMO-RS-Parameter.

Optimierung	c_{mf}	$s_{\text{hb}} [\text{e/\AA}]$	$c_{\text{hb}} [\text{kcal}/(\text{mol \AA})/\text{e}^2]$	f_{corr}	$r_{\text{av}} [\text{\AA}]$	SOR
1	2.20	0.0010	54000	1.90	0.36	412
2	2.20	0.0000	45000	1.10	0.34	404
3	2.80	0.0000	46600	0.70	0.37	406
4	2.80	-0.0004	43900	1.00	0.43	410
5	4.00	0.0002	47200	1.90	0.35	406
6	2.60	0.0000	45000	1.30	0.36	405

In diesen Fällen sind die Summen der Fehlerquadrate nahezu gleich. Lässt man Optimierung Nummer vier wegen des negativen s_{hb} außer acht, ergibt sich der gemittelte optimierte Radius zu $r_{\text{av}} = 0.36 \text{ \AA}$.

Eine erneute sechste Optimierung mit festgesetztem $f_{\text{corr}} = 1.30$, resultierend aus der bilinearen Regression (Gleichung 99 Seite 3-10) und mit $r_{\text{av}} = 0.36 \text{ \AA}$, produziert nur minimal unterschiedliche Resultate und wird deshalb nur am Ende zur Feinabstimmung angewendet.

D.h. f_{corr} wird in der Optimierung zunächst als freier gleichwertiger Parameter angesehen.

Zur Optimierung der COSMO-RS-Parameter wurde das Programm modifiziert, das von Dr. Klamt zur DFT-Parametrisierung entwickelt wurde. Die Veränderungen bezogen sich im ersten Schritt hauptsächlich auf den I/O-Teil des Programms, um das Einlesen der MOPAC-COSMO-Ausgabedateien zu ermöglichen und dem veränderten Moleküldatensatz mit der reduzierten Anzahl an physikochemischen Eigenschaften gerecht zu werden. Im zweiten Schritt wurden dann inhaltliche Veränderungen vorgenommen.

Abbildung 10 (Seite 3-13) zeigt das Flussdiagramm der Parameteroptimierungsroutine. Nachdem alle COSMO-Ausgabedateien und andere notwendige Informationen eingelesen sind, werden die chemischen Potentiale berechnet. Die optimalen linearen Parameter werden durch Matrizenoperationen bestimmt. Dann wird die Fehlerquadratsumme aller Größen gebildet und gespeichert.

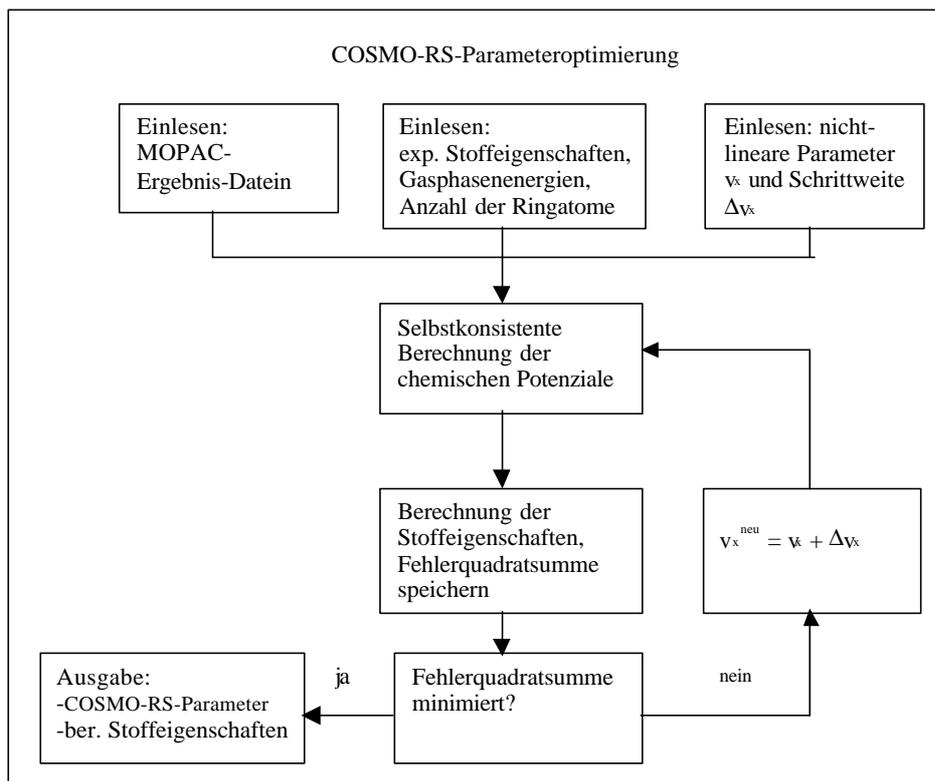


Abbildung 10: Flussdiagramm der COSMO-RS-Parameter-Optimierungsroutine.

Die nicht-linearen Parameter werden *nacheinander* in der Reihenfolge c_{mf} , s_{hb} , c_{hb} , f_{corr} , r_{av} um die vorgegebenen Schrittweiten Δv_x verändert und die Zielgrößen mit den zugehörigen Fehlerquadraten neu bestimmt. Wird keine Reduktion der Fehlerquadrate erreicht, wird das Ergebnis verworfen und Δv_x addiert. Ergibt sich eine kleinere Fehlerquadratsumme, wird schrittweise weiterverfahren. Erst wenn durch Veränderung dieses Parameters keine statistische Verbesserung mehr erfolgt, wird mit dem nächsten Parameter fortgefahren. Kann die Veränderung *aller* nicht-linearen Parameter in einem Zyklus (c_{mf} bis r_{av}) keine Verbesserung erbringen, werden die Ergebnisse ausgegeben.

3.3 Optimierungen mit dem CHO-Teildatensatz

Den Startoptimierungen lagen MOPAC-Rechnungen mit den DFT-analogen van der Waals-Radien $r_H = 1.30$, $r_C = 2.00$ und $r_O = 1.72$ zugrunde.

Die ersten Testrechnungen mit dem CHO-Datensatz (T1 –T355) lieferten Fehlerquadratsummen, die im Bereich von 570 – 650 lagen. In Tabelle 4 (Seite 3-14) sind die Parameter der Optimierungen gelistet. Es ergibt sich ein im Vergleich zur DFT-Parametrisierung verkleinerter f_{corr} -Wert. Die Optimierungen drei bis sechs sind in lokale Minima gelaufen, sie

liefern weitaus schlechtere SOR-Werte. Der negative s_{hb} -Parameter ist darüber hinaus unphysikalisch. Dies zeigt, dass die Startwerte das Erreichen der verschiedenen Minima stark beeinflussen. Deshalb wurde zuerst versucht, die Startwerte nach einem *factorial design* [50] zu generieren. Erste Tests zeigten jedoch, dass ein umfassendes Absuchen des gesamten Parameterraums wegen seiner hohen Dimensionalität ($n=5$ bis 7 , je nach Anzahl der nicht-linearen Parameter (Faktoren)) sehr rechenintensiv ist. Beispielsweise ergäbe ein befriedigendes *central composite design* für sechs Faktoren bereits 77 Startwertkombinationen. Eine Überprüfung zeigte, dass die meisten der Parameter-Startkombinationen zu lokalen Minima mit teilweise unphysikalischen Parametersätzen führten. Deshalb ist das systematische Absuchen des gesamten Parameterraums nur wenig sinnvoll.

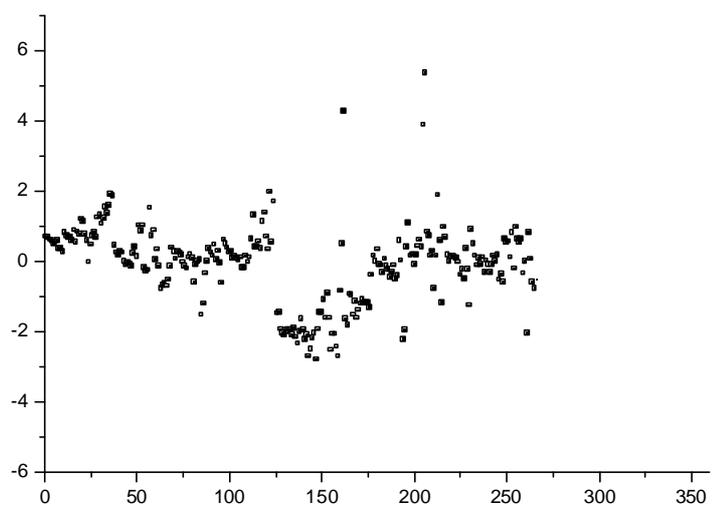
Tabelle 4: Fehlerquadratsummen (SOR) der Stoffeigenschaften und Endparameter von Optimierungen mit identischen COSMO-Radien aber unterschiedlichen Startwerten der übrigen Parameter.

Optimierung :	#1	#2	#3	#4	#5	#6
c_{mf}	2.40	1.20	2.60	2.60	3.00	1.90
s_{hb} [e/Å]	0.0066	0.0070	-0.0020	-0.0008	-0.0006	-0.0020
c_{hb} [kcal/mol Å ² /e]	114000	117000	36400	34000	33100	35400
f_{corr}	1.15	1.50	14.10	10.50	9.10	14.40
r_{av} [Å]	0.38	0.29	0.40	0.39	0.41	0.36
w [kcal/mol]	0.255	0.251	0.259	0.259	0.261	0.258
g_H [kcal/mol Å]	-0.0343	-0.0347	-0.0318	-0.0302	-0.0304	-0.0322
g_C [kcal/mol Å]	-0.0305	-0.0310	-0.0309	-0.0272	-0.0267	-0.0307
g_O [kcal/mol Å]	0.0090	0.0086	0.0781	0.0488	0.0446	0.0762
l	-0.3794	-0.3024	-0.5613	-0.6255	-0.5628	-0.6134
h	-1.1371	-0.9662	-1.3992	-1.5137	-1.3538	-1.5309
b [kcal/mol Å ²]	0.0950	0.0972	0.0453	0.0487	0.0510	0.0445
SOR	587	574	642	652	653	645

Tabelle 4 offenbart noch ein weiteres Problem: einen positiven Sauerstoff-Dispersionsparameter g_O . Dies ist unphysikalisch, würde es doch bedeuten, dass bei Molekülen mit Sauerstoffatomen die Dispersionswechselwirkungen abstoßender Natur sind. Der Radius r_{av} ist verglichen mit der DFT-Parametrisierung beachtlich verkleinert.

Die Standardabweichungen sind: $SD(\Delta G_{hydr}) = 0.92$ kcal/mol, $SD(\log K_{OW}) = 0.70$ und $SD(\log P_V) = 0.75$.

Die genaue Analyse der Fehler der vorhergesagten freien Hydratationsenthalpien (Abbildung 11 nächste Seite) zeigte einen systematischen Fehler der Alkohole.



å

Tabelle 5: Endparameter der CHO-Optimierung mit linearem Gasphasen-Korrekturfaktor für OH-Verbindungen τ_1 .

Parameter	Wert	Parameter	Wert
r_H	1.25 Å	γ_H	-0.0314 kcal/(mol Å ²)
r_C	2.00 Å	γ_C	-0.0364 kcal/(mol Å ²)
r_O	1.70 Å	γ_O	0.0259 kcal/(mol Å ²)
r_{av}	0.38 Å	τ_1	-1.42
c_{mf}	2.00 kcal/mol Å ² /e ²	β	0.099 kcal/mol Å ²
f_{corr}	2.60	λ	-0.12
c_{hb}	87400 kcal/(mol Å ²)/e ²	ω	-0.20 kcal/mol
σ_{hb}	0.0075 e/Å ²	η	-8.55

Der lineare OH-Parameter τ_1 ergab sich zu -1.42 . Zwar erhält man nun einen kleineren Sauerstoffdispersionsparameter, jedoch ist dieser immer noch positiv ($\gamma_O = 0.0259$).

Insgesamt ist diese Art der Lösung aber ziemlich unelegant, ist sie doch eine einfache Korrektur einer speziellen Stoffgruppe.

Die Verschlechterung der K_{OW} -Statistik und vor allem der Fakt, dass die OH-Verbindungen potenzielle Wasserstoffbrückenbildner sind, führten zum Versuch am Wasserstoffbrückenbindungsterm Veränderungen vorzunehmen.

Es zeigt sich, dass die Statistik durch Reskalierung der Abschirmladungsdichten am Sauerstoff im Wasserstoffbrückenbindungsterm ähnlich gut verbessert werden kann, wie mit dem linearen OH-Gasphasenkorrekturfaktor.

Eine spezielle *gruppenspezifische* (z.B. Sauerstoff in der Hydroxygruppe) Skalierung verbessert die Statistik weiter. Im COSMO-RS-Programm findet eine Abfrage statt, ob ein Segment potenzieller Wasserstoffbrückenbindungsteilnehmer ist. Dies ist für Wasserstoffsegmente mit negativen Abschirmladungsdichten und für an Sauerstoffen positionierte Segmente mit positiven Abschirmladungsdichten der Fall. Diese Abschirmladungsdichten werden für die Verwendung im Wasserstoffbrückenbindungsterm mit $m = 1$ multipliziert, andere mit $m = 0$. Nun wurde dieser Multiplikator nicht auf eins festgesetzt, sondern je nach chemischer Funktionalität verändert. Mit den in Tabelle 6 (Seite 3-17) gelisteten Multiplikatoren für die Abschirmladungsdichten der Sauerstoffe in verschiedenen Funktionalitäten verringerte sich die Fehlerquadratsumme auf 348. Dies entspricht einer Verbesserung um 21% im Vergleich zur Optimierung mit linearem OH-Gasphasenkorrekturfaktor t_1 .

Tabelle 6: Abschirmladungsdichten-skalierende Multiplikatoren für Sauerstoffe im COSMO-RS-Wasserstoffbrückenbindungsterm. In Klammern sind die funktionellen Gruppen angegeben (z.B. O(C=O) ist Sauerstoff in einer Carbonyl-Funktion).

Sauerstoff	Multiplikator m	Sauerstoff	Multiplikator m
O (OH)	1.00	O (R-O-R)	0.80
O (C=O)	0.75	O (R-O-O-R)	0.80

Mit diesen Skalierungen wird der Sauerstoffdispersionsparameter g_o nun wie gewünscht negativ. Auch die übrigen COSMO-RS-Parameter liegen in einem physikalisch sinnvollen Bereich.

Alternativ zur vorgenommenen Skalierung hätten wahrscheinlich *funktionalitätsabhängige* COSMO-Sauerstoffradien anstelle *eines* universellen Sauerstoffradius bei der Konstruktion der SAS die Verbesserung ebenfalls erbringen können, denn die Skalierung entspricht je nach Multiplikator einer Verkleinerung oder Vergrößerung der erzeugten Oberflächenladungsdichten, die sich wohl auch mit differenzierten Radien ergeben hätte. Dieser Weg wurde in der vorliegenden Arbeit jedoch nicht verfolgt. Zwar wird diese Methodik bereits erfolgreich in anderen Kontinuumsmodellen angewendet, doch diese Änderung wirkt global und nicht nur im Wasserstoffbrückenbindungsterm. Durch die vorgenommene Reskalierung kann aber gerade im verwendeten CHO-Minimaldatensatz die Sauerstoff-Wasserstoffbrückenbindung energetisch verbessert werden, da keine Wasserstoffbrückenanteile anderer Atome (z. B. N...H-Brückenbindungen) stören können.

Trotzdem sollte die funktionalitätsabhängige Variation der COSMO-Radien bei weiteren methodischen Verbesserungen geprüft werden.

Vorteil der nachträglichen Skalierung ist der weitaus geringere Rechenaufwand. Nachteil ist, dass die modelltheoretisch richtigen Abschirmladungsdichten im perfekten Leiter (für den vorgegeben Sauerstoffradius) nachträglich verändert werden.

Schließlich führte eine Optimierung, bei der ausschließlich die Abschirmladungsdichten der Sauerstoffatome in Hydroxygruppen mit t skaliert wurden, zu einem äquivalenten statistischen Resultat. Gleichung 103 zeigt den COSMO-RS-Wasserstoffbrückenbindungsenergieterm mit dem Multiplikator t .

$$E_{hb}(\mathbf{s}, \mathbf{s}') = c_{hb} \max[0, t \cdot \mathbf{s}_{acc} - \mathbf{s}_{hb}] \cdot \min[0, \mathbf{s}_{don} + \mathbf{s}_{hb}]$$

Gleichung 103

mit :

$$t = \begin{cases} 0.00 & \text{wenn } \mathbf{s} \notin \text{Sauerstoff} \\ 1.21 & \text{wenn } \mathbf{s} \in O(\text{Hydroxy}) \\ 1.00 & \text{wenn } \mathbf{s} \in O(\text{sonst}) \end{cases} \quad \text{Gleichung 104}$$

Anschaulich bedeutet dies, dass die Wasserstoffbrückenbindungsakzeptorstärke von Hydroxygruppen relativ zu den übrigen Sauerstofffunktionalitäten (Carbonylen, Ethern etc.) verstärkt wird. Die Gesamtbeiträge pro Segmentpaar sind im Vergleich zur Optimierung nullter Ordnung² ($c_{\text{hb}} = 117000 \text{ kcal/mol } \text{\AA}/e^2$ (siehe Tabelle 4 Seite 3-14)) allerdings kleiner ($c_{\text{hb}} = 52900 \text{ kcal/mol } \text{\AA}/e^2$). Bei dieser Optimierung wurde ein signifikant vergrößerter Sauerstoffradius von $r_{\text{O}} = 1.90 \text{ \AA}$ verwendet. Anscheinend können so kleinere systematische Abweichungen der Ether- und Carbonylfunktionen von den Standard-COSMO-RS-Parametern abgefangen werden. Abbildung 12 zeigt die absoluten Fehler der Zielgrößen für diese Optimierung zweiter Ordnung² und offenbart gleichzeitig ein Manko dieser Parametrisierung: Die eigentlich strukturell einfachen Alkane (#1- #60) zeigen einen systematischen Fehler bei den freien Hydratationsenthalpien.

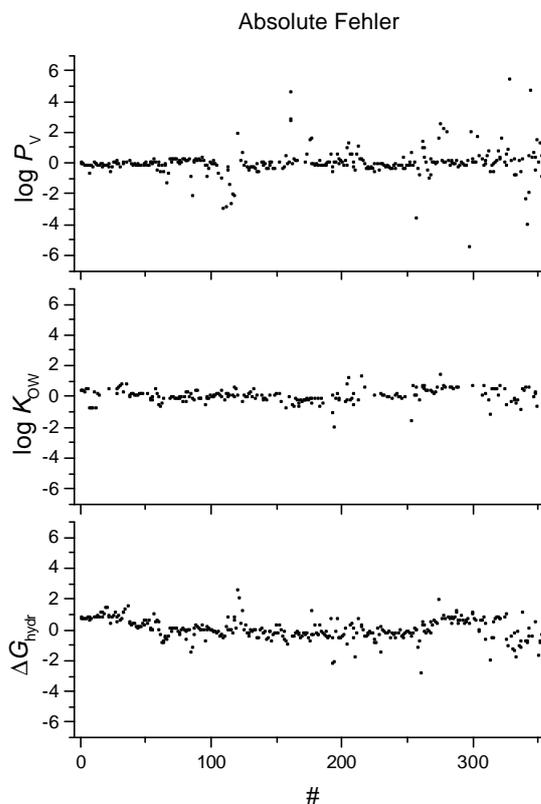


Abbildung 12: Absolute Fehler (exp-ber) der drei physikochemischen Stoffeigenschaften bei der CHO-Optimierung (Substanzen T1 bis T355 des Trainingsdatensatzes).

² Definitionen: **Optimierungen nullter, erster und zweite Ordnung** siehe Seite 3-2

Auf diesen Fehler wird im folgenden Kapitel noch einmal genauer eingegangen. Nachfolgende Tabelle zeigt abschließend die COSMO-RS-Parameter dieser Optimierung zweiter Ordnung:

Tabelle 7: COSMO-RS-Parameter der CHO-Optimierung mit skalierendem Hydroxysauerstoff-Parameter τ .

Parameter	Wert	Parameter	Wert
r_H	1.25 Å	γ_H	-0.0289 kcal/(mol Å ²)
r_C	2.00 Å	γ_C	-0.0503 kcal/(mol Å ²)
r_O	1.90 Å	γ_O	-0.0399 kcal/(mol Å ²)
r_{av}	0.29 Å	$\tau(\text{OH})$	1.21
β	0.075 kcal/(mol Å ²)	c_{mf}	3.2
λ	0.16	f_{corr}	0.1
ω	-0.213 kcal/mol	c_{hb}	52900 kcal/(mol Å ²)/e ²
η	-9.07	σ_{hb}	0.0026 e/ Å ²

Die Tabelle zeigt, dass die Parameter analog den DFT-Parametrisierungen in einem physikalisch sinnvollen Bereich liegen. Eine Ausnahme ist jedoch der f_{corr} -Parameter der mit 0.1 außerordentlich klein ist. Da der Parameter nahe Null ist, bedeutet dies aber lediglich, dass die zur Berücksichtigung lokaler Korrelationen eingeführten Ladungsdichten \mathbf{s}_n^\perp (vgl. Gleichung 55 und Gleichung 56 auf Seite 2-18) bei skalierten Wasserstoffbrücken-Sauerstoff-Abschirmladungsdichten hinfällig werden. Der früher diskutierte Sauerstoffdispersionsparameter ist nun negativ, nur c_{hb} ist im Vergleich zur DFT-Parametrisierung signifikant vergrößert.

Mit Hilfe der optimierten COSMO-Radien und dem eingeführten Parameter t konnte die Fehlerquadratsumme der Zielgrößen im Vergleich zur Optimierung nullter Ordnung um ca. 40% verringert werden. Die Standardabweichungen der physikochemischen Stoffeigenschaften sind in Tabelle 8 gelistet.

Tabelle 8: Vergleich der Standardabweichungen (SD) der drei Stoffeigenschaften. Gegenübergestellt sind die Optimierungen nullter und zweiter Ordnung.

	Optimierung nullter Ordnung	Optimierung zweiter Ordnung
$SD^{\Delta G_{hydr}}$	0.98	0.66
$SD^{\log KOW}$	0.71	0.59
$SD^{\log PV}$	0.77	0.67

An diesem Punkt wurde der Datensatz um N-, F- und Cl-haltige Verbindungen erweitert.

3.4 CHONFCI-Optimierungen

Nachdem ein erster Parametersatz für die CHO-Verbindungen ermittelt wurde, wurde in einem zweiten Schritt der Datensatz um Verbindungen erweitert, die auch Stickstoff, Fluor und Chlor beinhalten. Für alle folgenden Optimierungen wurden die Wasserstoff- und Kohlenstoff-COSMO-Radien der CHO-Optimierung ($r_H = 1.25 \text{ \AA}$ bis $r_C = 2.00 \text{ \AA}$) beibehalten. Da sich der Sauerstoffradius im Bereich von $r_O = 1.80 \text{ \AA}$ bis $r_O = 1.90 \text{ \AA}$ durch geringe Sensitivität auszeichnete, wurden die Optimierungen mit dem erweiterten Datensatz für diese beiden alternativen Sauerstoffradien durchgeführt.

Als Test, ob die Gasphasen- und COSMO-Geometrieoptimierungen in sinnvoller energetischer Beziehung stehen, bietet es sich an, die Differenzen der Energien $\Delta^X = E_{\text{gas}} - E_{\text{COSMO}}$ gegen die negativen dielektrischen Energien E_{diel} (siehe Gleichung 60 Seite 2-20) aufzutragen. Nachfolgende Abbildung zeigt den Betrag des Nettogewinns an elektrostatischer Energie Δ^X aufgetragen gegen die negative dielektrische Energie.

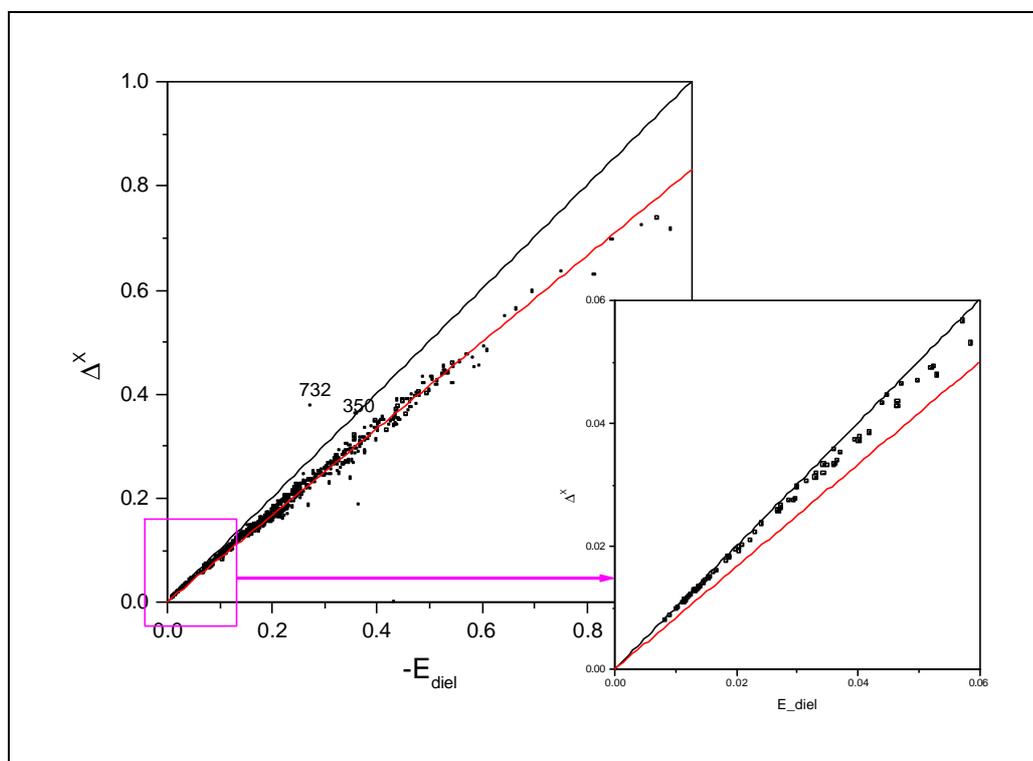


Abbildung 13: Korrelation der negativen dielektrischen Energie mit dem elektrostatischen Energiegewinn $\Delta^X = E_{\text{gas}} - E_{\text{COSMO}}$.

Analog zur von Klamt vorgestellten Korrelation [33] erhält man auch auf AM1-Niveau eine Regressionsgerade mit der Steigung 0.8. Der Korrelationskoeffizient ist $r^2 = 0.98$. Die rote

Gerade markiert die Winkelhalbierende und stellt ein Schranke dar, welche nicht überschritten werden darf. Strukturen der Verbindungen, die über dieser Schranke liegen (in diesem Fall nur die Verbindung T732 und T350), müssen überprüft werden, denn für diese würde gelten, dass nach Entfernung des Kontinuums ein Zustand erreicht werden könnte, der *stabiler als der Zustand in der Gasphase* ist. Dies ist natürlich nicht möglich, vorausgesetzt, dass bei den Gasphasenoptimierungen tatsächlich die globalen Minima lokalisiert wurden.

Wieder wurden mit der Optimierungsroutine die COSMO-Radien schrittweise ($\Delta r = 0.05 \text{ \AA}$) verändert. Von allen getesteten Radienkombinationen konnte mit $r_H = 1.25 \text{ \AA}$, $r_C = 2.00 \text{ \AA}$, $r_O = 1.90 \text{ \AA}$, $r_N = 1.98 \text{ \AA}$, $r_{Cl} = 2.15 \text{ \AA}$ und $r_F = 1.70 \text{ \AA}$ das beste statistische Ergebnis erzielt werden. Abbildung 14 zeigt die Fehler dieser Optimierung erster Ordnung.

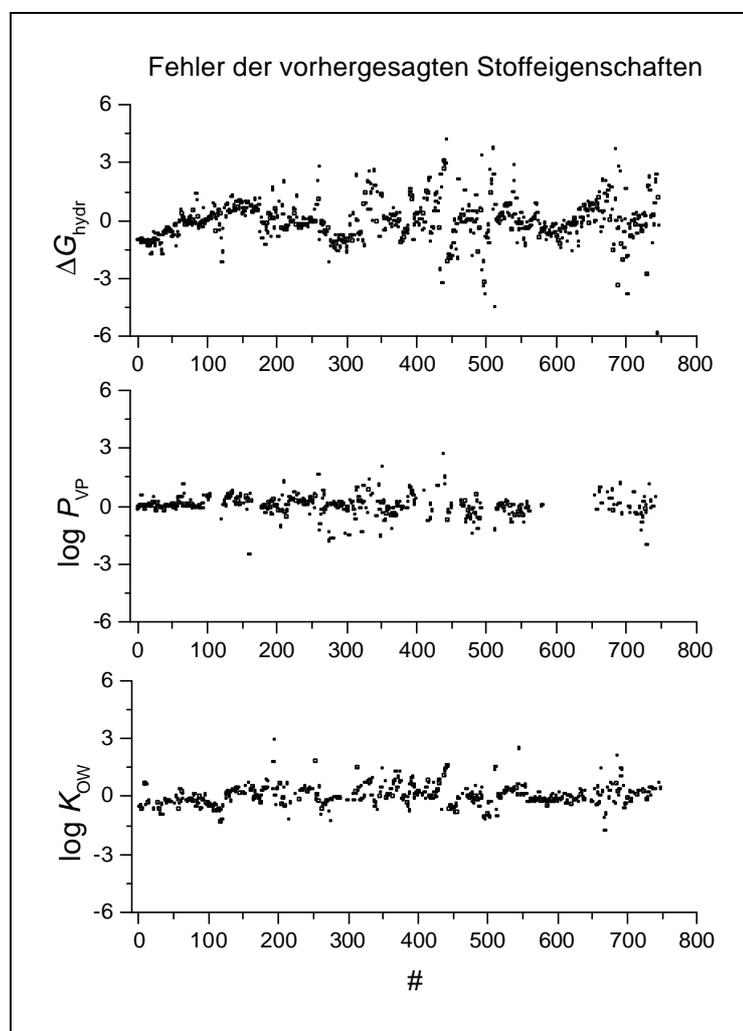


Abbildung 14: Fehler der Stoffeigenschaften aller Moleküle (# = Molekülnummer) der Optimierung erster Ordnung (Gesamtdatensatz).

Wiederum ist zu erkennen, dass die freien Hydratationsenthalpien mit den größten Fehlern behaftet sind, wohingegen die experimentellen Dampfdrücke weitaus besser reproduziert

werden. Da der Unterschied der beiden Verteilungsgrößen, das chemische Potenzial *in Lösung* ist (Wasser bzw. Reinstoff), wurde nun versucht, den Wasserstoffbrückenbindungsterm nicht wie im letzten Kapitel beschrieben skalierend für einzelne Stoffgruppen zu modifizieren, sondern grundlegend zu erweitern.

3.4.1 Lokale Grenzorbitale

Es ist bekannt, dass die Stärke der Wasserstoffbrückenbindung weniger mit den Partialladungen als vielmehr mit den Säure-Base-Eigenschaften der Atomzentren im Molekül zusammenhängt [94]. Diese Tatsache wurde von COSMO-RS bisher nur ungenügend berücksichtigt. Die Wasserstoffbrückenbindungsenergie wird umso größer, je gegensätzlicher die Ladungsdichten auf den Oberflächen der beiden Atome sind. Dies entspricht einer ausschließlich ladungsgesteuerten Wasserstoffbrückenbindungsenergie. Für die Beschreibung der Säure-Base-Eigenschaften an einzelnen Atomzentren sollten die an diesen Zentren lokalisierten Anteile der Molekülorbitale ein gutes Maß darstellen.

Dr. Klamt hat zur Bestimmung von Geschwindigkeitskonstanten für den Abbau organischer Substanzen durch Hydroxylradikale in der Gasphase erstmalig die lokalen Grenzorbital-Deskriptoren LFODs (*Local Frontier Orbital Descriptors*) eingeführt [51, 52], die auch als Maß der lokalen Polarisierbarkeit der Atomzentren verwendet werden können.

Theorie der LFODs

Die Suche nach geeigneten Deskriptoren für die Reaktivität an bestimmten atomaren Zentren veranlasste Fukui zur Vorstellung der (Super)-Delokalisierbarkeiten [53,54]. Die Superdelokalisierbarkeit – erstmals für Hückelrechnungen eingeführt – ist die Summe der MO-Ladungen q^A der besetzten π -Orbitale am Zentrum A gewichtet mit den zugehörigen inversen MO-Energien. Nach diesem Berechnungsverfahren ist die Delokalisierbarkeit ein Maß für den Anteil der Grenzorbitale an der lokalen Elektronendichte an den Atomzentren. Durch Anwendung dieses Konzepts auf semiempirische MO-Rechnungen formulierte Klamt drei neue Arten von LFODs:

- Die ladungsbegrenzte effektive HOMO-Energie (charge limited effective HOMO energy $ECH^A(q_{occ})$)
- Die energiegewichtete effektive HOMO-Energie (energy weighted effective HOMO energy $EEH^A(q_{occ})$)
- Die energiegewichtete effektive HOMO-Ladung (energy limited HOMO charge $CH^A(E)$)

Für alle drei Arten sind auch die LUMO-Analoga definiert. Der erste Deskriptor beschreibt die mittlere Energie der Ladungsmenge am Atom A. Berechnet wird $ECH^A(q_{occ})$ als gewichtete mittlere Summe der Orbitalenergien E_a , ausgehend vom HOMO zu energetisch niedrigeren Orbitalen, bis ein vorgegebener Ladungsbetrag q_{occ} erreicht wird. Formal ausgedrückt ist $ECH^A(q_{occ})$ gleich:

$$ECH^A(q_{occ}) = q_{occ}^{-1} \sum_{i=HOMO}^1 E_n g \left(q_{occ}, \underbrace{2 \sum_{j=HOMO}^{i+1} \sum_n (c_{j,n}^A)^2}_x, \underbrace{\sum_m (c_{i,m}^A)^2}_y \right) \quad \text{Gleichung 105}$$

mit

$$g(q_{occ}, x, y) = \begin{cases} y & \text{wenn } x + y < q_{occ} \\ q_{occ} - x & \text{wenn } x \leq q_{occ} \leq x + y \\ 0 & \text{wenn } x > q_{occ} \end{cases} \quad \text{Gleichung 106}$$

Griechische Indizes markieren Atomorbitale und lateinische Indizes Molekülorbitale. Großbuchstaben bezeichnen die Atomzentren. Die etwas komplizierte Formel gibt eine verhältnismäßig einfache Rechenabfolge wieder. Ausgehend vom HOMO werden die MO-Energien, gewichtet mit den doppelten quadratischen Orbitalkoeffizienten (y), aufaddiert. Die letzte zu berücksichtigende MO-Energie wird nicht mit $2 \cdot c^2$ gewichtet, sondern mit dem Restbetrag, der zur Vervollständigung der vorgegebenen Ladungsschranke q_{occ} notwendig ist: $(q_{occ} - x)$.

Soll die ladungsgewichtete effektive LUMO-Energie $EEL^A(q_{unocc})$ berechnet werden, wird ausgehend vom LUMO über alle unbesetzten Orbitale summiert.

Die energiegewichtete HOMO-Energie berechnet sich zu:

$$EEH^A(\mathbf{e}_{occ}) = \frac{\sum_{i=1}^{HOMO} \sum_n (c_{j,n}^A)^2 E_i e^{\frac{E_i}{\mathbf{e}_{occ}}}}{\sum_{i=1}^{HOMO} \sum_n (c_{i,n}^A)^2 e^{\frac{E_i}{\mathbf{e}_{occ}}}} \quad \text{Gleichung 107}$$

Jetzt werden die quadratischen Orbitalkoeffizienten über alle besetzten Orbitale summiert und mit einer exponentiellen Energiefunktion mit vorgegebenen Energiewert \mathbf{e}_{occ} gewichtet. Im Limes von $\mathbf{e}_{occ} \rightarrow 0$ geht $EEH^A(\mathbf{e}_{occ})$ gegen E_{HOMO} . Analog gilt dies für ECH für $q_{occ} \rightarrow 0$. Die

korrespondierende energiegewichtete LUMO-Energie wird durch Summation über alle unbesetzten Orbitale gewichtet mit $\exp(-E_i/e_{unocc})$ berechnet.

Schließlich kann noch die energiegewichtete HOMO-Ladung nach

$$QH(E_{occ})^A = 2 \sum_{HOMO}^1 \sum_n (c_{i,n}^A)^2 W(E_i, E_{occ}) \quad \text{Gleichung 108}$$

mit

$$W(E_i, E_{occ}) = \begin{cases} 1.0 & \text{wenn } E_i > E_{occ} + 0.5 \\ E_i - E_{occ} + 0.5 & \text{wenn } E_{occ} + 0.5 \geq E_i \geq E_{occ} - 0.5 \\ 0.0 & \text{wenn } E_i < E_{occ} - 0.5 \end{cases} \quad \text{Gleichung 109}$$

berechnet werden. Der Gewichtungsfaktor $W(E_i, E_{occ})$ garantiert den linearen Übergang des Grenzwichtungsfaktors zwischen 0 eV und 1 eV. Zur Berechnung des zugehörigen $QL^A(E_{unocc})$ -Deskriptors werden die Koeffizienten aller unbesetzten Orbitale aufsummiert.

Um zu klären ob die Grenzorbitaldeskriptoren die Elemente auf natürliche Weise nach verschiedenen funktionellen Gruppen aufschlüsseln, und damit eventuell auch zur Korrektur systematischer Abweichungen geeignet sind, wurden nach Elementen sortierte LFOD-Statistiken angefertigt. Hierzu wurde ein Programm entwickelt, mit dem die LFODs nach Extraktion der Eigenvektoren und Eigenenergien aus den MOPAC-Ausgabedateien automatisch berechnet werden.

Die nächste Abbildung zeigt eine Gesamtstatistik der LFODs aller im Trainingsdatensatz vorhandenen Wasserstoffatome mit den vorgegebenen Grenzwerten: $q_{occ} = q_{unocc} = 0.6$ e, $e_{occ} = -2.5$ eV, $e_{unocc} = 2.5$ eV, $E_{occ} = -15.0$ eV und $E_{unocc} = 3$ eV. Es zeigt sich, dass die LFODs mit den verwendeten Grenzbedingungen für eine Anwendung im Wasserstoffbrückenbindungsterm prinzipiell geeignet erscheinen, da sich die an Sauerstoff und Stickstoff gebundenen Wasserstoffatome (Nr. 6639-6857) durch die ECH- und QH-Deskriptoren (siehe Abbildung 15 nächste Seite) deutlich von den an Kohlenstoffen gebundenen Wasserstoffen separieren lassen. Dadurch ist es beispielsweise möglich durch Verwendung eines Schwellwertes nur die an Heteroatome gebundenen Wasserstoffe zu erfassen.

Lokale Grenzorbital-Deskriptoren der Wasserstoffatome

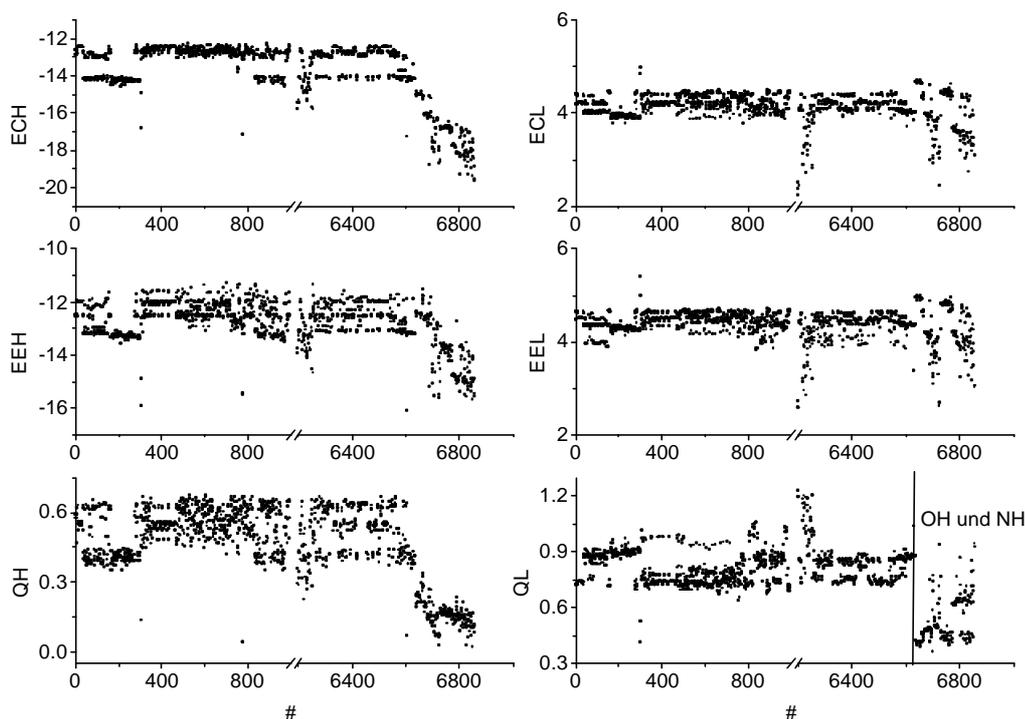


Abbildung 15: Lokale Grenzorbital-Deskriptoren aller Wasserstoffatome im Trainingsdatensatz. An N und O gebundene Wasserstoffe ($\# > 6639$) können durch die LFODs identifiziert werden.

Für eine Verbesserung des Wasserstoffbrückenbindungsterms sind die lokalen Grenzorbital-deskriptoren an den Kohlenstoffatomen nur von untergeordneter Bedeutung, da Kohlenstoff bekanntermaßen keine Wasserstoffbrückenbindung eingeht. Dennoch ist die Statistik aufschlussreich. So zeigt Abbildung 16 eine deutliche Separation von sp^2 - und sp^3 -hybridisierten Kohlenstoffatomen bei den ECH/ECL- und EEH/EEL-Deskriptoren. Falls sich bei der Optimierung systematische Fehler dieser Stoffgruppen ergeben, könnten die LFODs als zusätzliche natürliche Parameter zum Fehlerausgleich herangezogen werden.

Auch bei den Sauerstoffatomen (siehe Abbildung 17) sind die LFODs Parameter, mit denen verschiedene funktionelle Gruppen und reaktive Zentren identifiziert werden können. Dies eröffnet ein neues Anwendungsfeld. Anstatt die Anzahl von Bindungen und die Art und Anzahl von Bindungspartnern zur Identifikation von funktionellen Gruppen zu verwenden, könnten hierzu auch die verschiedenen lokalen Grenzorbitaldeskriptoren verwendet werden.

Lokale Grenzorbital-Deskriptoren der Kohlenstoffatome

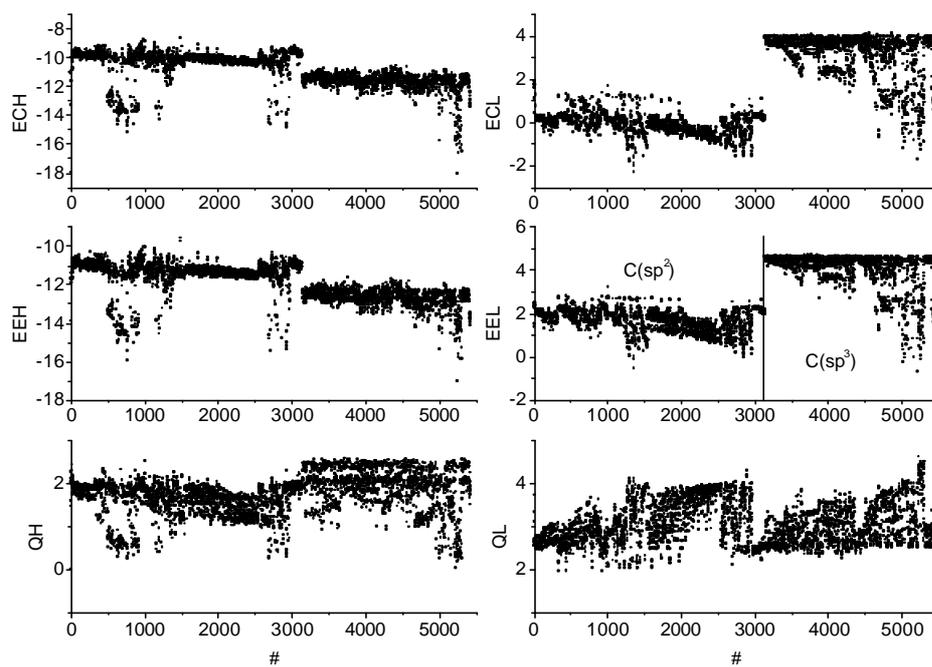
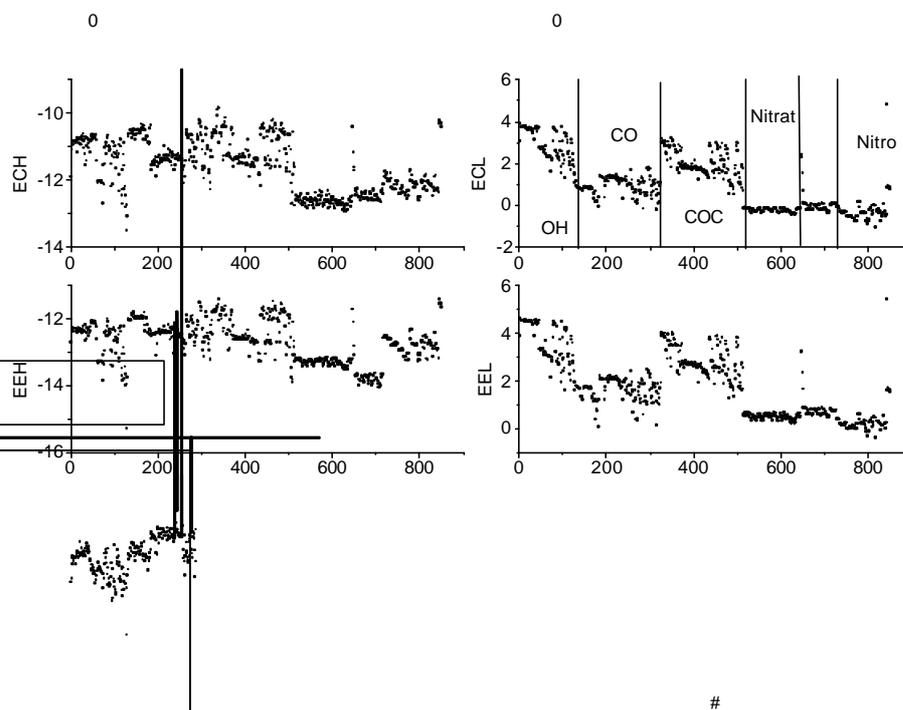


Abbildung 16: Lokale Grenzorbitaldeskriptoren aller Kohlenstoffatome im Trainingsdatensatz.



Der ECL-Deskriptor der Sauerstoffe steigt von den Nitro- und Nitratfunktionen über Carbonyl- bis zu Ether- und Hydroxygruppen. Für den Ausgleich eventueller systematischer Fehler einzelner Stoffgruppen könnten also auch die LFODs der Sauerstoffatome verwendet werden.

Im Falle des Stickstoffs sollten sich vor allem die ECH-Deskriptoren zur Stoffklassenunterscheidung eignen. Hier separieren sich Stickstoffatome von Nitro- und Nitratverbindungen deutlich von Stickstoffatomen in Cyano- und Amifunktionen.

Lokale Grenzorbital-Deskriptoren der Stickstoffatome

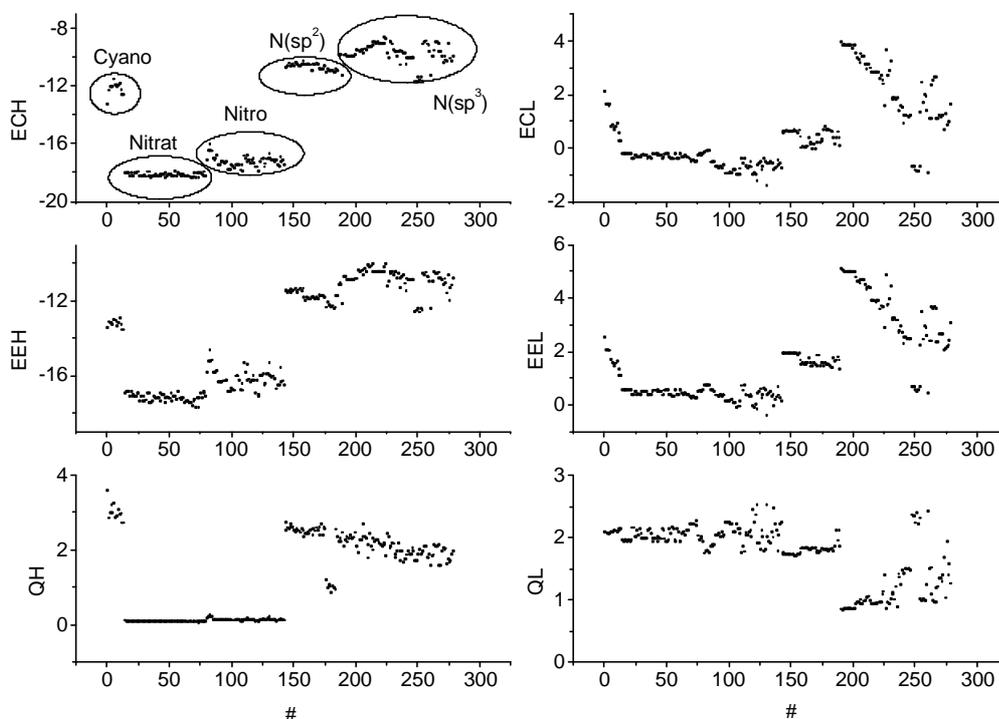


Abbildung 18: Lokale Grenzorbitaldeskriptoren an den Stickstoffzentren des Trainingsdatensatzes.

Obwohl mehrere Strategien zur Verbesserung des COSMO-RS-Wasserstoffbrückenbindungsterms verfolgt wurden, konnte mit Hilfe der lokalen Grenzorbitaldeskriptoren keine entscheidende Verbesserung herbeigeführt werden. Weder als additiver Beitrag der Deskriptoren L (Gleichung 110)

$$E_{hb}(\mathbf{s}, \mathbf{s}', L, L') = c_{hb} \max[0, \mathbf{s}_{acc} - \mathbf{s}_{hb}] \cdot \min[0, \mathbf{s}_{don} + \mathbf{s}_{hb}] + d_{hb} (L_{don} \cdot L'_{acc}) \quad \text{Gleichung 110}$$

noch als skalierender Beitrag zum bestehenden Wasserstoffbrückenbindungsenergieterm

$$E_{hb}(\mathbf{s}, \mathbf{s}', L, L') = c_{hb} (L_{don} \cdot L'_{acc}) \max[0, \mathbf{s}_{acc} - \mathbf{s}_{hb}] \cdot \min[0, \mathbf{s}_{don} + \mathbf{s}_{hb}] \quad \text{Gleichung 111}$$

konnten die Grenzorbitale die Statistik verbessern. In den Gleichungen steht L allgemein für die lokalen Grenzorbitaleskriptoren (z. B. $L_{don} = \text{ECL}$; $L_{acc} = \text{ECH}$). Neben den in den Gleichungen angegebenen multiplikativen Ansätzen ($L_1 \cdot L_2$) wurden auch Differenzen ($L_1 - L_2$) gebildet. Auch die spezielle Berücksichtigung für nur ein Element brachten keine Verbesserungen.

Natürlich könnten die Deskriptoren auch außerhalb des H-Brückenbindungsterms zur Skalierung der chemischen Potenziale analog dem bei den CHO-Verbindungen eingeführten Skalierungsfaktor τ verwendet werden, doch soll dies nicht Gegenstand der vorliegenden Untersuchung sein, da hier versucht wurde, die Natur der Wasserstoffbrückenbindung im COSMO-RS-Bild zu verbessern.

Es bleibt festzuhalten, dass sich Atome in unterschiedlichen Funktionalitäten mittels lokaler Grenzorbitaleskriptoren zwar separieren lassen, aber mit den getesteten Ansätzen weder systematische noch nicht-systematische Stoffgruppenfehler in COSMO-RS beheben lassen. Deshalb wird im nachfolgenden Kapitel eine andere Art der Modifikation vorgeschlagen, die nicht am Wasserstoffbrückenbindungsterm ansetzt.

3.4.2 Core-Ausgleichsladungen

Ein völlig anderer Ansatz zur Verbesserung der Statistik ist es, nicht zu versuchen einen neuen COSMO-RS-spezifischen Parameter einzuführen, sondern schon bei den zugrundeliegenden AM1-Rechnungen Modifikationen vorzunehmen. Durch gezielte Veränderung der Ladungsverhältnisse innerhalb des Moleküls können die resultierenden COSMO-Abschirmladungsdichten auf der SAS und damit die Basis der nachfolgenden RS-Rechnungen beeinflusst werden. Betrachtet man nochmals Abbildung 14 auf Seite 3-21, so fällt unter anderem die unbefriedigende Vorhersage der freien Hydratationsenergien von Nitro- und Nitratverbindungen (#450 - #515) auf. Dieses Resultat ist nicht unerwartet, ist doch hinlänglich bekannt, dass mit semiempirischen AM1-Rechnungen die Bildungsenthalpien und Dipolmomente von Nitroverbindungen nur unbefriedigend reproduziert werden können. Das Dipolmoment gibt in erster Näherung die Ladungsverteilung im Molekül wieder. Ist also das berechnete Dipolmoment eines Moleküls ungenau, so ist die gesamte Ladungsverteilung wahrscheinlich ungenau beschrieben, und auch die Abschirmladungsdichten auf der SAS werden verfälscht. Es liegt der Versuch nahe, die Ladungsverteilung im Molekül bei den

AM1-Rechnungen zu modifizieren und das berechnete Dipolmoment als Qualitätskriterium der Veränderung zu verwenden.

Die wohl einfachste Methode eine Veränderung der Ladungsverteilung zu erreichen, ist das Hinzufügen zusätzlicher Ladungen auf die Atomzentren. Die gesamte addierte Ladungsmenge muss null sein, da die Ladungsneutralität des Moleküls bewahrt werden muss. Im Falle der Nitrogruppen wird also zur Core-Ladung (= Kernladung + Ladung der Rumpfelektronen) des Stickstoffs eine Ladung q und zu den Core-Ladungen der Sauerstoffe jeweils die Ladung $-q/2$ addiert.

Zur praktischen Bestimmung der Ausgleichsladungen wurden sechs Nitroverbindungen mit bekannten Gasphasendipolmomenten ausgewählt und mit Ausgleichsladungen versehen, deren Betrag systematisch verändert wurde. Die Standardabweichung wird mit Stickstoffzusatzladungen $q_N = -0.32$ e bis -0.34 e ($e = \text{Elementarladung}$), und Sauerstoffzusatzladungen $q_O = 0.16$ e bis 0.17 e minimal (siehe Tabelle 9).

Für die Rechnungen wurde das MOPAC-Unterprogramm *dipol.f* derart modifiziert, dass eine Datei mit den Ladungen eingelesen werden konnte.

Tabelle 9: Experimentelle (EXP) und mit AM1 berechnete Gasphasendipolmomente [Debye] mit Standardabweichung SD und mittlerem Fehler (MF).

	EXP	AM1	AM1 mit $q_N = -0.25$ $q_O = 0.125$	AM1 mit $q_N = -0.30$ $q_O = 0.15$	AM1 mit $q_N = -0.31$ $q_O = 0.155$	AM1 mit $q_N = -0.32$ $q_O = 0.16$	AM1 mit $q_N = -0.33$ $q_O = 0.165$	AM1 mit $q_N = -0.34$ $q_O = 0.17$
NO ₂	0.32	0.41	0.11	0.21	0.24	0.26	0.28	0.30
Nitromethan	3.46	4.17	3.48	3.34	3.31	3.28	3.25	3.23
Nitroethan	3.23	4.39	3.69	3.56	3.53	3.50	3.47	3.45
Nitrobenzol	4.22	5.24	4.54	4.40	4.37	4.34	4.32	4.29
1-Nitropropan	3.66	4.50	3.80	3.67	3.64	3.61	3.58	3.56
2-Nitropropan	3.73	4.54	3.85	3.71	3.68	3.66	3.60	3.60
MF		-0.77	-0.14	-0.05	-0.03	0.01	0.02	0.03
SD		0.92	0.28	0.18	0.17	0.16	0.17	0.16

Wie man leicht erkennt, werden alle Dipolmomente mit den AM1-Rechnungen systematisch überschätzt. Interessanterweise unterscheidet sich Stickstoffdioxid von den übrigen Nitroverbindungen: Während die Dipolmomente der übrigen Verbindungen bei Erhöhung der Ausgleichsladungen sich stetig verringern, zeigte das Dipolmoment von NO₂ bei einer Stickstoffzusatzladung von $q_N = -0.25$ e ein Minimum. Schließlich wurden im weiteren die Ladungen $q_N = -0.32$ e und $q_O = 0.16$ e verwendet, da hier das Mittel der Fehlersummen (MF)

minimal ist. In der nachfolgenden Abbildung sind experimentelle gegen berechnete Dipolmomente aufgetragen.

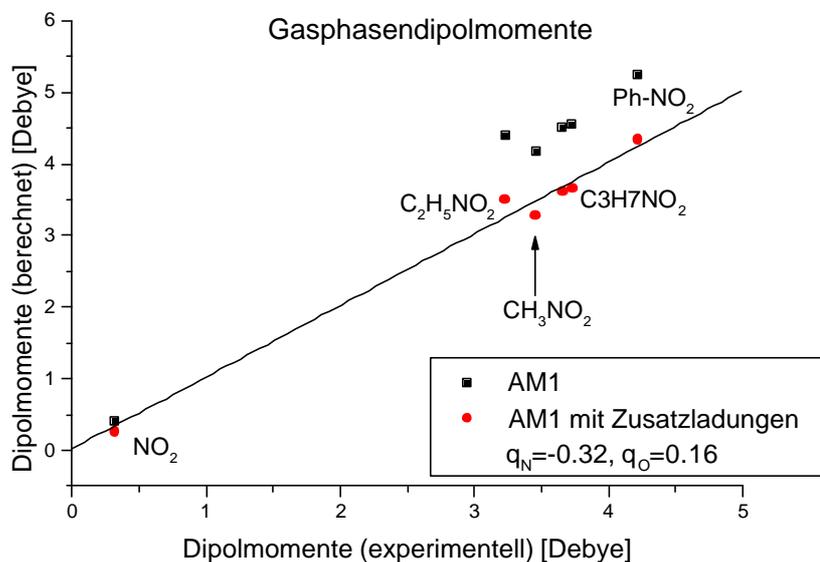


Abbildung 19: Auftragung experimenteller und berechneter Gasphasendipolmomente.

Eine größere Anzahl an experimentellen Daten wäre wünschenswert, um größere statistische Genauigkeit zu erzielen, da es sich hier um einen rigorosen Ausgleich handelt, der später für *alle* Nitroverbindungen angewendet werden soll. Deshalb wurde die Qualität der Core-Ausgleichsladungen anhand von 12 Nitroaromaten bezüglich der mit ab initio-Methoden berechneten Dipolmomente noch einmal überprüft. Tabelle 10 (Seite 3-6) zeigt AM1-Dipolmomente, mit und ohne Ausgleichsladungen und im Vergleich dazu verschiedene ab initio-Dipolmomente. Zusätzlich sind für die ab initio-Methoden jeweils die Standardabweichungen zu den AM1-Rechnungen angegeben. Es zeigt sich eine bessere Übereinstimmung der modifizierten AM1-Dipolmomente mit DFT/B3LYP- und MP2-Dipolmomenten. So bewirkten die Zusatzladungen eine Verkleinerung der Standardabweichung um 0.16 Debye bezüglich der B3LYP-Dichtefunktionalrechnungen, und eine Verbesserung von 0.78 Debye bezüglich des störungstheoretischen MP2-Niveaus. Dagegen werden die BP-Dipolmomente (BP = B88-PW91-Funktional) von AM1 ohne Ausgleichsladungen statistisch genauer reproduziert. Die große Standardabweichung bezüglich der AM1-Methode mit Ausgleichsladungen wird hauptsächlich durch das sehr große BP-Dipolmoment von 3-Nitro-o-Phenylendiamin verursacht.

Tabelle 10: Berechnete Gasphasendipolmomente [Debye] von 12 Nitroaromaten

#	Substanz	Exp	AM1	AM1*	B3LYP ¹	BP ²	MP2 ³
1	Nitrobenzol	4.22	5.20	4.34	4.60	4.58	4.05
2	1-Chlor-2-Nitrobenzol	4.64	5.40	4.52	5.00	4.67	4.54
3	1-Chlor-3-Nitrobenzol	3.73	4.70	3.80	3.80	3.93	3.41
4	1-Chlor-4-Nitrobenzol	2.83	4.10	3.23	2.90	3.35	2.38
5	3-Nitroanilin		5.90	5.08	5.70	6.04	4.99
6	4-Nitroanilin		7.30	6.40	7.20	7.76	6.64
7	p-Nitrotoluol		5.70	4.83	5.20	5.40	4.54
8	3,4-Dichlornitrobenzol		3.80	2.92	2.70	3.10	2.21
9	3-Nitro-o-Phenylendiamin		5.60	4.69	5.60	6.48	4.99
10	2-Nitro-4-Chloranilin		5.30	4.58	5.00	4.75	4.50
11	2-Chlor-4-Nitrotoluol		4.40	3.50	3.30	3.73	2.77
12	4-Nitro-2-Methylanilin		7.40	6.47	7.30	7.26	6.27
Mittelwerte			5.40	4.53	4.86	5.09	4.27
SOR ^{AM1} (SD)					5.59 (0.71)	4.43 (0.63)	16.77 (1.23)
SOR ^{AM1*} (SD)					3.34 (0.55)	7.13 (0.80)	2.29 (0.45)

*mit Core-Ausgleichsladungen $q_N=-0.32$ und $q_O=0.16$; ¹6-31G** (Gaussian98); ²DNP (DMOL96) ³6-31G** (Gaussian94)

Zur praktischen Anwendung der Ausgleichsladungen in den Optimierungen mussten die MOPAC-Quelltexte abgeändert werden. Nach der Initialisierung der COSMO-Rechnungen im Unterprogramm *rmopac.f* wurde ein neu entwickeltes Unterprogramm *readcar.f* aufgerufen, mit dem es möglich ist, verschiedene im Moleküldatensatz vorkommende funktionelle Gruppen automatisch zu erkennen. Grundlage hierfür bilden die kartesischen Koordinaten der Atome. Nach einer Bestimmung der Art und Anzahl der Bindungspartner ist es möglich, den einzelnen Atomen gezielt unterschiedliche Core-Ausgleichsladungen zuzuweisen. Dies geschieht mit Hilfe eines Vektors (*dipvec*), welcher in einer Common-Anweisung an das COSMO-Unterprogramm *addnuc* übergeben wird. Hier wird dieser zum *core*-Vektor addiert.

Obwohl die Ausgleichsladungen mit Hilfe von Gasphasenrechnungen bestimmt wurden, erfolgt der Ladungsausgleich nun also *bei den COSMO-Rechnungen*, die Gasphasenrechnungen bleiben dagegen unberührt. Es wird also eine Veränderung der Elektrostatik relativ zur Gasphase bewirkt. Grund für die Entscheidung war, dass das *allen* semiempirischen Methoden gemeine Prinzip festgesetzter Core-Ladungen (= Kernladung + Ladung der Rumpfelektronen) nicht verletzt wird, wohingegen in Kontinuumsmodellen die Veränderung

der Ladungsverteilung üblich ist. So werden beispielsweise bei den SMx-Solvatationsmodellen nach Cramer und Truhlar die sogenannten „Class IV-Charges“ [78,79] verwendet.

Wie in Abbildung 19 (Seite 3-30) gezeigt, eignet sich die Core-Ladungsmodifikation einzelner funktioneller Gruppen aber prinzipiell sehr wohl zur Verbesserung der AM1-Methode hinsichtlich der Vorhersage von Dipolmomenten.

Nachfolgende Graphik zeigt die absoluten Fehler der freien Hydratationsenthalpien, welche mit den Ausgleichsladungen $q_N = -0.32 e$ und $q_O = 0.16 e$ für die Nitrogruppen ermittelt wurden:

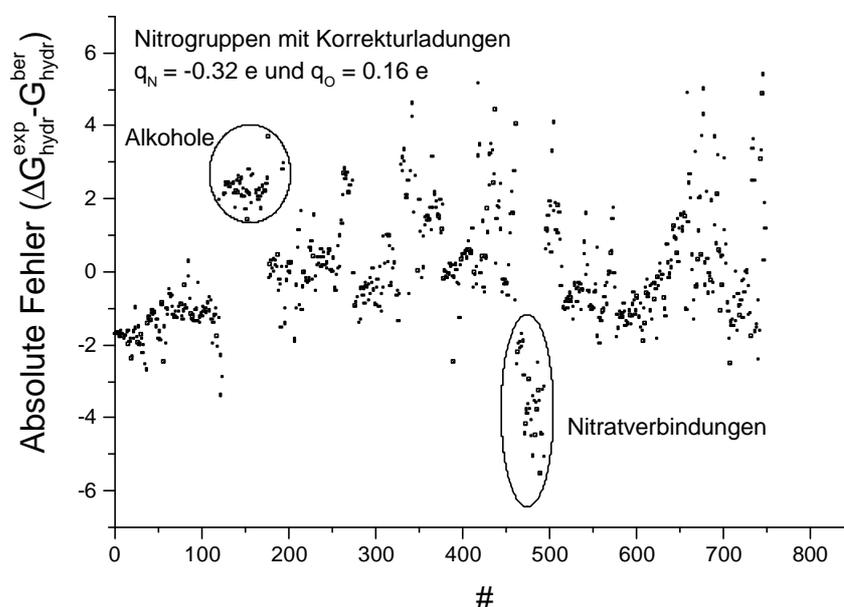


Abbildung 20: Absolute Fehler der freien Hydratationsenthalpie (mit Core-Ausgleichsladungen für Nitroverbindungen). Salpetersäureester und Alkohole (und andere Stoffgruppen) zeigen systematische Abweichungen.

Die Summe der Fehlerquadrate aller Zielgrößen ist 1513. Dies entspricht einer Erhöhung um 8% im Vergleich zur Optimierung erster Ordnung (SOR = 1401)! Deutlich erkennbar in Abbildung 20 ist aber die systematische Abweichung der berechneten freien Hydratationsenthalpien der Salpetersäureester. Deshalb wurden in einem zweiten Schritt die Core-Ausgleichsladungen ebenso auf die Nitratverbindungen aufgebracht. Formal können Salpetersäureester auch als Nitroverbindungen angesehen werden, nur ist die Nitrogruppe nun nicht an ein Kohlenstoffatom, sondern an ein weiteres Sauerstoffatom gebunden. Abbildung 21 zeigt wiederum die absoluten Fehler der Hydratationsenthalpien.

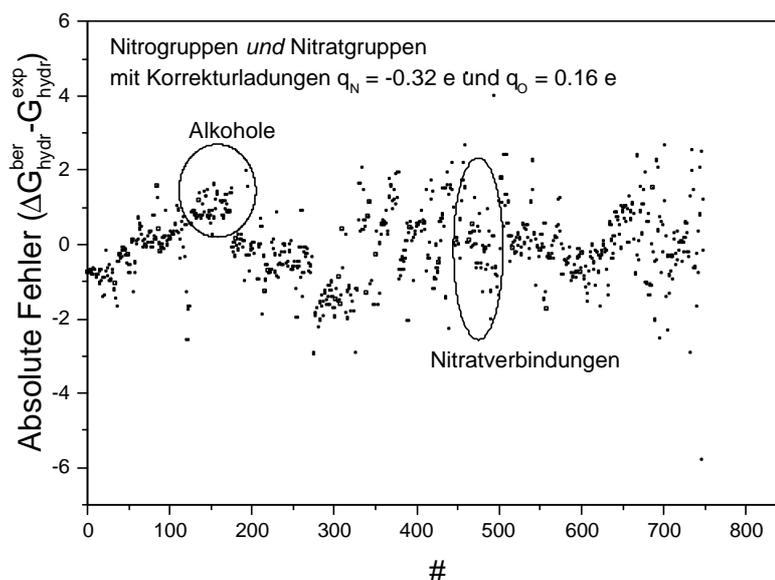


Abbildung 21: Absolute Fehler der freien Hydratationsenthalpien (mit Core-Ausgleichsladungen für Nitro- und Nitratverbindungen).

Es ist zu erkennen, dass das Konzept der Core-Ausgleichsladungen nun mit Anwendung auf Nitratverbindungen *und* Nitroverbindungen zu einer verbesserten Statistik führt, da der systematische Fehler der Nitrate nun ebenfalls ausgeglichen wurde. Die Fehlerquadratsumme beträgt nunmehr nur noch 1210, was einer Verringerung um 13.6% im Vergleich zur Optimierung erster Ordnung mit identischen COSMO-Radien entspricht. Eine genaue Analyse zeigt, dass eine Verbesserung bei allen drei Stoffeigenschaften erzielt wurde. So verringern sich die Fehlerquadratsummen für ΔG_{hydr}^- , $\log P_V^-$, und $\log K_{\text{OW}}$ um 14.1 %, 9.0 % beziehungsweise 16.3 %.

Der Einfluss der Core-Ausgleichsladungen auf die optimierten COSMO/AM1-Strukturen ist gering. Bindungslängen ändern sich nur um ca. 0.002 Å und ONO-Winkel um ca. 1°. Dies ist natürlich wünschenswert, denn eine Verbesserung der elektrostatischen Verhältnisse auf Kosten einer extremen strukturellen Verzerrung soll vermieden werden. Nachfolgende Abbildung zeigt einen Vergleich der COSMO/AM1-Nitromethanstrukturen.

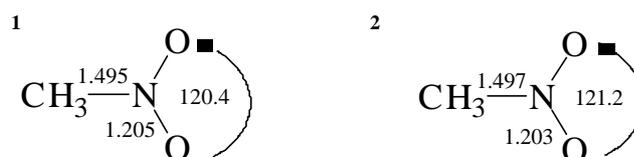


Abbildung 22: Mit COSMO/AM1 optimierte Nitromethanstrukturen. Ohne (1) und mit Core-Ausgleichsladungen (2).

Auch ein Vergleich der Orbitalenergien zeigt nur geringe Unterschiede zwischen COSMO/AM1-Rechnungen und COSMO/AM1-Rechnungen mit Ausgleichsladungen.

Die folgende Abbildung zeigt schematisch die Orbitalenergien der zwei höchsten besetzten und der vier niedrigsten unbesetzten COSMO/AM1-Orbitale im Vergleich zur Gasphasenrechnung.

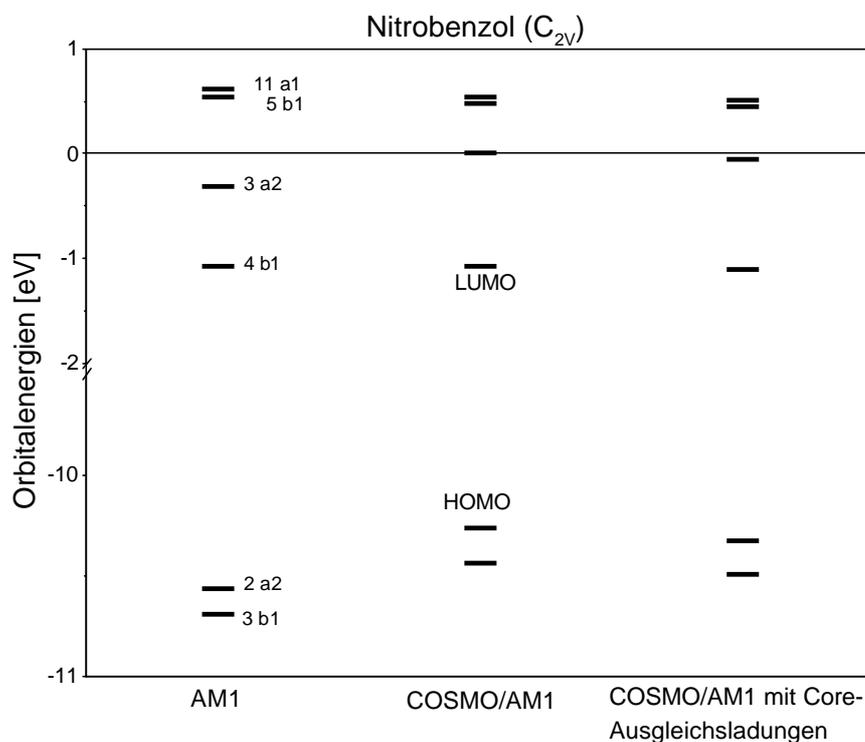


Abbildung 23: Orbitalenergien von Nitrobenzol: Gasphase im Vergleich zu COSMO/AM1 und COSMO/AM1 mit Core-Ausgleichsladungen.

Es zeigt sich, dass die durch die Ausgleichsladungen hervorgerufene Verschiebung der Orbitalenergien nur sehr klein ist. Die energetische Verschiebung beim Nitrobenzol ist systematisch: Alle Orbitalenergien verringern sich durch Addition der zusätzlichen Core-Ladungen. Interessanterweise verliert das zweitniedrigste unbesetzte Molekülorbital (3a2) seinen antibindenden Charakter und wird analog zur Gasphasenrechnung bindend.

Wegen der systematischen Verschiebung bleibt auch die HOMO-LUMO-Energiedifferenz nahezu konstant.

Wegen der verbesserten Fehlerstatistik für ΔG_{hydr} , $\log P_V$ und $\log K_{\text{OW}}$ ist nicht bewiesen, dass die verwendeten Core-Ausgleichsladungen tatsächlich zu einer besseren Beschreibung der elektronischen Verhältnisse beigetragen haben. Es ist ebenso möglich, dass Defizite des COSMO-RS/AM1-Modells nur durch eine interne Fehlerkompensation ausgeglichen wurden.

Deshalb wurden die COSMO/AM1-Dipolmomente (mit und ohne Ausgleichsladungen) mit COSMO/DFT-B3LYP-Dipolmomenten verglichen (Abbildung 24). Es zeigt sich, dass die Dipolmomente der AM1-Rechnungen mit Core-Ausgleichsladungen tatsächlich näher an den COSMO/B3LYP-Dipolmomenten liegen. Es ergeben sich Standardabweichungen von 0.62 (AM1 mit Zusatzladungen) beziehungsweise 0.66 Debye (AM1). Natürlich sollte beachtet werden, dass die DFT-Dipolmomente auf Basis anderer van der Waalscher Radien berechnet wurden.

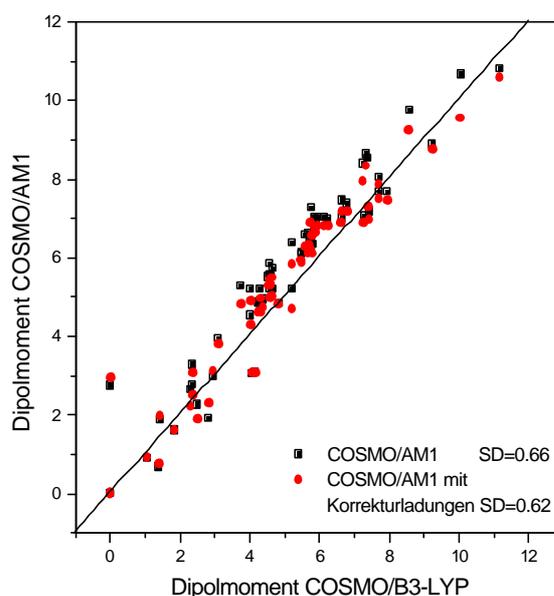


Abbildung 24: Vergleich von COSMO/DFT- und COSMO/AM1-Dipolmomenten [Debye] der Nitro- und Nitratverbindungen im Trainingsdatensatz.

Die Ausgleichsladungen wurden mit einem sehr kleinen Satz experimenteller *Gasphasendipolmomente* ermittelt, und deshalb wurde erwartet, dass die Ladungen nicht zwingend auch die optimalen Ladungen für die *COSMO*-Rechnungen sind. Es liegt deshalb nahe, die Core-Ausgleichsladungen zu variieren. Abbildung 21 (Seite 3-33) lässt vermuten, dass die verwendeten Dipolmomentausgleichsladungen vom Betrag zu groß gewählt wurden, da die Nitro- und Nitratverbindungen von negativen Abweichungen der freien Hydratationsenthalpien zu leicht positiven Abweichungen, also einer Überschätzung (vor allem im Vergleich zu Säuren und Estern #262 - #326), wechselten. Daher wurde der Betrag der Zusatzladungen wieder schrittweise reduziert. Nachfolgende Tabelle zeigt die Fehlerquadratsummen der Stoffeigenschaften mit den verwendeten Core-Ausgleichsladungen.

Tabelle 11: Fehlerquadratsummen (SOR) der berechneten physikochemischen Stoffeigenschaften für verschiedene Core-Ausgleichsladungen (q_N = Stickstoffausgleichsladung, q_O = Sauerstoffausgleichsladung) für Nitro- und Nitratgruppen.

q_N [e]	q_O [e]	SOR
-0.32	0.16	1226
-0.28	0.14	1196
-0.24	0.12	1172
-0.20	0.10	1158
-0.16	0.08	1153
-0.12	0.06	1166

Im Bereich um $q_N = -0.16$ e wird die Fehlerquadratsumme minimal, und damit sind die optimalen Ausgleichsladungen für die Nitro- und Nitratgruppen gefunden.

Nachdem die Methode der Core-Ausgleichsladungen für Nitro- und Nitratverbindungen erfolgreich angewendet wurde, wurde geprüft, ob mit dieser Methode auch das Problem der systematischen Abweichungen der Hydroxyverbindungen (vgl. Abbildung 20 Seite 3-32, und Abbildung 21 Seite 3-33) behoben werden kann.

Da die Vorooptimierung der Ausgleichsladungen an dem kleinen Nitrodatensatz mit Hilfe der experimentellen Gasphasendipolmomente nicht zu den schließlich optimalen Core-Ausgleichsladungen führte, wurden die Ladungen für die Hydroxylgruppen ganz pragmatisch durch schrittweise Variation ermittelt. Eine COSMO/AM1-Rechnung mit den Zusatzladungen $q_{O(OH)} = -0.03$ e und $q_{H(OH)} = 0.03$ e verbesserte die Statistik noch einmal signifikant: Die Summe der Fehlerquadrate sinkt auf 1051, dies entspricht einer Verringerung von 8.8% im Vergleich zu den Ergebnissen mit den optimierten Core-Ladungen für Nitro- und Nitratgruppen. Im Detail erhält man um 13%, 2.9% und um 4.4% verbesserte Fehlerquadratsummen für ΔG_{hydr} , $\log K_{OW}$ und $\log P_V$. Die Standardabweichungen sind $SD^{\Delta G_{hydr}} = 0.89$, $SD^{\log K_{OW}} = 0.74$ und $SD^{\log P_V} = 0.70$. In der folgenden Tabelle sind die Ausgleichsladungen mit den übrigen optimierten COSMO-RS-Parametern aufgeführt.

Tabelle 12: COSMO-RS-Parameter der Optimierung mit Nitro-, Nitrat- und Hydroxy-Ausgleichsladungen.

Parameter	Wert	Parameter	Wert
γ_H	-0.0319 kcal/(mol Å ²)	γ_N	-0.0628 kcal/(mol Å ²)
γ_C	-0.0444 kcal/(mol Å ²)	γ_{Cl}	-0.0501 kcal/(mol Å ²)
γ_O	-0.0441 kcal/(mol Å ²)	γ_F	-0.0176 kcal/(mol Å ²)
r_{av}	0.38 Å		
β	0.1239 kcal/(mol Å ²)	c_{mf}	2.6
λ	0.17	f_{corr}	2.4
ω	-0.1877 kcal/mol	c_{hb}	41500 kcal/(mol Å ²)/e ²
η	-7.55	σ_{hb}	0.0020 e/Å ²

Die nächste Abbildung zeigt die Fehler für die einzelnen Zielgrößen in graphischer Form.

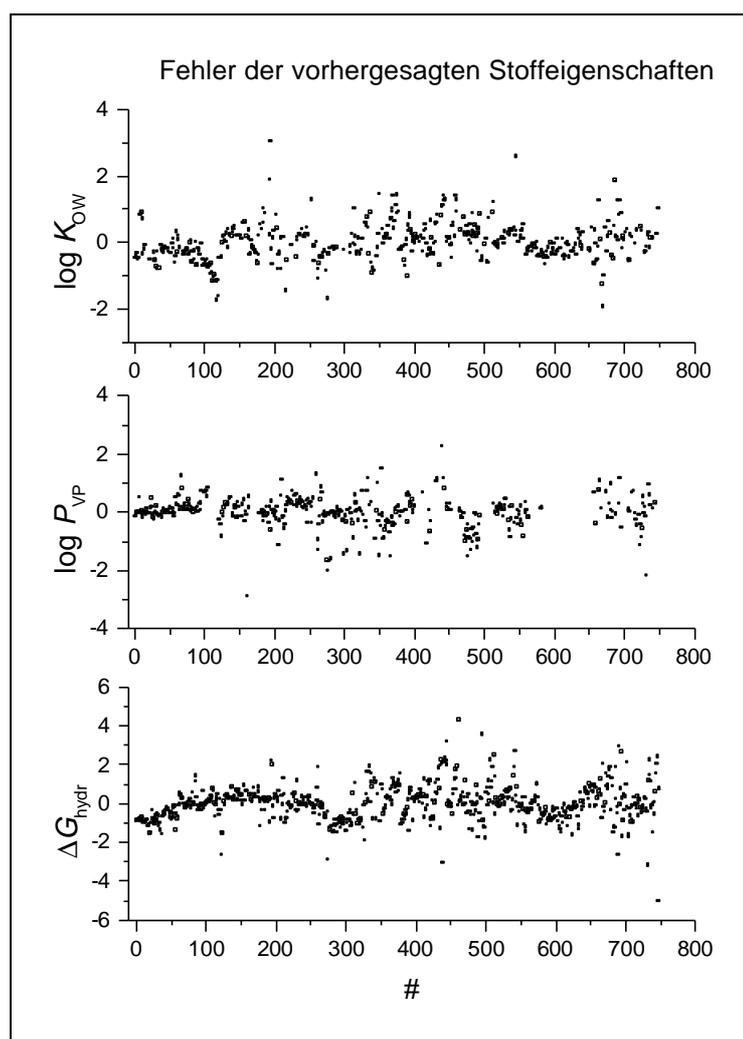


Abbildung 25: Absolute Fehler für ΔG_{hydr} , $\log K_{OW}$ und $\log P_V$ RS-Rechnungen basierend auf COSMO/AM1-Rechnungen mit Ausgleichsladungen für Nitro-, Nitrat- und Hydroxygruppen.

Ein Vergleich mit Abbildung 21 (Seite 3-33) offenbart die verbesserte Vorhersage für die Alkohole. Aus den Abbildungen ist aber auch ersichtlich, dass die freien Hydratationsenthalpien und die Oktanol/Wasser-Verteilungskoeffizienten der reinen Kohlenwasserstoffe - und hier vor allem die der Alkane (#1 bis #37) - systematische Abweichungen zeigen. Auch hier wurde deshalb versucht, eine Ergebnisverbesserung mit Hilfe von Core-Ausgleichsladungen zu erzielen. Ungeachtet der Tatsache, ob eine Verbesserung der Statistik mit dieser Methode zu erzielen ist, ist es interessant, ob die Ursache wiederum die mit AM1 mangelhaft beschriebene Ladungsverteilung im Molekül ist. Ein Vergleich von AM1- und DFT-Dipolmomenten bestätigt diesen Befund:

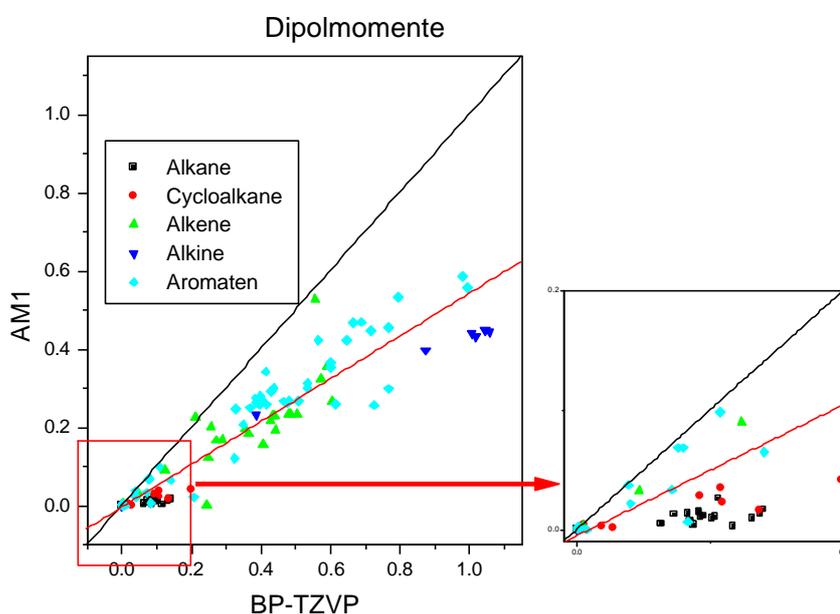


Abbildung 26: Korrelation von AM1- und DFT-Dipolmomenten. (DFT mit BP-Funktional und TZVP-Basis)

In Abbildung 26 sind die AM1-Dipolmomente der reinen Kohlenwasserstoffe des Trainingsdatensatzes mit den DFT-Dipolmomenten korreliert. Es zeigt sich, dass mit AM1 im Schnitt um ca. 45 % kleinere Dipolmomente als mit DFT ermittelt werden (rote Regressionsgerade). Vor allem bei den Alkanen (diese haben sehr kleine Dipolmomente) ist die Abweichung im Vergleich zu den übrigen Kohlenwasserstoffen noch einmal vergrößert. Der Grund für die systematische Abweichung ist also wiederum die ungenaue AM1-Ladungsverteilung.

In einem ersten Schritt wurden Ausgleichsladungen ausschließlich auf sp^3 -hybridisierte Kohlenstoffatome und auf an diese gebundene Wasserstoffatome *reiner* CH-Verbindungen aufgebracht. Die Statistik verbesserte sich daraufhin noch einmal signifikant. Für das folgende

Set von nun erweiterten Ausgleichsladungen ergibt sich eine Fehlerquadratsumme von $SOR = 1007$.

Tabelle 13: Core-Ausgleichsladungen der besten COSMO-RS/AM1-Optimierung. N = Anzahl der am C-sp³-Atom gebundenen H-Atome.

Position der Ladung q	Ladung [e]
$q_{N(\text{Nitro})}$	-0.16
$q_{O(\text{Nitro})}$	0.08
$q_{H(\text{OH})}$	-0.05
$q_{O(\text{OH})}$	0.05
$q_{C(\text{CH})}$	$N^* q_{H(\text{CH})} $
$q_{H(\text{CH})}$	-0.02

Allerdings entspricht dieses Vorgehen nicht der COSMO-RS-Philosophie größtmöglicher Allgemeingültigkeit, da dies eine Korrektur einer speziellen Stoffgruppe (der Alkane) und nicht wie vorangegangen einer chemischen Funktionalität ist. Aus diesem Grund wurden die Core-Ausgleichsladungen in einem zweiten Schritt auf *alle* nicht an Heteroatome gebundene sp³-hybridisierte Kohlenstoffe (und zugehörige Wasserstoffe) positioniert. Die Qualität der Statistik verschlechterte sich dadurch nur unwesentlich, und deshalb wurde die Optimierung der Vorangegangenen vorgezogen.

Tabelle 14 zeigt die Standardfehler (summiert über alle drei Zielgrößen) der Alkane, der reinen Alkohole und der Nitroverbindungen.

Tabelle 14 Standardfehler der Stoffgruppen für die Ausgleichsladungen eingeführt wurde (Trainingsdatensatz).

	ohne Ausgleichsladungen	mit Ausgleichsladungen
Alkane	0.62	0.54
OH	0.50	0.40
Nitroverb	0.95	0.57

Tabelle 14 veranschaulicht die Verbesserungen der Stoffgruppen für die Ausgleichsladungen eingeführt wurden. Die Gesamtstatistik wird allerdings nicht nur wegen der besseren Beschreibung dieser Gruppen verbessert, vielmehr ergeben sich mit den Ausgleichsladungen auch veränderte COSMO-RS-Parameter, und so werden auch andere Stoffgruppen teilweise besser beschrieben. Vor allem die gute Beschreibung der Nitrogruppen erbrachte prozentual die größte Verbesserung.

Eine abschließende Auflistung aller COSMO-RS/AM1-Parameter zeigt die nächste Tabelle. Mit angegeben sind abgeschätzte COSMO-Radien und Dispersionsparameter für Schwefel und Phosphor die in Kapitel 4.4 (Abschätzung des Blut/Hirnschranken-Übergangskoeffizienten $\log P_{BB}$ Seite 4-21 ff.) verwendet werden.

Tabelle 15: Übersicht der abschließend gewonnenen optimierten COSMO-RS/AM1-Parameter inklusive Core-Ausgleichsladungen für Nitro-, Nitrat-, Hydroxy- und sp^3 -hybridisierte Kohlenstoffgruppen.

Parameter	Wert	Parameter	Wert
r_H	1.25 Å	γ_H	-0.033 kcal/(mol Å ²)
r_C	2.00 Å	γ_C	-0.0464 kcal/(mol Å ²)
r_O	1.90 Å	γ_O	-0.039 kcal/(mol Å ²)
r_N	1.98 Å	γ_N	-0.095 kcal/(mol Å ²)
r_{Cl}	2.15 Å	γ_{Cl}	-0.052 kcal/(mol Å ²)
r_F	1.70 Å	γ_F	-0.021 kcal/(mol Å ²)
r_{av}	0.32 Å	$q_N(NO)$	-0.16
c_{mf}	1.9	$q_O(NO)$	0.08
f_{corr}	2.4	$q_H(OH)$	-0.05
c_{hb}	47000 kcal/(mol Å ²)/e ²	$q_O(OH)$	0.05
σ_{hb}	0.0030 e/Å ²	$q_H(Csp^3)$	$N^* q_H(Csp^3) $
a_{eff}	4.34 Å ²	$q_C(Csp^3)$	-0.02
λ	0.147	r_S, r_P (geschätzt)	2.20 Å, 2.25 Å
ω	-0.1801 kcal/mol	γ_S (geschätzt)	0.51 kcal/(mol Å ²)
η	-8.27	γ_P (geschätzt)	0.45 kcal/(mol Å ²)

Im nächsten Kapitel wird die vorgestellte Parametrisierung mit alternativen Rechenmethoden verglichen.

3.5 Methodenvergleiche

Die COSMO-RS/AM1-Vorhersagen sollen nicht singulär im Raum stehen, sondern werden in diesem Kapitel mit weiteren Verfahren verglichen, die die Stoffeigenschaften ebenfalls aus der chemischen Struktur vorhersagen können. Die im Folgenden verwendeten Abkürzungen AM1_1 und AM1_2 bezeichnen die eben vorgestellte Optimierung erster Ordnung (AM1_1) und die Parametrisierung mit Core-Ausgleichsladungen (= Optimierung zweiter Ordnung: AM1_2).

Neben COSMO-RS/DFT-Parametrisierungen wurden auch andere semiempirische Kontinuumsmodelle zur Berechnung der Henry-Konstante verwendet. Diese Vergleiche sind besonders interessant, sind sie doch wegen vergleichbarer Rechenintensität die direkten Konkurrenten von COSMO-RS/AM1. Schließlich wird die Vorhersagekraft von COSMO-RS/AM1 noch mit Hilfe traditioneller Gruppenbeitragsmethoden kontrolliert, die zurzeit die Standardanwendungen im Umweltbereich sind.

3.5.1 COSMO-RS/DFT

Um die Qualität der Berechnungen mit den neuen Parametersätzen einzuschätzen, wurden für den Trainings- und den Vorhersagedatensatz umfangreiche COSMO-RS-Rechnungen auf quantenchemischen DFT-Niveau (DFT = Dichtefunktionaltheorie) [55,56] durchgeführt. Hierzu wurden zwei unterschiedliche Programme verwendet, in denen das COSMO-Kontinuumsmodell bereits implementiert ist und für die bereits COSMO-RS-Parametrisierungen für verschiedene Dichtefunktionale und Basissätze existieren. Das Programm DMOL³ [57,58] wurde für Dichtefunktionalrechnungen mit dem Perdew-Wang-Korrelationsfunktional in Verbindung mit dem Becke-Austausch-Funktional (BP = B88-PW91) [59,60,61] und numerischen Basissätzen von double-zeta-Qualität inklusive Polarisationsfunktionen (dnp) verwendet. Bei den numerischen DMOL-Rechnungen wurde ein sehr feines Integrationsnetz verwendet. Es zeigten sich in unerwartet vielen Fällen Rechenabbrüche und erhebliche Konvergenzprobleme was dazu führte, dass die unvollständigen Ergebnisse nachfolgend nicht diskutiert werden.

Als Alternative wurden mit der TURBOMOLE-Software [62] nochmals zwei Dichtefunktionale mit Basissätzen unterschiedlicher Qualität angewendet. Es wurde noch einmal das BP-Funktional gewählt, jetzt allerdings mit einer triple-zeta-Valenzbasis inklusive Polarisationsfunktionen (TZVP) [63]. Um die hohen Rechenzeitanforderungen (wegen der anspruchsvollen Basen) abzufangen, wurden die Rechnungen mit der RI-Option (RI = Resolution of Identity) zur effizienten Berechnung der zwei-Elektronen-Coulomb-Potenziale unter Zuhilfenahme von Hilfsbasen (auxiliary basis sets) [64,65] durchgeführt. Schließlich wurde noch das B3LYP-Funktional [66] mit einer *split-valence* Basis plus Polarisationsfunktionen (SVP) [67] für Gasphasen- und COSMO-Optimierungen verwendet.

Die DFT-Rechnungen beanspruchten die meiste Rechenzeit, diese war weitaus höher, als die Rechenzeit der gesamten AM1-Optimierungsprozedur. Insgesamt wurden über 1900 Strukturen optimiert. Nach grober Schätzung beläuft sich die Rechenzeit auf insgesamt ca.

2000 CPU-Tage (SGI-R10000-Prozessoren). Wegen der hohen Rechenzeitanforderungen wurde auf Frequenzanalysen verzichtet.

In den folgenden Statistiken wurden einige Moleküle im Trainings- und Testdatensatz nicht berücksichtigt. Die Gründe hierfür sind meistens nicht erfolgte Konvergenz und einige wenige Fälle von COSMO-Rechenabbrüchen aufgrund von numerischen Fehlern bei der Käfigkonstruktion.

3.5.2 Weitere Kontinuumsmodelle

Neben den erwähnten COSMO-RS/DFT-Rechnungen wurden noch weitere quantenchemische Kontinuumsmodelle benutzt, um die freien Hydratationsenthalpien zu berechnen. Eine neue Parametrisierung des MST-Modells (MST = Miertus Scrocco Tomasi) zur Berechnung der freien Lösungsenthalpie in Oktanol wurde erst kürzlich veröffentlicht [68] und konnte deshalb nicht mehr zum Vergleich beim Oktanol/Wasser-Verteilungskoeffizienten herangezogen werden.

Das neben COSMO-RS/DFT verwendete Modell mit den anspruchsvollsten quantenchemischen Berechnungen ist das PCM-Modell (PCM = Polarizable Continuum Model) nach Barone, Cossi und Tomasi [69-71] mit dem UAHF-Verfahren (UAHF = United Atom for Hartree-Fock) [72] zur Konstruktion der SAS in der GAUSSIAN98-Implementierung [73]. Die Anzahl der *tesseræ* (= Anzahl der Flächenstücke des Polyeders mit dem die sphärischen Atomhüllen angenähert werden) wurde vom Standardwert (60) auf 196 erhöht, um sie der vergleichbaren Anzahl der NSPA (Number of Segments per Area) bei den COSMO/AM1-Rechnungen anzunähern.

Bei der UAHF-Konstruktion werden um Wasserstoffatome keine separaten Polyeder gebildet, sondern die Wasserstoffatome sind mit in den Sphären der Schweratome enthalten. Dies hat zur Folge, dass der im Trainingsdatensatz enthaltene molekulare Wasserstoff nicht nach diesem Verfahren berechnet werden kann.

Ausgehend von den B3LYP/TURBOMOLE-Gasphasenstrukturen wurden die Grundzustandsenergien (und Geometrien) mit der 6-31G*-Basis auf HF/SCF-Niveau (HF = Hartree Fock; SCF = Self Consistent Field) neu ermittelt. Mit den HF/SCF-Strukturen wurden dann MP2-, PCM/HF- und PCM/MP2-*single-point*-Rechnungen durchgeführt. Man erhält also Hydratationsenthalpien auf MP2- und HF-Niveau basierend auf den HF-Gasphasengeometrien.

Neben der beschriebenen ab initio-PCM-Methode wurden noch vier semiempirische SMx-Kontinuumsmodelle nach Cramer und Truhlar zum Vergleich herangezogen. Bei diesen

werden zur Berechnung der Solvatationsenthalpien zwei Beiträge addiert. Zum einen wird der elektrostatische Anteil nach der *Generalized Born Approximation* (GBA) berechnet, zum anderen wird ein zweiter Beitrag mit Hilfe der SAS berechnet, der nicht-elektrostatische Anteile der Solvatationsenthalpie quantifizieren soll. Dieser Beitrag wird bei den SMx-Modellen als ST-Term (ST = Surface Tension) oder CDS-Term (CDS = Cavity-Dispersion-Solvent-structure) bezeichnet. Letztere Bezeichnung nennt die Energiebeiträge, die der Term beinhalten soll: Die Energie zur Bildung eines Hohlraums im Lösungsmittel, Dispersionsenergiebeiträge, Strukturänderungen des Lösungsmittels, und darüberhinaus auch die nicht-elektrostatischen Energieanteile der Wasserstoffbrückenbindung.

Die verwendeten SMx-Modelle unterscheiden sich einerseits in der Wahl des semiempirischen Hamiltonoperators und andererseits auch in der Parametrisierung und Art des Kontinuummodells selbst. Auch die aus der Dichtematrix des Moleküls im Kontinuum gewonnenen Ladungen werden unterschiedlich berechnet.

Das SM2-Modell [74] ist für den AM1-Hamiltonian parametrisiert, wohingegen das SM3-Verfahren [75] auf dem PM3-Operator basiert. Beide Verfahren wurden mit AMSOL 4.0 [76] durchgeführt. Neben diesen Verfahren, die ausschließlich für wässrige Lösungen parametrisiert sind, wurden noch das SM5.4/AM1- und das SM5.4/PM3-Modell [77] zur Berechnung der Henry-Konstanten herangezogen. Diese Methoden bedienen sich der *class-IV-charges* [78,79] zur Berechnung der elektrostatischen Energiebeiträge. Schließlich wurden Henry-Konstanten mit dem SM5.0R-Modell [80] berechnet. Hier wird die freie Hydratationsenthalpie ausschließlich aus der Molekülgröße und der Molekülgestalt ermittelt. Es wird völlig auf eine elektrostatische Behandlung verzichtet, vielmehr ist der elektrostatische Anteil hier implizit in den Oberflächenspannungsparametern enthalten. Die SM5.*-Rechnungen wurden mit AMSOL 6.5.3 [81] durchgeführt. Mit Ausnahme des letzten Verfahrens wurden bei allen Methoden die Geometrien im Kontinuum optimiert. Für die SM5.0R-Rechnungen wurden dagegen MOPAC/AM1-Gasphasengeometrien verwendet.

3.5.3 Strukturinterpolierende Methoden

Schließlich werden die mit COSMO-RS berechneten Verteilungsgrößen mit den Resultaten einiger nicht Quantenchemie-basierten Gruppenbeitragsmethoden statistisch verglichen. Bei diesen werden ebenfalls physikochemische Eigenschaften ohne Zuhilfenahme experimenteller Daten direkt aus der chemischen Struktur berechnet.

Die verschiedenen Verteilungsgrößen werden entweder direkt oder mit Hilfe physikalischer Gleichungen bestimmt. So wird bei der Bindungsbeitragsmethode nach Meylan & Howard die Henry-Konstante direkt aus Bindungsbeiträgen additiv berechnet, wohingegen bei den UNIFAC-Methoden zuerst Aktivitätskoeffizienten bei unendlicher Verdünnung mit der UNIFAC-Matrix nach Tiegs et al. [82] bestimmt werden, aus denen der K_{OW} nach den verschiedenen angegebenen Methoden (siehe Referenzen in Tabelle 16) berechnet wird. Allen Methoden ist jedoch gemein, dass die Stoffeigenschaften direkt (oder alle zur Berechnung benötigten Größen) aus Strukturfragmenten extrapoliert werden. Alle strukturinterpolierenden Methoden und die nachfolgend beschriebenen statistischen Auswertungen wurden mit dem Programmsystem ChemProp [93] durchgeführt.

Da mit den Fragmentmethoden meistens nur jeweils eine Stoffeigenschaft berechnet werden kann und die Modelle sich erheblich nach der Art der Inkrementierung und nach der Anzahl der eingeführten Parameter unterscheiden, wird an dieser Stelle auf die Originalliteratur verwiesen. Die nächste Tabelle zeigt eine Übersicht aller verwendeter Methoden mit den Literaturhinweisen für die Fragmentmethoden.

3.5.4 Vergleichende Statistik

In den nachfolgenden Unterkapiteln werden die verschiedenen Rechenverfahren verglichen. Zur leichteren Übersicht werden für die einzelnen Verfahren Abkürzungen eingeführt, die in der folgenden Tabelle erklärt werden:

Tabelle 16: Abkürzungsverzeichnis der Rechenmethoden.

Rechenmethode	Abkürzung	berechnete Größe	Referenz
COSMO-RS/AM1 Optimierung erster Ordnung	RS/AM1_1	alle	
COSMO-RS/AM1 mit Core-Ausgleichsladungen	RS/AM1_2	alle	
COSMO-RS/RI-DFT (BP) TZVP-Basis (TURBOMOLE)	RS/BP	alle	
COSMO-RS/DFT (B3LYP) SVP-Basis (TURBOMOLE)	RS/B3LYP	alle	
PCM/HF 631G*-Basis (GAUSSIAN98)	PCM/HF	$\log H'$	
PCM/MP2 631G*-Basis (GAUSSIAN98)	PCM/MP2	$\log H'$	
SM2 (AM1) (AMSOL 4.0)	SM2	$\log H'$	
SM3 (PM3) (AMSOL 4.0)	SM3	$\log H'$	
SM5.4A (AM1, CM1, Solvent=Wasser AMSOL 6.5.3)	SM5.4A	$\log H'$	
SM5.4P (PM3, CM1, Solvent=Wasser AMSOL 6.5.3)	SM5.4P	$\log H'$	
SM5.0R (nicht elektrostatisch AMSOL 6.5.3)	SM5.R	$\log H'$	
Bindungsbeitragsmethode nach Meylan & Howard	MH_b	$\log H'$	[91]

Gruppenbeitragsmethode nach Meylan & Howard	MH_g	$\log H'$	[7]
Dampfdruckvorhersage nach Grain	Gr	$\log P_V$	[83]
Dampfdruckvorhersage nach Lee & Kesler	LK	$\log P_V$	[103]
Dampfdruckvorhersage nach Mackay	Ma	$\log P_V$	[84]
CLOGP	CLOGP	$\log K_{OW}$	[11]
KOW-UNIFAC	UNIFAC_KOW	$\log K_{OW}$	[89]
UNIFAC nach Banerjee and Howard	UNIFAC_BH	$\log K_{OW}$	[85]
UNIFAC nach Arbuckle	UNIFAC_A	$\log K_{OW}$	[86]
UNIFAC nach Mackay	UNIFAC_M	$\log K_{OW}$	[87]

3.5.4.1 Henry-Konstanten

Ein statistischer Vergleich der mit den verschiedenen Modellen berechneten Henry-Konstanten ist in Tabelle 17 auf Seite 3-46 gezeigt. Mit angegeben ist der quadratische Regressionskoeffizient:

$$r^2 = 1 - \frac{\sum_i (y_i - y_i^{fit})^2}{\sum_i (y_i - y_0)^2} \quad \text{Gleichung 112}$$

und der vorhersagende quadratische Regressionskoeffizient:

$$p^2 = 1 - \frac{\sum_i (y_i - y_i^{calc})^2}{\sum_i (y_i - y_0)^2} \quad \text{Gleichung 113}$$

Bei beiden Größen steht im Nenner die Summe der quadratischen Differenzen der experimentellen Werte y_i und dem Mittelwert y_0 . Im Unterschied zum p^2 -Wert steht beim r^2 -Wert im Zähler jedoch nicht die Summe der Fehlerquadrate sondern die Summe der Fehlerquadrate mit aus den ursprünglich berechneten Werten y^{calc} durch finale lineare Regression gefitteten Werten y^{fit} .

Die neue COSMO-RS/AM1-Parametrisierung zeigt statistische Überlegenheit. Von ebenfalls hoher Qualität sind die COSMO-RS/DFT-Rechnungen mit dem BP-Funktional. Im Vergleich hierzu hat COSMO-RS/B3LYP nur geringe Vorhersagekraft. Das B3LYP-Funktional in Verbindung mit dem kleinen SVP-Basissatz offenbart erhebliche Mängel. Eine genaue Analyse zeigt, dass die großen Fehler vor allem durch die schlechtere Vorhersage der Nitro- und Nitratverbindungen zu erklären ist. Auch die anspruchsvollen PCM-Verfahren werden den hohen Rechenzeitanforderungen nicht gerecht. Das PCM-Modell liefert auf störungstheo-

retischem MP2-Niveau statistisch sogar schlechtere Ergebnisse als auf HF-Niveau. Ein möglicher Grund hierfür kann die Verwendung von HF-Geometrien sein. Erfreulich ist, dass COSMO-RS/AM1 auch den anderen semiempirischen Verfahren überlegen ist, da diese ähnliche Rechenzeitanforderungen stellen.

Tabelle 17: Methodenvergleich (Trainingsdatensatz) für die logarithmische **Henry-Konstante** geordnet nach steigendem Standardfehler.

Rechenmethode ⁱ	N ^a	r ² ^b	p ² ^c	MAF ^d	Bias ^e	SE ^f	Err _{Min} ^g	Err _{Max} ^h
RS/AM1_2	720	0.91	0.91	0.47	-0.0	0.65	-3.80	2.74
RS/AM1_1	720	0.88	0.88	0.55	-0.02	0.77	-4.28	3.08
RS/BP*	717	0.88	0.87	0.56	0.23	0.79	-2.60	5.62
MH_b	675	0.89	0.85	0.47	-0.07	0.87	-7.16	3.71
SM2*	718	0.82	0.77	0.78	0.41	1.06	-3.90	6.57
SM5.4P	720	0.73	0.65	0.82	0.54	1.31	-4.34	10.46
SM2	720	0.75	0.65	0.82	0.36	1.31	-15.56	6.57
SM3*	718	0.73	0.64	1.01	-0.09	1.33	-5.11	4.71
SM3	720	0.72	0.60	1.04	-0.11	1.40	-10.08	4.71
SM5.0R**	717	0.81	0.59	0.86	-0.16	1.26	-6.67	3.62
SM5.0R	720	0.79	0.58	0.89	-0.20	1.43	-13.59	3.62
SM5.4A	720	0.65	0.53	0.94	0.65	1.52	-8.00	9.65
RS/B3LYP	711	0.40	0.07	1.31	1.01	2.11	-2.86	16.21
PCM/HF*	718	0.43	-0.11	1.40	0.83	2.32	-5.39	10.72
PCM/HF	720	0.42	-0.20	1.43	0.79	2.43	-14.03	10.72
PCM/MP2	720	0.35	-0.52	1.77	1.49	2.75	-9.96	11.19

* symbolisiert die Auslassung von Ausreißern; (* #461 und #462), (** #747 und #695)

^a N = Anzahl der Moleküle

^b quadratischer Regressionskoeffizient (siehe Gleichung 112 auf Seite 3-45)

^c vorhersagender quadratischer Regressionskoeffizient (siehe Gleichung 113 auf Seite 3-45)

^d Mittlerer absoluter Fehler

^e systematische Abweichung

^f Standardfehler

^g größter negativer Fehler

^h größter positiver Fehler

ⁱ Abkürzungen der Rechenmodelle siehe Tabelle 16 Seite 3-44

Von nahezu ähnlicher Qualität ist auch die sehr recheneffiziente Bindungsbeitragsmethode nach Meylan und Howard. Die Methode zeigt nur eine minimale systematische Abweichung (Bias), und der Standardfehler (SE) ist nur um 0.22 Einheiten größer als bei der COSMO-RS/AM1_2-Parametrisierung.

Nachfolgende Abbildung zeigt eine Auftragung experimenteller gegen berechnete Henry-Konstanten.

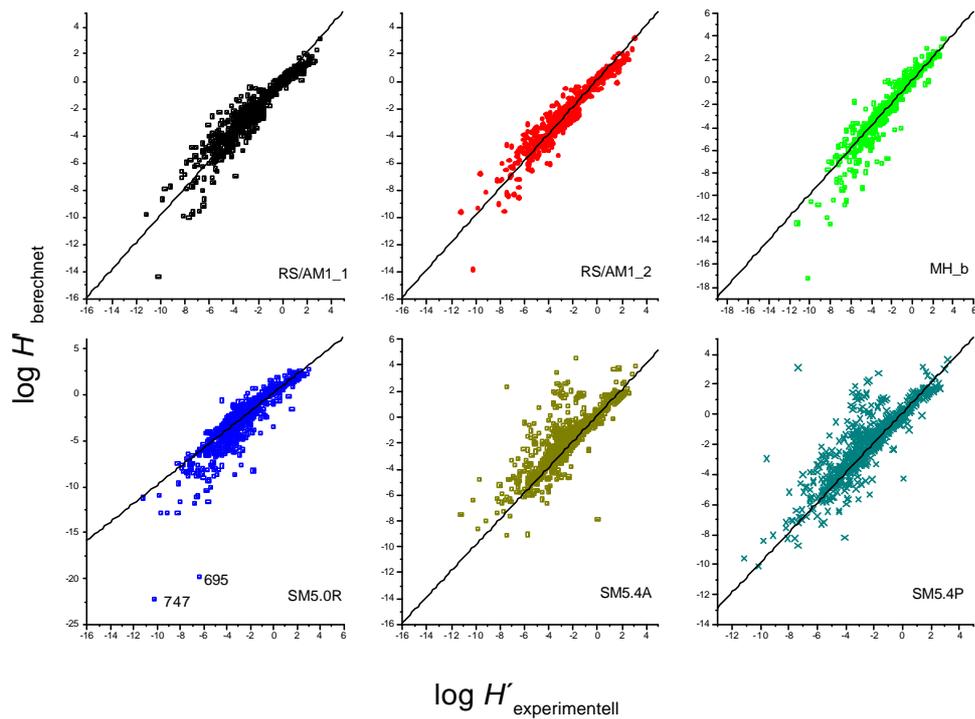


Abbildung 27: Sechs Methoden im Vergleich: Auftragung experimenteller gegen berechnete Henry-Konstanten.

Minimale und maximale experimentelle Henry-Konstanten sind gleich -11.16 , und 3.14 . Die Henry-Konstanten der Verbindungen 461, 462, 695 und 747 sind bei einigen Methoden Ausreißer, und wurden deshalb bei den betreffenden Verfahren aus der Statistik entfernt. In Abbildung 28 sind noch einmal die gleichen Methoden zusammengefasst. Allerdings sind jetzt die absoluten Fehler gegen die experimentellen $\log H'$ -Konstanten aufgetragen.

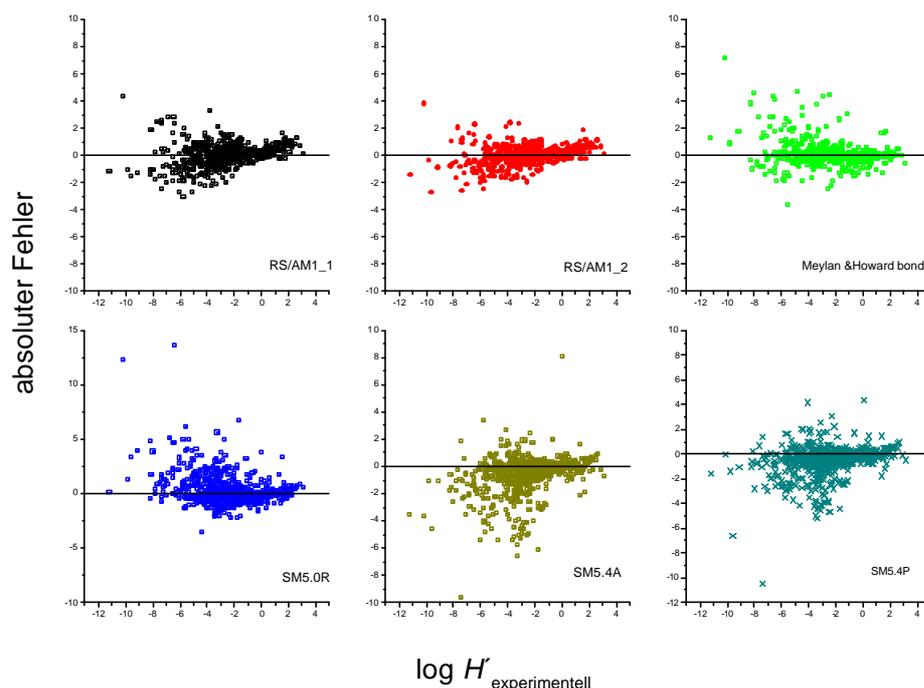


Abbildung 28: Absolute Fehler aufgetragen gegen experimentelle Henry-Konstanten.

Es offenbart sich, dass bei einigen der gezeigten Methoden, vor allem bei Substanzen mit kleinen und mittleren Henry-Konstanten, gehäuft große Fehler auftreten. Dies lässt vermuten, dass auch das experimentelle Datenmaterial fehlerbehaftet ist, da in der Tat die experimentelle Bestimmung kleiner Henry-Konstanten schwierig ist. Die nächste Abbildung (Seite 3-49) zeigt einen nach Stoffgruppen aufgeschlüsselten Vergleich der mittleren absoluten Fehler des COSMO-RS/AM1_2- und des SM3-Verfahrens. Es zeigt sich, dass COSMO-RS/AM1 nicht nur bei der Untergruppen mit Core-Ausgleichsladungen kleinere Fehler aufweist.

Dass die anderen semiempirischen Verfahren ebenfalls bei den NO-Verbindungen große Fehler zeigen, kann als Beweis angesehen werden, dass die schlechte Beschreibung der Nitro-Verbindungen nicht durch modelltheoretische, COSMO-RS-spezifische Defizite verursacht wird, sondern aus Defiziten der semiempirischen Methoden selbst resultiert. Auch bei den anderen Gruppen, mit Ausnahme der stickstoffhaltigen Halogenverbindungen, ist die Vorhersagekraft von COSMO-RS/AM1 größer als die von SM2.

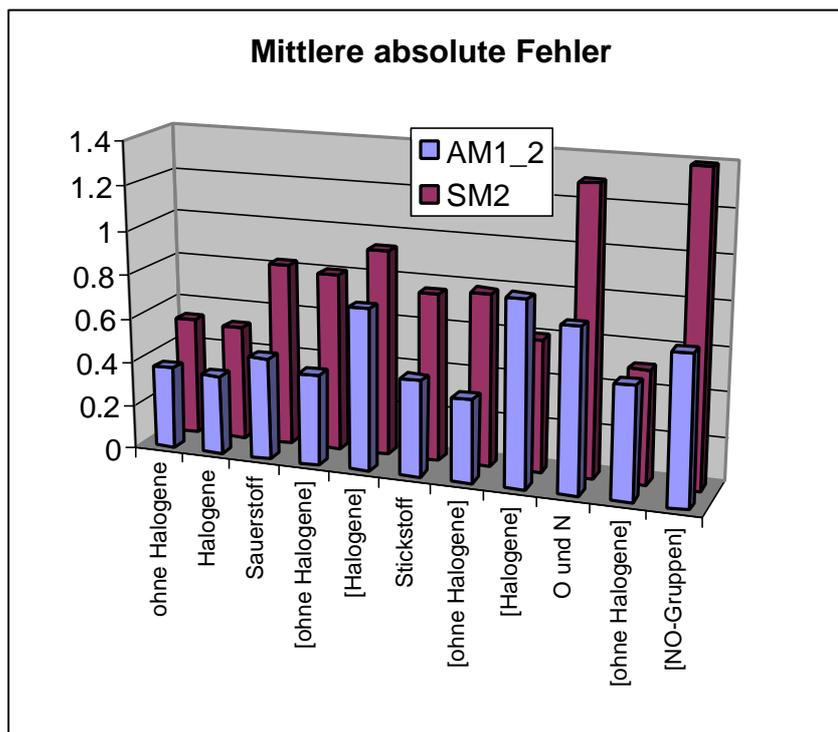


Abbildung 29: Fehler des Trainingsdatensatzes, geordnet nach Stoffgruppen. Gruppen in Klammern sind Untergruppen der vorangegangenen Gruppe (von links nach rechts).

Zu einer unabhängigen Beurteilung der statistischen Qualität muss aber der Vorhersagedatensatz (Testdatensatz) herangezogen werden. Die folgende Tabelle zeigt die vergleichende Statistik für den Testdatensatz.

Tabelle 18: Methodenvergleich (Testdatensatz) für die logarithmische **Henry-Konstante** geordnet nach steigendem Standardfehler.

Rechenmethode	N	r^2	p^2	MAF	Bias	SE	ExpMin	ExpMax	ErrMin	ErrMax
RS/AM1_2*	201	0.88	0.87	0.53	0.12	0.76	-9.44	2.70	-3.51	3.43
MH_b	194	0.90	0.87	0.42	-0.09	0.78	-9.44	2.70	-3.86	2.63
MH_g	171	0.90	0.83	0.42	-0.24	0.82	-7.88	2.70	-5.25	1.03
RS/AM1_2	202	0.81	0.79	0.57	0.07	0.98	-9.44	2.70	-8.71	3.43
SM2	202	0.82	0.75	0.82	0.49	1.06	-9.44	2.70	-3.32	3.71
SM5.4A	202	0.77	0.72	0.75	0.33	1.14	-9.44	2.70	-3.36	6.01
SM3	202	0.71	0.61	1.00	0.10	1.34	-9.44	2.70	-4.87	5.83

*ohne den Ausreißer Methaldehyd

Abkürzungen siehe Tabelle 16 Seite 3-44 und Tabelle 17 Seite 3-46

Wird Methaldehyd ($\log H'$ -Fehler = -8.71) vernachlässigt, kann COSMO-RS die Spitzenposition behaupten. Die beiden Gruppenbeitragsmethoden von Meylan und Howard zeigen allerdings einen größeren quadratischen Regressionskoeffizienten. Nachteil der Bindungsbeitragsmethode ist die fehlende Parametrisierung für acht Verbindungen im Datensatz.

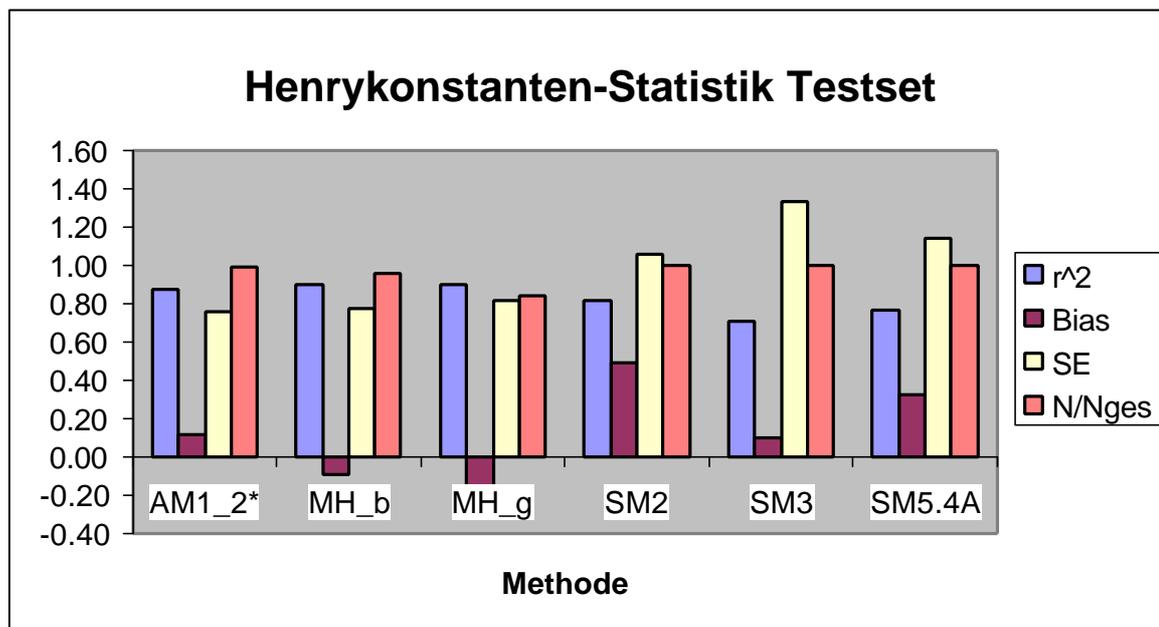


Abbildung 30: Methodenvergleich für die Henry-Konstanten des Test-Sets (*ohne V83).

Die Abbildung zeigt noch einmal die einzelnen statistischen Variablen im Überblick. Angegeben ist auch der prozentuale Anteil der berechenbaren Größen N/N_{ges} . Alle quantenchemischen semiempirischen Verfahren haben eine positive systematische Abweichung (Bias), wohingegen die beiden Gruppenbeitragsmethoden negative systematische Abweichungen aufweisen.

3.5.4.2 Dampfdruck

Ebenso wie im Falle der Henry-Konstanten zeigt COSMO-RS bei der Vorhersage des Dampfdrucks die statistisch besten Ergebnisse. Die systematische Abweichung (Bias), der mittlere absolute Fehler und der Standardfehler sind bei der COSMO-RS/AM1-Parametrisierung für den Trainingsdatensatz am geringsten. Auch die COSMO-RS/BP-Methode zeigt sich den Gruppenbeitragsmethoden nach Mackay und Grain überlegen.

Tabelle 19: Methodenvergleich (Trainingsdatensatz) für den logarithmischen **Dampfdruck** geordnet nach steigendem Standardfehler.

Rechenmethode	N	r ²	p ²	MAF	Bias	SE	ExpMin	ExpMax	ErrMin	ErrMax
RS/AM1_2	437	0.92	0.92	0.34	0.00	0.52	-9.72	1.70	-2.93	2.31
RS/AM1_1	437	0.92	0.92	0.35	0.03	0.53	-9.72	1.70	-2.51	2.68
RS/BP**	434	0.92	0.90	0.38	-0.02	0.57	-8.60	1.70	-4.03	1.85
RS/BP	436	0.80	0.77	0.42	0.02	0.87	-9.72	1.70	-4.03	10.15
Ma*	432	0.77	0.74	0.63	0.17	0.92	-9.72	1.70	-3.72	2.48

Ma	433	0.77	0.75	0.63	0.18	0.93	-9.90	1.70	-3.72	2.48
Gr	435	0.80	0.72	0.49	-0.18	0.99	-9.90	1.70	-5.14	1.32
Gr*	434	0.79	0.71	0.49	-0.18	0.99	-9.72	1.70	-5.14	1.32
RS/B3LYP	431	0.50	0.18	1.03	0.76	1.57	-9.72	1.70	-3.48	9.88
LeKe*	435	0.76	0.22	0.74	-0.49	1.61	-9.72	1.70	-9.41	2.79

* ohne T324; ** ohne T213 und T352

Abkürzungen siehe Tabelle 16 Seite 3-44 und Tabelle 17 Seite 3-46

Im Vergleich zu den Henry-Konstanten-Statistiken sind die Dampfdruckvorhersagen genauer. Der Standardfehler ist um 0.13 logarithmische Einheiten geringer. Die maximalen und minimalen Fehler sind ebenfalls geringer.

Ein anderes Bild zeigt der Methodenvergleich für den Testdatensatz (Abbildung 31). Hier scheint das Gruppenbeitragsmodell nach Grain der neuen AM1-Parametrisierung noch überlegen. Die Erklärung hierfür ist, dass die vom Grain-Modell eher ungenügend beschriebenen Nitro- und Nitratverbindungen ($SE = 2.52$) im Testdatensatz stark unterrepräsentiert sind (vgl. 3.1.2 Seite 3-7).

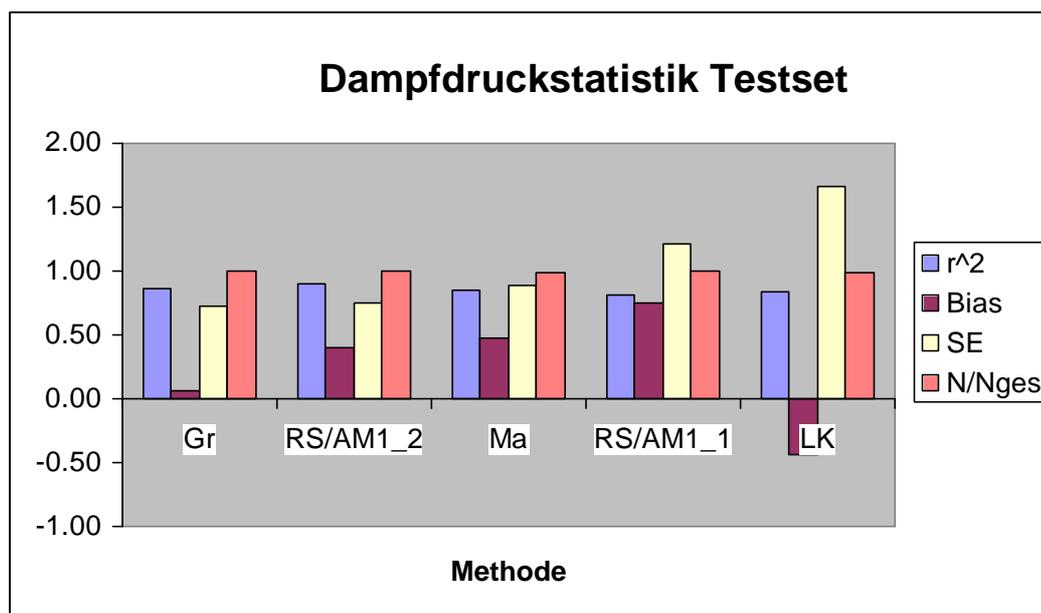


Abbildung 31: Dampfdruck-Methodenvergleich für den Vorhersagedatensatz.

Den Inkrementmethoden nach Mackay und nach Lee/Kesler ist die COSMO-RS/AM1-Parametrisierung trotz der wenigen NO-Verbindungen auch beim Vorhersagedatensatz statistisch überlegen.

3.5.4.3 Oktanol/Wasser-Verteilungskoeffizient

Die gute Vorhersagequalität des CLOGP-Programms kann mit COSMO-RS/AM1 nicht erreicht werden. Obwohl sich ein guter Regressionskoeffizient ergibt, ist der Standardfehler von COSMO-RS/AM1 doppelt so groß wie der von CLOGP. Eine bevorzugte Anwendung des COSMO-RS/AM1-Modells zur Berechnung des Oktanol/Wasser-Verteilungskoeffizienten bei 25°C empfiehlt sich hier also nicht. Immerhin können die UNIFAC-Methoden übertroffen werden. Der einzige Vorteil der COSMO-RS-Methode ist wiederum die umfassende Anwendbarkeit. So können beispielsweise die $\log K_{OW}$ -Werte von 27 Verbindungen (dies entspricht einem Anteil von 5.4%) wegen fehlender Parametrisierung mit CLOGP nicht berechnet werden. Die Anwendung der UNIFAC-Methoden ist nur für ca. 32% der Verbindungen möglich.

Tabelle 20: Methodenvergleich (Trainingsdatensatz) für den logarithmischen **Oktanol/Wasser-Verteilungskoeffizient** geordnet nach steigendem Standardfehler.

Rechenmodell	N	r ²	p ²	MAF	Bias	SE	ExpMin	ExpMax	ErrMin	ErrMax
CLOGP	471	0.99	0.99	0.15	0.00	0.26	-2.04	8.68	-1.15	1.24
UNIFAC_KOW	134	0.98	0.97	0.23	-0.08	0.39	-1.36	7.44	-2.88	0.66
RS/AM1_2	498	0.93	0.93	0.39	0.00	0.53	-2.04	8.68	-1.90	2.91
RS/AM1_1	498	0.93	0.93	0.40	0.02	0.55	-2.04	8.68	-1.78	2.93
UNIFAC_BH	157	0.96	0.95	0.42	-0.17	0.55	-1.36	8.68	-2.01	1.23
RS/BP	497	0.92	0.91	0.43	0.13	0.62	-2.04	8.68	-1.39	3.55
RS/B3LYP	492	0.87	0.84	0.48	0.25	0.81	-2.04	8.68	-1.25	5.34
ALOGP	463	0.89	0.85	0.55	-0.40	0.82	-2.04	8.68	-5.96	1.61
UNIFAC_A	157	0.95	0.89	0.70	0.51	0.86	-1.36	8.68	-1.32	2.53
UNIFAC_M	157	0.96	0.81	0.83	0.55	1.11	-1.36	8.68	-1.30	3.35

Abkürzungen siehe Tabelle 16 Seite 3-44 und Tabelle 17 Seite 3-46

Um eine bessere Vorhersagequalität für den Oktanol/Wasser-Verteilungskoeffizienten zu erhalten, muss COSMO-RS/AM1 also speziell für diese Größe angepasst werden. Beispielsweise kann der Standardfehler deutlich reduziert werden, wenn in der Anpassung ausschließlich der $\log K_{OW}$ als Zielgröße verwendet wird und ein Oberflächendispersionsterm:

$$\Delta m_{disp} = \sum_k g_k A_k^X \quad \text{Gleichung 114}$$

als Korrektur zur Differenz der chemischen Potentiale addiert wird. Der Oktanol/Wasser-Verteilungskoeffizient berechnet sich dann nach:

$$\log K_{ow} = \frac{\Delta G_{o \rightarrow w}}{2.3RT} = \frac{m_w^X - m_o^X + \Delta m_{disp}^X}{2.3RT} \quad \text{Gleichung 115}$$

Nach diesem Rechenverfahren verringert sich der Standardfehler um 0.11 logarithmische Einheiten auf $SE = 0.42$. Nachfolgende Abbildung zeigt eine Auftragung experimenteller gegen modellierte logarithmische K_{OW} -Werte. Verglichen werden COSMO-RS/AM1_2, CLOGP und die eben vorgestellte, ausschließlich an den Oktanol/Wasser-Verteilungskoeffizient angepasste COSMO-RS-Parametrisierung.

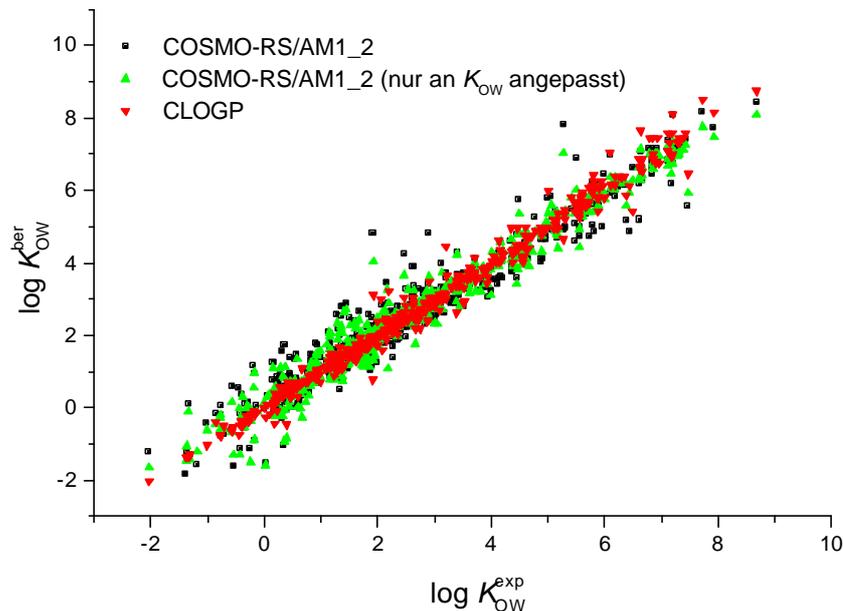


Abbildung 32: Methodenvergleich der $\log K_{OW}$ -Berechnung. Bei exklusiver Anpassung an den K_{OW} wird ein deutlich besseres Ergebnis erzielt, die Qualität von CLOGP kann dennoch nicht erreicht werden.

Trotz der exklusiven Anpassung kann die statistische Qualität von CLOGP nicht erreicht werden. Allerdings wurden auch keine großen Anstrengungen einer Spezialparametrisierung unternommen, da die Zielsetzung des Forschungsprojekts, eine umfassende Grundparametrisierung zu ermitteln, nicht verlassen werden sollte.

Es bleibt fraglich, ob die Qualität der CLOGP-Fragmentmethode überhaupt erreicht werden kann, enthält diese doch mehrere hundert Parameter und Korrekturfaktoren. Dies wird deutlich, wenn man sich beispielsweise das größte mit eigenen Parametern versehene Fragment der CLOGP-Methode (Abbildung 33) betrachtet:

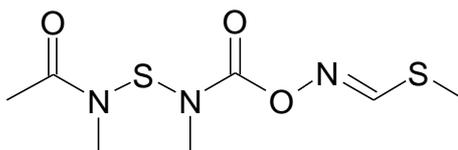


Abbildung 33: Größtes in CLOGP parametrisiertes Fragment.

Dieses „Fragment“ ist nicht mehr nur eine Funktionalität, vielmehr handelt es sich um ein spezielles Molekül, welches quasi extra parametrisiert ist.

Ein Methodenvergleich für den Testdatensatz zur Validierung der Methode ist nicht notwendig, da die experimentellen Oktanol/Wasser-Verteilungskoeffizienten der Daylight Datenbank entnommen wurden, die auch bei der Anpassung der CLOGP-Methode (angepasst an 10000 experimentelle Daten) verwendet wurde. Allerdings soll angemerkt werden, dass die Überprüfung mit dem Testdatensatz die Ergebnisse des Trainingsdatensatzes bestätigten. So weichen die statistischen Ergebnisse des Testdatensatzes nur minimal von denen des Trainingsdatensatzes ab. Der Standardfehler ist 0.01 log-Einheiten größer, und der r^2 -Wert hat um 0.02 abgenommen.

Es konnte gezeigt werden, dass die neue COSMO-RS/AM1-Parametrisierung bei der Vorhersage von Henry-Konstanten die statistisch geringsten Fehler aufweist. Dies wurde anhand eines Testdatensatzes bestätigt. Vorteil gegenüber der fast äquivalenten COSMO-RS/BP-Methode ist die geringe Rechenzeit. Die Methode ist anderen semiempirischen Verfahren überlegen. Auch die Dampfdruckvorhersagen sind von hoher statistischer Qualität. Größter Vorteil ist die umfassende Anwendbarkeit der Methode.

Im nächsten Kapitel werden einige Anwendungsbeispiele der neuen COSMO-RS/AM1-Methode vorgestellt, anhand derer einige Stärken des COSMO-RS-Verfahrens verdeutlicht werden. Neben der Berechnung von Stoffeigenschaften als Funktion der Temperatur, werden auch physikochemische Verteilungsgrößen berechnet, die nicht in der Optimierungsprozedur berücksichtigt wurden.

4 COSMO-RS-Anwendungen

4.1 Berechnung temperaturabhängiger Henry-Konstanten von 21 aromatischen Kohlenwasserstoffen mit COSMO-RS/AM1

Eine umweltproblematische Substanzklasse ist die Klasse der (polycyclischen) aromatischen Kohlenwasserstoffe, ist doch für einige Vertreter eine karzinogene Wirkung nachgewiesen, und andere Vertreter dieser Substanzklasse werden als mutagen verdächtigt.

Mit COSMO-RS ist es im Unterschied zu traditionellen Inkrement-Methoden (z.B. UNIFAC [88,89] nun möglich, freie Hydratationsenthalpien in fluiden Phasen direkt aus der Molekülstruktur ohne Zusatzinformationen temperaturabhängig zu berechnen.

Die hier präsentierten COSMO-RS/AM1-Rechnungen basieren auf einem Parametersatz, welcher während der Anpassung an CHO-Verbindungen gewonnen wurde. Da ausschließlich Kohlenwasserstoffe behandelt werden, wird eine Parametrisierung ohne skalierenden Sauerstoffparameter (vgl. Kapitel 3.3 Seite 3-15 ff.) verwendet. Tabelle 21 enthält eine vollständige Liste aller verwendeter COSMO-RS/AM1-Parameter. Zum Vergleich sind die COSMO-RS/DFT-Parameter mit angegeben. AM1- und COSMO/AM1-Rechnungen wurden mit MOPAC98 [36] durchgeführt.

Tabelle 21: Verwendeter COSMO-RS/AM1 Parametersatz.

Parameter	COSMO-RS/AM1	COSMO-RS/DFT
r_H	1.25 Å	1.30 Å
r_C	2.00 Å	2.00 Å
r_O	1.70 Å	1.72 Å
γ_H	-0.039 kcal/(mol Å ²)	-0.041 kcal/(mol Å ²)
γ_C	-0.031 kcal/(mol Å ²)	-0.037 kcal/(mol Å ²)
γ_O	0.037 kcal/(mol Å ²)	-
r_{av}	0.28 Å	0.5 Å
α'	1282 kcal/mol Å ² /e ²	1288 kcal/mol Å ² /e ²
f_{corr}	3.4	2.4
c_{hb}	12782 kcal/(mol Å ²)/e ²	7400 kcal/(mol Å ²)/e ²
σ_{hb}	0.0081 e/Å ²	0.0082 e/Å ²
a_{eff}	6.5 Å ²	7.1 Å ²
λ	0.17	0.14
ω	-0.25 kcal/mol	-0.21 kcal/mol
η	-8.31	-9.15

Die Henry-Konstanten von 21 aromatischen Kohlenwasserstoffen (Tabelle 22) wurden für Temperaturen zwischen 4°C und 31°C mit COSMO-RS/AM1 berechnet und mit experimentellen Henry-Konstanten [90] verglichen. Für die ersten 13 Verbindungen (PAKs = Polycyclische aromatische Kohlenwasserstoffe) lagen experimentelle Daten bei 4.1, 11, 13, 25 und 31°C vor. Experimentelle Henry-Konstanten der Substanzen 14 bis 21 (Benzol und Benzolderivate) wurden dagegen verschiedenen Quellen entnommen. Während bei 4°C die Henry-Konstanten der 13 PAKs um fast drei Größenordnungen ($\Delta \log H = 2.81$) variieren, verringert sich der $\log H$ -Unterschied zwischen den Substanzen **1** und **13** bei Temperaturerhöhung auf 31°C kontinuierlich auf $\Delta \log H = 1.79$ (vgl. Tabelle 22).

Tabelle 22: Logarithmische Henry-Konstanten. H [Pa L mol⁻¹]; experimentell und mit COSMO-RS/AM1 berechnet (T in °C).

#	Substanz	Temperatur									
		4.1°C		11°C		18°C		25°C		31°C	
		exp	ber	exp	ber	exp	ber	exp	ber	exp	ber
1	2-Methylnaphthalin	4.11	4.33	4.32	4.52	4.52	4.70	4.71	4.86	4.87	5.00
2	1-Methylnaphthalin	4.00	4.32	4.24	4.50	4.47	4.68	4.68	4.84	4.86	4.98
3	Acenaphthylen	3.38	3.46	3.63	3.67	3.87	3.86	4.10	4.05	4.29	4.20
4	Acenaphthen	3.55	3.87	3.80	4.07	4.04	4.26	4.27	4.44	4.46	4.59
5	Fluoren	3.31	3.52	3.55	3.74	3.78	3.95	3.99	4.14	4.17	4.30
6	Phenanthren	2.97	3.17	3.20	3.41	3.42	3.63	3.63	3.84	3.80	4.01
7	Anthracen	3.10	3.22	3.33	3.46	3.54	3.68	3.75	3.89	3.92	4.06
8	1-Methylphenanthren	3.20	3.24	3.37	3.48	3.54	3.71	3.70	3.93	3.83	4.11
9	Fluoranthren	2.75	2.57	2.94	2.83	3.12	3.08	3.29	3.32	3.43	3.51
10	Pyren	2.63	2.57	2.84	2.82	3.04	3.07	3.23	3.30	3.39	3.49
11	Benzo[a]fluoren	2.94	2.42	3.11	2.70	3.28	2.97	3.43	3.22	3.56	3.43
12	Benz[a]anthracen	2.18	2.15	2.49	2.44	2.80	2.73	3.09	2.99	3.32	3.21
13	Chrysen	1.30	2.12	1.85	2.41	2.28	2.69	2.72	2.95	3.08	3.17
	Temp:	5°C		10°C		20°C		25°C		30°C	
14	Benzol	5.44		5.52		5.66		5.73		5.79	
15	Toluol	5.51		5.60		5.76		5.83		5.90	
16	o-Xylol	5.54		5.64		5.82		5.90		5.98	
17	Ethylbenzol	5.69		5.79		5.96		6.05		6.12	
18	m-Xylol	5.55		5.65		5.83		5.92		6.00	
19	p-Xylol	5.55		5.65		5.83		5.92		6.00	
20	Propylbenzol	5.84		5.95		6.14		6.23		6.31	
21	Butylbenzol	6.01		6.12		6.33		6.43		6.52	

Dieser Trend wird von COSMO-RS/AM1 mit $\Delta \log H = 2.21$ (4°C) und $\Delta \log H = 1.83$ (31°C) qualitativ gut wiedergegeben. Ohne Chrysen mit einer bei tiefen Temperaturen ungewöhnlich kleinen Henry-Konstante ist die Übereinstimmung sogar nahezu quantitativ ($SD^{4°C}(\text{PAK}) = 0.24$). Die zwei Methyl-Naphthalinisomere (Abbildung 34 Seite 4-3) haben im betrachteten Temperaturbereich die geringste Affinität aller PAKs zu Wasser. Obwohl der

Standardfehler bei 25°C 0.16 logarithmische Einheiten beträgt, werden auch viel kleinere H -Unterschiede zwischen ähnlichen Strukturen prinzipiell richtig wiedergeben. So bestätigen die Rechnungen den Befund, dass bei 25°C der $\log H$ von 2-Methylnaphthalen etwas größer ist als der $\log H$ von 1-Methylnaphthalen.

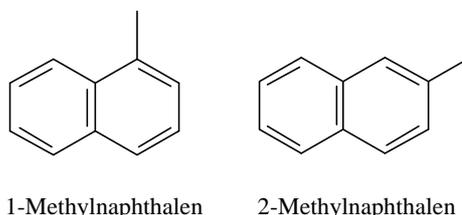


Abbildung 34: Die zwei Methylnaphthalen-Isomere.

Dies ist ein grundsätzlicher Vorteil von COSMO-RS, da traditionelle Gruppenbeitragsmethoden i.a. nicht zwischen Strukturisomeren unterscheiden können. So erhält man für beide Isomere mit dem auf Bindungsincrementen basierenden Verfahren nach Meylan und Howard [91] den gemeinsamen $\log H$ -Wert 4.74.

Interessanterweise ist die Übereinstimmung tendenziell besser für höhere Temperaturen. Die Standardfehler für $T = 4.1^\circ\text{C}$, 11°C , 18°C , 25°C und 31°C betragen in $\log H$ -Einheiten 0.32, 0.25, 0.20, 0.16 und 0.14. Zum Vergleich wurde $\log H$ als (P_v/S_w) gemäß Gleichung 64 (Seite 2-21) mit aus der Molekülstruktur berechneten Werten für den temperaturabhängigen Dampfdruck [13] und für die näherungsweise als temperaturunabhängig behandelte Wasserlöslichkeit [92] bei 25°C mit Hilfe des Programmsystems ChemProp [93] berechnet. Die entsprechenden $\log H$ -Standardfehler betragen 0.82, 0.73, 0.63, 0.54 und 0.48 und sind damit bei gleichem Trend fast dreimal so groß wie mit COSMO/RS-AM1.

Physikalische Einsichten gewährt die Aufteilung der freien Hydratationsenthalpie in einen elektrostatischen (ΔG_{hydr}^{el}) und einen nicht-elektrostatischen Anteil (ΔG_{hydr}^{nel}):

$$\Delta G_{hydr} = \Delta G_{hydr}^{el} + \Delta G_{hydr}^{nel} - C$$

Gleichung 116

Die Konstante C ist eine vom jeweiligen Konzentrationsmaß abhängige Größe. Um Werte bezogen auf bar (Druck) für die Gasphase und mol/mol (Molenbruch) für die Wasserphase, (dies sind die bei COSMO-RS verwendeten Konzentrationsmaße) in die gebräuchliche Einheit mol/L umzurechnen, erhält C den Wert $RT \ln(\text{Anzahl der Mole Wasser in einem Liter} \cdot \text{molares Gasphasen-Volumen})$ (vgl. Gleichung 67 bis Gleichung 77 Seite 2-21 ff.). Für die COSMO-RS-Referenzzustände ist also $C = 0$, so dass bei Verwendung dieser Einheiten ΔG_{hydr} als einfache Summe von ΔG_{hydr}^{el} und ΔG_{hydr}^{nel} geschrieben werden kann. Hierbei ist der

elektrostatische Anteil die Summe der idealen Abschirmenergie und der sogenannten mittleren Misfitenergie (Gleichung 117).

$$\Delta G_{hydr}^{el} = \Delta^X + \frac{1}{2} \sum_i -s_i \frac{\sum_j E_{misfit}(s_i, s_j) s_j e^{\left(\frac{-E_{misfit}(s_i, s_j) + m_s(s_i)}{kT}\right)}}{\sum_j s_j e^{\left(\frac{-E_{misfit}(s_i, s_j) + m_s(s_i)}{kT}\right)}} \quad \text{Gleichung 117}$$

Gleichung 118 zeigt Bestandteile des nicht-elektrostatischen Energieanteils. Auf der rechten Seite stehen (von links nach rechts): die Dispersionsenergieanteile, ein Term für entropische Anteile, die Kavitationsenergie (= das um den elektrostatischen Anteil reduzierte chemische Potenzial) und der Korrekturterm für Ringverbindungen, welcher systematische Fehler der Hydratationsenergien von zyklischen Verbindungen kompensiert [94].

$$\Delta G_{hydr}^{nel} = \sum_k g_k A - hRT + (m_s^X - \Delta G_{hydr}^{el}) - wn_{ra}^X \quad \text{Gleichung 118}$$

Eine Anmerkung muss an dieser Stelle zur Entropie gemacht werden. Obwohl ein Entropie-term in ΔG_{hydr}^{nel} enthalten ist, sind dennoch auch Entropieanteile in ΔG_{hydr}^{el} enthalten. Der Grund hierfür ist die Quantifizierung der Misfitenergie durch Paarwechselwirkung der SAS-Segmente. Hierbei wird die Entropie durch die Bildung aller möglichen Paare erhöht. Im Realfall werden nicht alle Paarungen wegen sterischen Restriktionen vorkommen. Dies verdeutlicht noch einmal, dass es sich bei der Aufspaltung der Hydratationsenthalpie um keine streng physikalische sondern modelltheoretische Aufteilung handelt.

Im Vergleich zu den ΔG_{hydr}^{nel} -Werten, die mit steigender Temperatur stark zunehmen, erhöht sich ΔG_{hydr}^{el} bei höheren Temperaturen nur wenig (Tabelle 23).

Tabelle 23: Elektrostatische und nichtelektrostatische Anteile der freien Hydratationsenthalpien ΔG_{hydr} aller 21 Substanzen. Freien Enthalpien sind in kJ mol^{-1} für die COSMO-RS-spezifischen Referenzzustände angegeben.

#	Temperatur (°C)									
	4.1	11	18	25	31	4.1	11	18	25	31
	ΔG_{hydr}^{el}					ΔG_{hydr}^{nel}				
1	-4.31	-4.17	-4.03	-3.89	-3.77	10.12	11.08	12.04	12.97	13.77
2	-4.43	-4.29	-4.16	-4.02	-3.90	10.18	11.12	12.07	13.00	13.79
3	-7.64	-7.51	-7.37	-7.23	-7.12	8.81	9.77	10.73	11.67	12.47
4	-4.68	-4.53	-4.39	-4.25	-4.13	8.05	9.02	9.99	10.94	11.75
5	-6.22	-6.07	-5.92	-5.77	-5.64	7.71	8.72	9.74	10.74	11.58
6	-7.41	-7.26	-7.10	-6.95	-6.81	7.07	8.11	9.17	10.19	11.06
7	-7.06	-6.90	-6.75	-6.59	-6.45	6.97	8.03	9.10	10.13	11.01

8	-6.99	-6.82	-6.66	-6.49	-6.34	6.98	8.08	9.18	10.26	11.17
9	-9.03	-8.86	-8.69	-8.52	-8.37	5.46	6.58	7.70	8.79	9.71
10	-8.88	-8.72	-8.55	-8.38	-8.24	5.30	6.40	7.50	8.57	9.49
11	-8.77	-8.59	-8.40	-8.22	-8.06	4.44	5.62	6.80	7.95	8.93
12	-9.53	-9.34	-9.15	-8.96	-8.80	3.75	4.96	6.18	7.36	8.37
13	-9.79	-9.61	-9.42	-9.23	-9.07	3.84	5.04	6.25	7.42	8.42
14	-1.80	-1.72	-1.63	-1.55	-1.46	13.43	14.14	14.90	15.61	16.19
15	-1.38	-1.26	-1.17	-1.05	-0.96	13.35	14.10	14.90	15.69	16.36
16	-1.17	-1.05	-0.96	-0.84	-0.71	13.31	14.14	15.02	15.86	16.53
17	-0.21	-0.08	0.04	0.17	0.25	13.10	13.97	14.81	15.69	16.36
18	-1.05	-0.92	-0.79	-0.67	-0.59	13.22	14.06	14.94	15.82	16.48
19	-0.96	-0.84	-0.75	-0.63	-0.50	13.18	14.06	14.90	15.77	16.48
20	0.75	0.88	1.00	1.13	1.26	12.97	13.89	14.81	15.73	16.48
21	1.84	1.97	2.13	2.30	2.43	12.76	13.77	14.73	15.73	16.53

So ist der über alle Verbindungen gemittelte Zuwachs von ΔG_{hydr}^{el} (0.56 kJ mol^{-1}) nur etwa 14% des Zuwachses von ΔG_{hydr}^{nel} (3.76 kJ mol^{-1}). Einige weitere Trends sind offensichtlich: Je größer das Molekulargewicht einer Substanz ist, desto größer ist die elektrostatische Stabilisierung. Aliphatische Substituenten verringern die elektrostatische Stabilisierung (bei ähnlichem Molekulargewicht). So hat ΔG_{hydr}^{el} von Acenaphthylen (**3**) bei 25°C den Wert $-7.23 \text{ kJ mol}^{-1}$, während der elektrostatische Energiegewinn durch Solvation ΔG_{hydr}^{el} beim zweifach hydrierten Acenaphthen (**4**) nur $-4.25 \text{ kJ mol}^{-1}$ beträgt (siehe Abbildung 35).

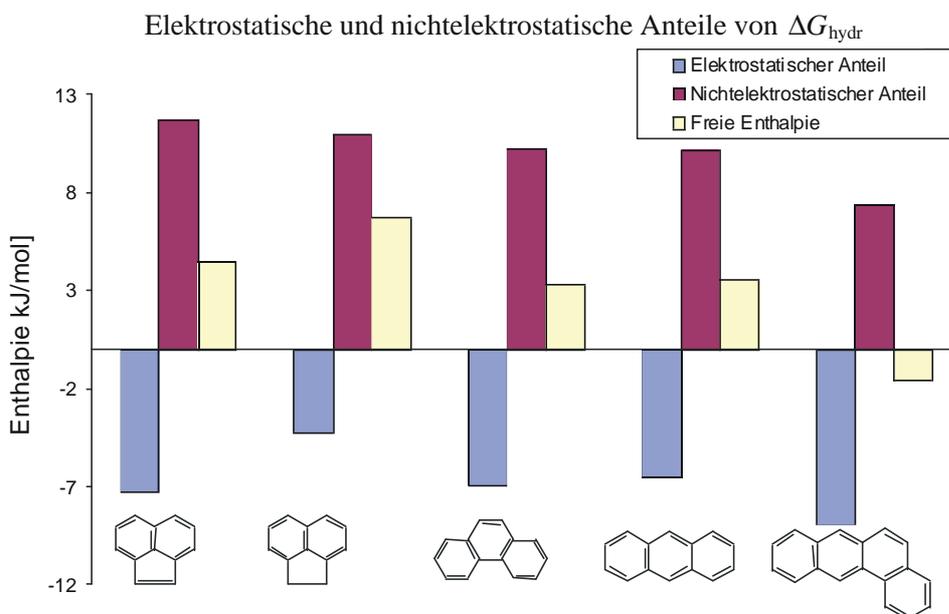


Abbildung 35: Elektrostatische und nicht-elektrostatische Anteile der freien Solvatationsenthalpie (Einheiten in kJ/mol für COSMO-RS-spezifische Referenzzustände).

Die Differenz der nicht-elektrostatischen Anteile beträgt nur 0.73 kJ mol^{-1} , und so erklärt sich der Unterschied in der freien Hydratationsenthalpie hauptsächlich durch die unterschiedliche Elektrostatik. Die nicht-elektrostatischen Beiträge verringern sich ebenfalls bei steigendem Molekulargewicht. Dies ist auf eine Zunahme der Dispersionsenergie zurückzuführen, die innerhalb der COSMO-RS-Theorie mit der Größe der Moleküloberfläche korreliert ist. Zusätzlich macht sich bei den PAKs die Stabilisierung durch die Ringkorrektur bemerkbar, denn die Anzahl der Ringatome steigt von Benzol ($n_{\text{Ringatom}} = 6$) auf $n_{\text{Ringatom}} = 18$ für die Verbindungen **12** und **13**.

Eine Auftragung von ΔG_{hydr} gegen T liefert die Hydratationsenthalpie ΔH_{hydr} und die zugehörige Entropie ΔS_{hydr} aus Achsenabschnitt und Steigung. In Tabelle 24 sind die durch Regression erhaltenen Enthalpie- und Entropiewerte zusammengefasst.

Tabelle 24: Mit linearer Regression berechnete Entropien (ΔS_{hydr}) und Enthalpien (ΔH_{hydr}). ΔH ist in kJ mol^{-1} und ΔS in $\text{kJ mol}^{-1} \text{ K}^{-1}$ angegeben. Alle Angaben beziehen sich auf die COSMO-RS-Referenzzustände.

Substanz	ΔH^{exp}	$\Delta H^{\text{COSMO-RS/AMI}}$	ΔS^{exp}	$\Delta S^{\text{COSMO-RS/AMI}}$
1	-45	-37	-0.18	-0.16
2	-51	-37	-0.20	-0.15
3	-55	-42	-0.20	-0.16
4	-55	-40	-0.20	-0.16
5	-51	-44	-0.19	-0.17
6	-50	-48	-0.17	-0.17
7	-49	-48	-0.18	-0.17
8	-38	-50	-0.14	-0.18
9	-41	-54	-0.14	-0.18
10	-45	-53	-0.15	-0.18
11	-37	-58	-0.13	-0.19
12	-69	-61	-0.23	-0.20
13	-106	-61	-0.34	-0.20
14	-26	-19	-0.07	-0.05
15	-25	-21	-0.04	-0.06
16	-34	-24	-0.09	-0.07
17	-43	-23	-0.13	-0.07
18	-34	-24	-0.10	-0.07
19	-37	-24	-0.10	-0.07
20	-42	-26	-0.14	-0.08
21	-47	-28	-0.16	-0.09

Alle Solvatationsenthalpien (außer **13** und **11**) haben absolute Fehler unterhalb 20 kJ mol^{-1} , und die Regressionskoeffizienten sind über den betrachteten Temperaturbereich durchweg größer als 0.99. Der Fehler scheint groß, aber in der Literatur werden Enthalpien gefunden, die teilweise um mehr als 30 kJ/mol voneinander abweichen (Beispiel: Phenanthren 60.21

kJ/mol [95, 96] und 29.30 kJ/mol [97]; der COSMO-RS-Wert liegt mit 47.7 kJ/mol dazwischen). Die COSMO-RS-Ergebnisse können so bei der Validierung experimenteller Daten wertvolle Hilfe leisten. Dies lässt sich am Beispiel homologer Reihen verdeutlichen: ΔH^{exp} fällt kontinuierlich bei formaler Kondensation eines aromatischen Ringes: Benzol ($\Delta H^{\text{exp}} = -26.32$ kJ/mol), Anthracen ($\Delta H^{\text{exp}} = -49.47$ kJ/mol), Benz[a]anthracen ($\Delta H^{\text{exp}} = -68.91$ kJ/mol) (siehe Abbildung 36). Dieser Trend wird von COSMO-RS wiedergegeben.

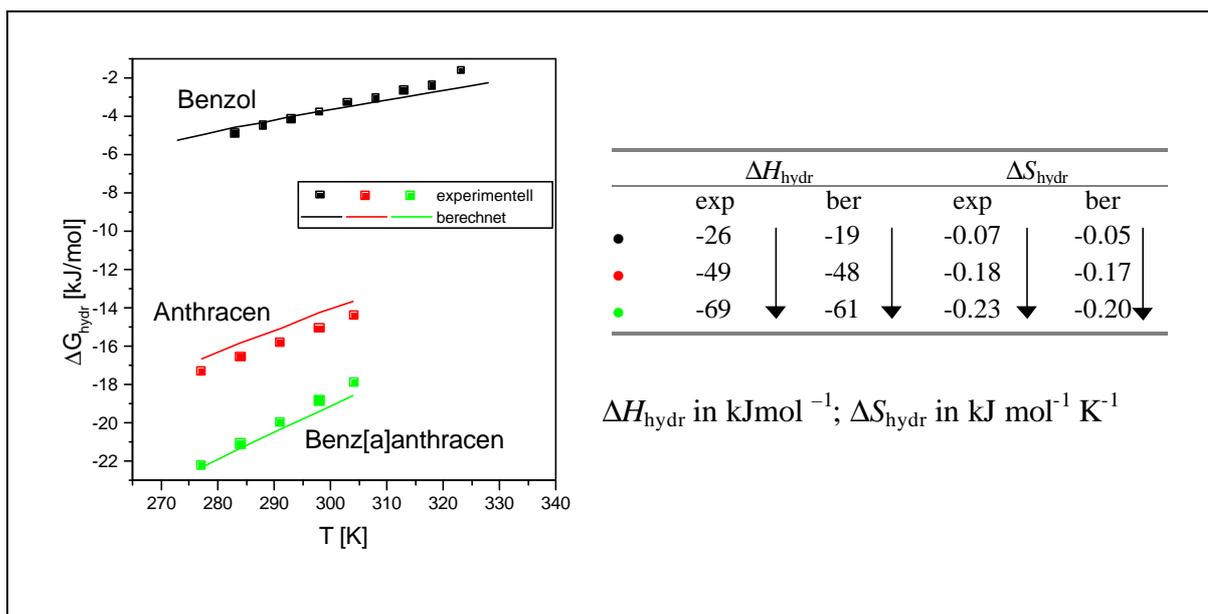
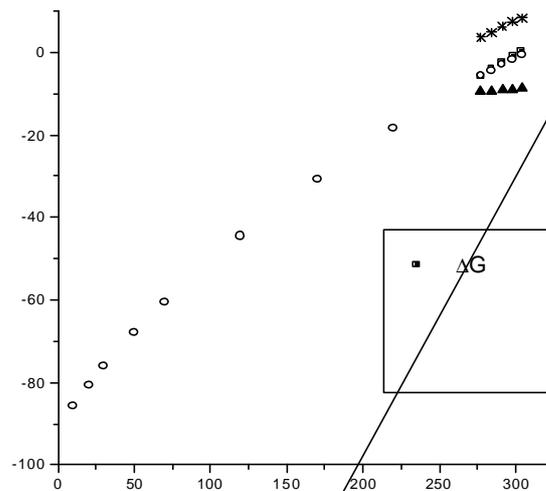


Abbildung 36: Berechnung der Enthalpie ΔH_{hydr} und Entropie ΔS_{hydr} aus dem Temperaturverlauf der freien Enthalpie ΔG_{hydr} .

Für eine zweite Reihe strukturell verwandter Moleküle (Abbildung 37 auf Seite 4-8) prognostiziert COSMO-RS aus der linearen Regression wiederum steigende ΔH -Werte bei Kondensation eines aromatischen Ringes gemäß der Erwartung. Auf den ersten Blick erscheint die Übereinstimmung mit den experimentellen Daten gut. Allerdings liefert die lineare Regression der experimentellen Daten für 1-Methylphenanthrene einen größeren ΔH -Wert als für 1-Methylnaphthalene. Dies bedeutet, dass sich die Regressionsgeraden für die experimentellen Daten bei einer Temperatur $T < 0$ °C schneiden. Glaubt man der physikalisch sinnvoll erscheinenden COSMO-RS-Vorhersage, so kann die Richtigkeit der experimentell ermittelten Temperaturabhängigkeit von 1-Methylnaphthalen oder 1-Methylphenanthren angezweifelt werden. Diese Beobachtung steht in Übereinstimmung mit den Befunden aus der Optimierungsprozedur. Hier wurde gezeigt, dass vor allem bei kleinen Henry-Konstanten die Fehler statistisch größer sind als bei den größeren Henry-Konstanten.

Abbildung 37: Berechnung der Enthalpie und Entropie aus dem Temperaturverlauf der freien Enthalpie.

Mit COSMO-RS/AM1 wird über einen größeren Temperaturbereich ein nicht-lineares Verhalten von ΔG_{hydr} vorhergesagt. So liefern Simulationsrechnungen für die unterkühlte Schmelze von Benz[a]anthracen im Tieftemperaturbereich $T = 1.0 \text{ K}$ bis 30.0 K die Werte $\Delta H_{\text{hydr}} = -91.26 \text{ kJ mol}^{-1}$ und $\Delta S_{\text{hydr}} = -0.52 \text{ kJ mol}^{-1} \text{ K}^{-1}$, die erheblich von den entsprechenden Ergebnissen im Temperaturbereich 277 K bis 304 K (Tabelle 24) abweichen. Abbildung 38 zeigt am Beispiel dieser Verbindung experimentelle und berechnete freie Hydratationsenthalpien mit den entsprechenden elektrostatischen und nicht-elektrostatischen Anteilen im oberen Temperaturbereich.



Anhand der vorgestellten Rechnungen konnte gezeigt werden, dass mit Hilfe der verwendeten COSMO-RS/AM1-Parametrisierung es nun erstmals möglich ist, Henry-Konstanten von aromatischen Kohlenwasserstoffen qualitativ, wenn nicht sogar quantitativ *temperaturabhängig* direkt aus der chemischen Struktur vorherzusagen. Aus dem Temperaturverlauf der freien Hydratationsenthalpie eines Stoffes können dann dessen Hydratationsenthalpie und Entropie mittels linearer Regression bestimmt werden.

Durch Betrachtung homologer Reihen können experimentell ermittelte Henry-Konstanten bewertet werden. Eine Aufspaltung der freien Hydratationsenthalpie in elektrostatische und nicht-elektrostatische Anteile zeigt, dass die elektrostatische Stabilisierung der Solvation bei Vergrößerung des aromatischen Systems zunimmt, und die nicht-elektrostatische Destabilisierung innerhalb der untersuchten homologen Reihe abnimmt.

Im nachfolgenden Anwendungsbeispiel wird nicht *eine* physikochemische Stoffeigenschaft *einer* Substanzklasse temperaturabhängig untersucht, sondern es werden *verschiedene* Stoffeigenschaften von Stoffen mit unterschiedlichen funktionellen Gruppen vergleichend analysiert.

4.2 Temperaturabhängigkeit von S_W , P_V und H' am Beispiel von vier Chemikalien

COSMO-RS/AM1 und COSMO-RS/DFT wurden verwendet, um Henry-Konstanten, Dampfdrücke und Wasserlöslichkeiten von Propan, Benzaldehyd, Acetophenon und 1,4-Dioxan in einen Temperaturbereich von 0°C bis 95°C zu berechnen. Für diese vier Substanzen wurden die berechneten Eigenschaften mit alternativen Berechnungsmethoden verglichen. Die Diskussion beinhaltet auch Vergleiche mit experimentellem Datenmaterial. Die Ergebnisse zeigen erstaunliche Qualitätsunterschiede der Methoden.

Nach Mackay [8] sind alle drei physikochemischen Eigenschaften zusammen zur Vorhersage der Stoffverteilung in der Umwelt hilfreich. Diese makroskopischen Eigenschaften hängen von externen Umweltbedingungen wie Temperatur und Druck ab. Mackay und Shiu [4] schätzten eine Verdopplung von H bei einer Temperaturerhöhung von 10°C auf 65°C, betonten aber, dass dieser Wert sehr stark vom gelösten Molekül abhängen kann und deshalb nicht generalisiert werden darf. Burkhard et al. [5] schätzten eine Zunahme von H um eine Größenordnung bei einer Temperaturerhöhung um 25°C, wohingegen Staudinger et al. [6] eine mittlere Zunahme um 60 % pro 10°C aufgrund eines Vergleichs verschiedener

Kohlenwasserstoffe vorhersagen. Temperaturabhängige Dampfdrücke werden normalerweise mit Hilfe von abgewandelten Formulierungen der Clausius-Clapeyron-Gleichung modelliert, wie z. B. mit der drei-Parameter-Antoine-Gleichung.

Kompliziert ist auch die Temperaturabhängigkeit der Wasserlöslichkeit, da diese bei Temperaturänderung weder streng monoton steigt noch abnimmt. Vielmehr ist bekannt, dass einige organische Verbindungen eine minimale Löslichkeit zwischen 285 K und 320 K [98,99] aufweisen. Die meisten temperaturabhängigen Messungen werden nur für bestimmte Substanzklassen durchgeführt. Deshalb ist der Bedarf an zuverlässigen Vorhersagemodellen groß.

Die vier analysierten Substanzen Propan, Benzaldehyd, Acetophenon und 1,4-Dioxan wurden gewählt, da die oben erwähnten Daten experimentell zwischen 0°C and 100°C verfügbar sind. Propan wurde ausgewählt, da es bei Raumtemperatur gasförmig ist. Die anderen Verbindungen sind abgesehen von Acetophenon, dessen Festpunkttemperatur 19°C beträgt, über den gesamten betrachteten Temperaturbereich flüssig. Die Substanzen offerieren strukturelle Variabilität: drei Ringverbindungen (aromatisch und nicht-aromatisch) mit Sauerstoff in verschiedenen Funktionalitäten (Ether, Aldehyd und Keton).

Methoden:

Erste 3D-Geometrien wurden mit der SYBYL-Software [100] generiert, mit Kraftfeldmethoden voroptimiert und mit semiempirischen AM1-Rechnungen (MOPAC [101]) verfeinert. Schließlich wurden sowohl Gasphasen- als auch COSMO-Geometrieoptimierungen mit DMol³ [57,58] mit dem BP-Funktional und dnp-Basissätzen (vergleiche Kapitel 3.5.1 Seite 3-41) durchgeführt. Zur Berechnung der physikochemischen Eigenschaften wurde COSMO-RS98.31 [102] mit den DMol- bzw. den neuen AM1-Parametern (siehe Tabelle 15 Seite 3-40) verwendet.

Mit der ChemProp-Software [93] wurden die Stoffeigenschaften auch mit Gruppenbeitragsverfahren berechnet. Temperaturabhängige Berechnungen nach Lee and Kesler [103] werden im folgenden P_V^{Lee} genannt. An dieser Stelle soll darauf hingewiesen werden, dass Acetophenon unterhalb des Schmelzpunkts (19°C) als unterkühlte Schmelze [104] behandelt wird. Wasserlöslichkeiten von Propan, Benzaldehyd und Acetophenon wurden mit der KOW-UNIFAC-Methode [89] bestimmt (S_w^{UNIFAC}). Dieses Verfahren konnte wegen fehlender Parametrisierung nicht für 1,4-Dioxan verwendet werden. Temperaturabhängige Henry-Konstanten ($H^{\text{PV/SWH}}$) ergeben sich danach als Quotient aus Wasserlöslichkeit (S_w^{UNIFAC}) und Dampfdruck (P_V^{Lee}) (siehe Gleichung 64 Seite 2-21).

Ergebnisse:

Berechnete und experimentelle Größen von Propan sind in Tabelle A-3 (Seite A-28 im Anhang) aufgeführt. Die Henry-Konstante variiert um 0.69 logarithmische Einheiten. Dieser Trend wird von COSMO-RS qualitativ richtig wiedergegeben, die vorhergesagte Änderung beträgt 0.33 (DFT) bzw. 0.42 (AM1) log-Einheiten bei Temperaturerhöhung von 0°C auf 75°C.

Obwohl dieser Trend gut wiedergegeben wird, überschätzt COSMO-RS/DFT die $\log H'$ -Werte ein wenig. Im Gegensatz hierzu werden logarithmische Henry-Konstanten, die aus Wasserlöslichkeiten und Dampfdrücken berechnet wurden ($\log H^{\text{PV/SW}}$), systematisch unterschätzt. Mittlere Abweichungen sind 0.27 für $\log H^{\text{DFT}}$ beziehungsweise -0.96 für $\log H^{\text{PV/SW}}$. Die Standardabweichung für $\log H^{\text{DFT}}$ (SD = 0.30) ist im Vergleich zur Alternativmethode nur ca. 33% (SD($\log H^{\text{PV/SW}}$) = 0.96). Diese große Standardabweichung kann durch die signifikante Überschätzung der zugrundeliegenden mit KOW-UNIFAC berechneten Wasserlöslichkeiten ($S_{\text{W}}^{\text{UNIFAC}}$) erklärt werden. Für diese ist die Standardabweichung (SD = 1.5) wieder größer als bei COSMO-RS/DFT (SD=1.09).

Die großen Standardabweichungen resultieren aus dem Umstand, dass Propan innerhalb des studierten Temperaturbereichs gasförmig ist. Sowohl COSMO-RS als auch KOW-UNIFAC behandeln Propan formal als Flüssigkeit, und deshalb ist die Beschreibung der Wasserlöslichkeit S_{W} wegen nicht enthaltener Verdampfungseffekte fehlerhaft. Im Falle gasförmiger Stoffe gilt die Beziehung zwischen H' , P_{v} , and S_{W} wie sie in Gleichung 64 (Seite 2-21) formuliert ist nicht. In der allgemeingültigen Form steht anstelle des Dampfdrucks der Partialdruck. Für Flüssigkeiten und Feststoffe ist der Partialdruck in erster Näherung gleich dem Dampfdruck, für Gase ist der Partialdruck allerdings gleich dem Normaldruck $P^* \approx 10^5$ Pa. Die physikalisch korrekte logarithmische Propan-Wasserlöslichkeit muss deshalb um die Differenz aus Dampfdruck und Normaldruck korrigiert werden. Die Standardabweichungen verringern sich dadurch auf 0.47 (AM1), 0.31 (DFT) und 1.03 (UNIFAC). Interessant ist der Modellvergleich der Temperaturabhängigkeit. S_{W} von Propan sinkt um 0.75 log-Einheiten bei einer Temperaturerhöhung von 5 °C auf 60 °C. Alle Modelle zeigen den richtigen Trend abnehmender Wasserlöslichkeiten bei steigender Temperatur (siehe Abbildung 39 auf Seite 4-12), generieren aber eine geringere Temperaturabhängigkeit (z.B. $\Delta \log S_{\text{W}}^{\text{UNIFAC}}(5-60^\circ\text{C}) = 0.4$). COSMO-RS sagt eine geringere Temperaturabhängigkeit voraus aber die absolute Übereinstimmung mit den experimentellen Wasserlöslichkeitsdaten ist besser.

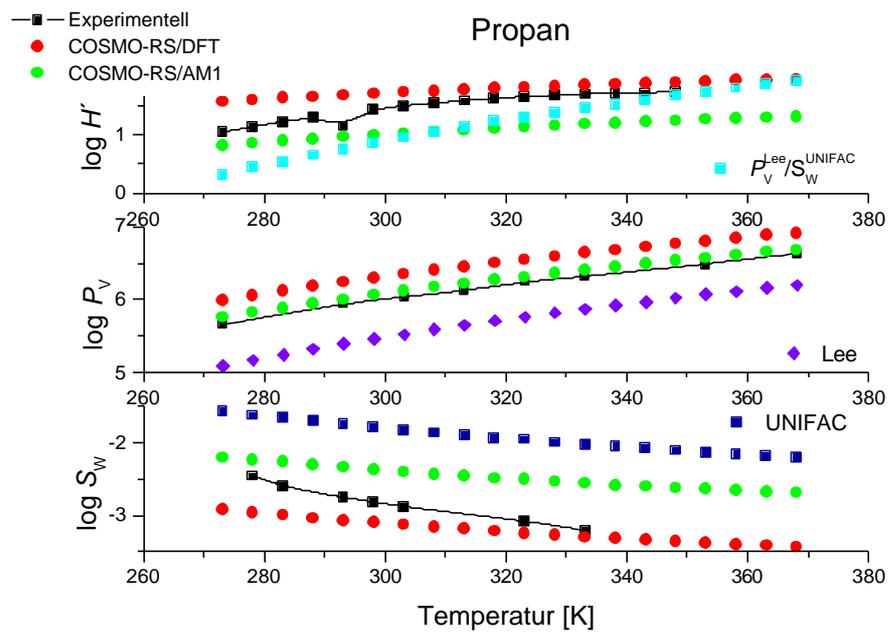


Abbildung 39: Temperaturverlauf der Verteilungskoeffizienten von Propan.

Sowohl Trend als auch Absolutwerte des Dampfdrucks von Propan werden mit COSMO-RS (sowohl mit DFT als auch mit AM1) gut reproduziert. COSMO-RS/AM1 zeigt hier sogar die kleinste systematische Abweichung. Absolute Fehler der Henry-Konstante verringern sich bei steigender Temperatur für COSMO-RS/DFT, bei der COSMO-RS/AM1-Vorhersage sind die Verhältnisse entgegengesetzt (steigender Fehler bei steigender Temperatur). Die Dampfdruckfehler aller Methoden bleiben über den betrachteten Temperaturbereich nahezu konstant. Im Falle der aus Dampfdruck (Lee) und Wasserlöslichkeit (UNIFAC) berechneten Henry-Konstanten wird eine Fehlerkompensation beobachtet (vgl. SE in Tabelle A-3 auf Seite A-28 im Anhang).

Benzaldehyd hat ein geringeres Bestreben sich in der Gasphase anzureichern als in der wässrigen Phase. Dieses Verhalten spiegelt sich in der negativen logarithmischen Henry-Konstante wieder. Abbildung 40 zeigt die temperaturabhängige Auftragung der Stoffeigenschaften von Benzaldehyd. Hier ist die COSMO-RS/DFT-Vorhersage für alle physikochemischen Eigenschaften sehr gut, wohingegen AM1-Rechnungen nur für Henry-Konstanten sehr gute Ergebnisse liefern (vgl. Tabelle A-4 auf Seite A-29 im Anhang). Der Fehler bei Dampfdruck und Wasserlöslichkeit beträgt ca. eine logarithmische Einheit.

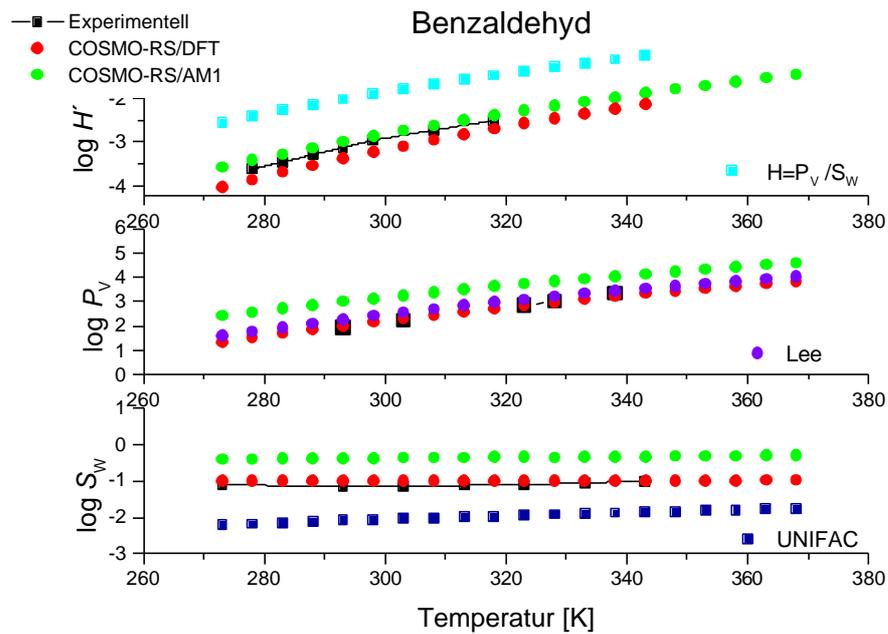


Abbildung 40: Temperaturabhängige experimentelle und berechnete Stoffeigenschaften von Benzaldehyd.

Experimentelle logarithmische Henry-Konstanten, Dampfdrücke und Wasserlöslichkeiten werden von COSMO-RS/DFT mit Standardfehlern von 0.24, 0.07, und 0.11 reproduziert. Es kann angenommen werden, dass die Fehler, speziell Letztere, innerhalb der experimentellen Bestimmungsgenauigkeit liegen. COSMO-RS reproduziert den Trend steigender Wasserlöslichkeit bei steigender Temperatur, die absolute Änderung der Wasserlöslichkeit ($\Delta \log S_w = 0.09$) wird allerdings etwas unterschätzt. Es gilt: $\Delta \log S_w^{\text{DFT}} = 0.01$, im Vergleich zu $\Delta \log S_w^{\text{UNIFAC}} = 0.35$ zwischen 0°C und 70°C . Den experimentellen Trend reproduziert AM1 am besten ($\Delta \log S_w^{\text{AM1}} = 0.06$), wenn auch die Absolutwerte stärker von den experimentellen Daten abweichen. Im Gegensatz hierzu überschätzt KOW-UNIFAC die S_w -Temperaturabhängigkeit von Benzaldehyd. Der Standardfehler der KOW-UNIFAC-Methode ist um 0.80 log-Einheiten größer als bei COSMO-RS/DFT. Dieser Fehler spiegelt sich bei der mit S_w^{UNIFAC} berechneten Henry-Konstante ($\text{SE} = 1.10$) wieder.

Während im Falle des Benzaldehyds Dampfdruck und Wasserlöslichkeit systematisch leicht überschätzt werden, werden diese bei Acetophenon von COSMO-RS/DFT unterschätzt (siehe Tabelle A-5 auf Seite A-30 im Anhang). Die Standardfehler von $\log P_v^{\text{DFT}}$ and $\log S_w^{\text{DFT}}$ sind 0.21 bzw. 0.17. Hier hat die KOW-UNIFAC-Methode etwas kleinere Standardfehler, aber es wird ein falscher Trend für die Temperaturabhängigkeit von $\log S_w$ prognostiziert! COSMO-

RS/DFT zeigt wachsende Ungenauigkeit bei höheren Temperaturen. Während bei 20°C der Fehler nur 0.04 log-Einheiten beträgt, ist er 0.29 bei 80°C. Der Grund ist wieder die unterschätzte Temperaturabhängigkeit von S_w : Die experimentell ermittelte Änderung der Wasserlöslichkeit ($\Delta \log S_w = 0.27$) für $\Delta T = 60^\circ\text{C}$ ist um einen Faktor zehn größer als die mit COSMO-RS/DFT berechnete ($\Delta \log S_w = 0.02$). Für COSMO-RS/AM1 ergibt sich ein ähnliches Bild wie im Falle des Benzaldehyds: gute Reproduktion experimenteller Henry-Konstanten mit systematischen Abweichungen bei Dampfdruck und Wasserlöslichkeit.

Die vierte untersuchte Substanz ist Dioxan. Anhand dieser Verbindung soll beispielhaft die Visualisierung der COSMO-Abschirmladungsdichten gezeigt werden: In Abbildung 41 ist die Ladungsdichte von 1,4-Dioxan (Abbildung: mitte) dargestellt. Blaue Bereiche markieren positiv polarisierte Oberflächenabschnitte (negative Abschirmladungsdichte s) und rote Bereiche repräsentieren negativ polarisierte Abschnitte. Grün und gelb gefärbte Abschnitte sind nur schwach polarisiert. Es ist sehr gut zu erkennen, dass vor allem bei den Sauerstoffen stark polarisierte Regionen vorliegen. In der rechten Abbildung sind dagegen die Beträge der Abschirmladungsdichten, also ein direktes Maß für hydrophile Bereiche (großes $|s|$) und lipophile Regionen (kleines $|s|$) dargestellt. Die lipophilen Bereiche der Oberfläche sind rot dargestellt, wohingegen blaue und violette Abschnitte stark hydrophile Abschnitte der SAS markieren.

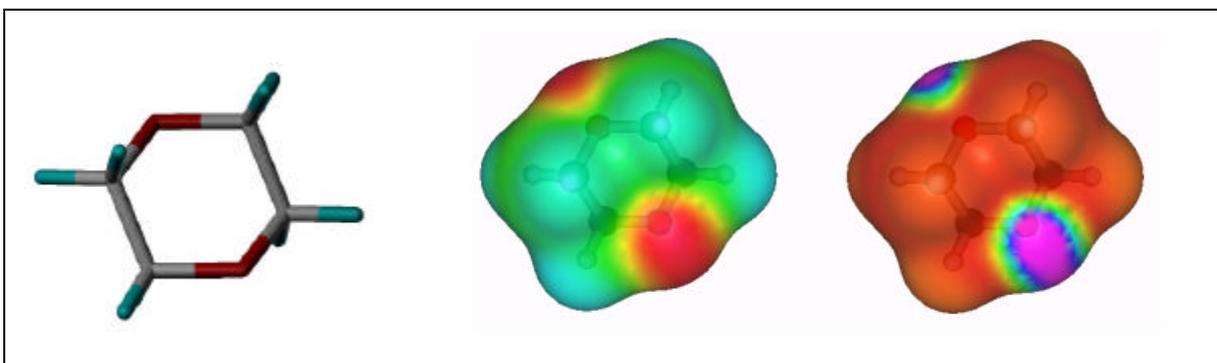


Abbildung 41: Visualisierung der COSMO-RS/DFT-Abschirmladungsdichten auf der SAS für 1,4-Dioxan ($\text{C}_4\text{H}_8\text{O}_2$).

Die experimentell bestimmten Henry-Konstanten von 1,4-Dioxan können mit COSMO-RS/DFT gut reproduziert werden (siehe nachfolgende Tabelle). Bei Raumtemperatur ist der absolute Fehler nur 0.1 log-Einheiten.

Tabelle 25: Henry-Konstante H' , Dampfdruck P_V [Pa] und Wasserlöslichkeit S_W [mol L⁻¹] von Dioxan mit Standardabweichung (SE) und mittlerem Fehler (MF). Alle Angaben in logarithmischen Einheiten.

T [K]	H'	H'^{AMI}	H'^{DFT}	P_V	P_V^{AMI}	P_V^{DFT}	P_V^{Lee}	S_W^{AMI}	S_W^{DFT}
273.15		-3.25	-4.40		3.81	2.69	3.15	0.68	0.74
278.15		-3.13	-4.25		3.93	2.83	3.29	0.69	0.72
283.15		-3.02	-4.11		4.04	2.96	3.41	0.68	0.70
288.15		-2.91	-3.97		4.15	3.09	3.54	0.68	0.68
293.15	-3.93	-2.80	-3.83	3.62	4.25	3.22	3.66	0.67	0.66
298.15	-3.71	-2.70	-3.70		4.35	3.34	3.77	0.67	0.65
303.15		-2.60	-3.58	3.83	4.45	3.46	3.88	0.67	0.63
308.15		-2.50	-3.46		4.54	3.57	3.98	0.66	0.62
313.15	-3.21	-2.41	-3.34	4.00	4.63	3.68	4.08	0.67	0.60
318.15		-2.32	-3.23		4.72	3.78	4.18	0.66	0.58
323.15	-3.12	-2.24	-3.12	4.22	4.80	3.88	4.28	0.66	0.57
328.15		-2.15	-3.01		4.89	3.98	4.37	0.66	0.56
333.15	-2.81	-2.07	-2.91		4.96	4.08	4.46	0.65	0.55
338.15		-1.99	-2.82	4.47	5.04	4.17	4.54	0.65	0.54
343.15	-2.62	-1.92	-2.72		5.12	4.26	4.62	0.65	0.53
348.15		-1.85	-2.63		5.19	4.35	4.70	0.65	0.52
353.15	-2.46	-1.78	-2.54		5.26	4.43	4.78	0.66	0.51
358.15		-1.71	-2.46		5.32	4.52	4.85	0.66	0.50
363.15		-1.65	-2.38		5.39	4.60	4.92	0.66	0.49
368.15		-1.58	-2.30		5.45	4.67	4.99	0.65	0.49
SE		0.86	0.09		0.61	0.35	0.06		
MF		0.85	-0.05		0.61	-0.34	0.06		

Mit steigender Temperatur steigt die Affinität von Dioxan zur Gasphase, und COSMO-RS reproduziert diesen Trend. Bei 50°C wird der absolute Fehler minimal ($\Delta \log H'^{\text{DFT}} = 0.00$) und steigt wieder mit wachsender Temperatur. Es ist bekannt, dass Dioxan mit Wasser beliebig mischbar ist, und deshalb ist die Berechnung von S_W nach Gleichung 97 (Seite 2-27) strenggenommen nicht mehr zulässig, da diese nur für unendliche Verdünnung gilt. Trotzdem ist die Berechnung von S_W sehr interessant, um die Anwendbarkeit von COSMO-RS in Grenzbereichen auszuloten. Bei Raumtemperatur ist $\log S_W^{\text{DFT}} = 0.65$. Das entspricht $4.47 \text{ mol L}^{-1} \cong 400 \text{ g L}^{-1}$. Wäre die Wasserlöslichkeit von Dioxan unbekannt, würde man erwarten, dass Dioxan und Wasser in fast gleichen Teilen (also unbegrenzt) mischbar sind.

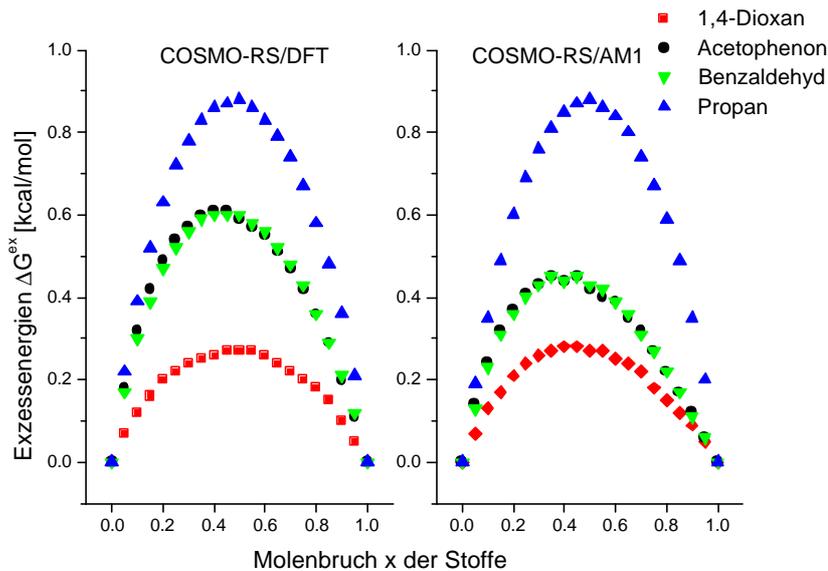


Abbildung 42: Exzessenergien [kcal/mol] des Stoff/Wassergleichgewichts in Abhängigkeit vom Molenbruch bei 25°C berechnet mit COSMO-RS/DFT und COSMO-RS/AM1.

Abbildung 42 zeigt die Exzessenergien ΔG^{ex} der Wasser/Substanzgleichgewichte für Molenbrüche von null bis eins bei 25°C. Die Exzessenergie ist ein Maß für die Abweichung einer Mischung vom idealen Verhalten. Für jeden Konzentrationspunkt x_i wird die Exzessenergie aus den chemischen Potenzialen m_i der Stoffe i in der Mischung und den chemischen Potenzialen der Reinstoffe m_i^* ($= m_i$ bei $x_i = 1$) nach Gleichung 119 berechnet.

$$\Delta G^{ex} = x_{H_2O}(m_{H_2O} - m_{H_2O}^*) + x_C(m_C - m_C^*)$$

Gleichung 119

Für Propan und Dioxan sagen COSMO-RS/AM1 und COSMO-RS/DFT ähnliche Exzessenergien voraus. Eine im Vergleich zur COSMO-RS/AM1-Prognose größere Abweichung vom idealen Verhalten prognostiziert COSMO-RS/DFT für Benzaldehyd und Acetophenon. Die Abbildung veranschaulicht, dass Dioxan die geringste Abweichung vom idealen Verhalten zeigt. Dies ist ein weiterer Hinweis für die beliebige Mischbarkeit von Dioxan und Wasser.

Verglichen mit der großen Anzahl von Chemikalien, die in der Umwelt gefunden werden, sind nur wenige temperaturabhängige experimentelle umweltrelevante Stoffeigenschaften in der Literatur zu finden. Aus diesem Grund ist COSMO-RS (vor allem mit der DFT-Parametrisierung) als Vorhersagemodell für kleine Moleküle von großer Bedeutung, um unbekannte Größen vorherzusagen. Auch COSMO-RS/AM1 ist schon im derzeitigen Entwicklungsstadium ein vielversprechendes Verfahren zur Vorhersage für Stoffeigenschaften-

ten, nicht nur bei Raumtemperatur, sondern im Temperaturbereich zwischen 0°C und 100°C. Allerdings ist anzumerken, dass noch zusätzliche Verbesserungen bei Carbonylen eingeführt werden könnten.

Mit COSMO-RS können relative Trends, sowohl für spezielle Stoffgruppen (siehe letztes Kapitel) als auch für strukturell verschiedene Moleküle, richtig wiedergegeben werden und dies häufig mit größerer Genauigkeit als mit Alternativmethoden.

4.3 Modellierung des Fischöl/Wasser-Verteilungskoeffizienten

Nachdem schon im letzten Kapitel versucht wurde eine Stoffeigenschaft (die Wasserlöslichkeit) vorherzusagen, die nicht explizit in der Optimierung berücksichtigt wurde, wird COSMO-RS nun zur Modellierung des Fischöl/Wasser-Verteilungskoeffizienten K_{FW} von acht Stoffen (siehe Tabelle 26 Seite 4-20) angewendet. Die acht Substanzen sind keiner speziellen Klasse zuzuordnen, sondern sie enthalten verschiedene funktionelle Gruppen. Für die Stoffe wurden die Fischöl/Wasser-Verteilungskoeffizienten am Umweltforschungszentrum Leipzig-Halle im Labor von Herrn Dr. Paschke und Mitarbeitern bestimmt. Das Messverfahren und die genauen Bedingungen werden im Anhang A.3 (Seite A-26 ff.) beschrieben.

Clausen zeigte für 32 chlorierte und fluorierte Kohlenwasserstoffe, dass experimentelle Fischöl/Wasser-Verteilungskoeffizienten sehr gut mit experimentellen Oktanol/Wasser-Verteilungskoeffizienten korrelieren ($r^2 = 0.99$) [107]. Es gilt:

$$\log K_{FW} = \log K_{OW} + 0.40 \quad \text{Gleichung 120}$$

Nachfolgende Abbildung zeigt eine Gegenüberstellung der K_{FW} - K_{OW} -Korrelation anhand des Clausen-Moleküldatensatzes und anhand der neuen gemessenen Werte.

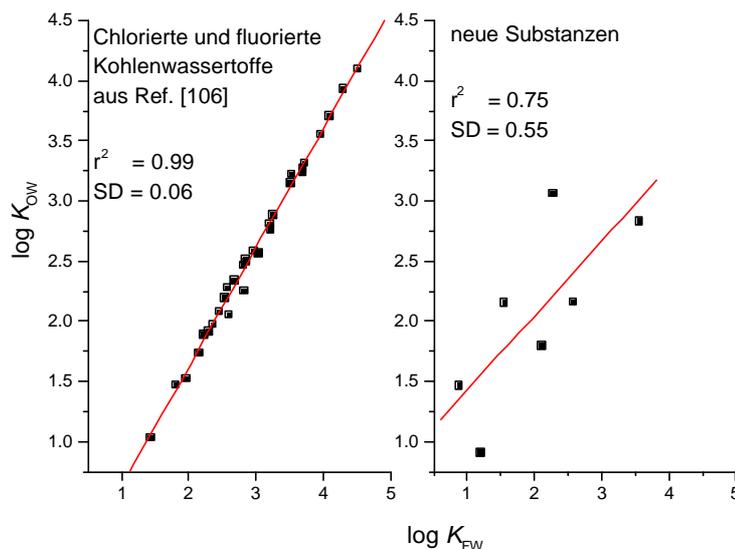


Abbildung 43: Korrelation experimenteller Fischöl/Wasser-Verteilungskoeffizienten (K_{FW}) mit experimentellen Oktanol/Wasser-Verteilungskoeffizienten (K_{OW})

Die Abbildung zeigt, dass der K_{FW} und der K_{OW} für die neu gemessenen Substanzen mit Stickstoff- und Sauerstoffatomen nicht mehr perfekt linear korreliert sind. Deshalb wurde versucht den K_{FW} mit COSMO-RS abzuschätzen.

Für die Rechnungen wurden COSMO-RS98 [102] mit der neuen AM1-Parametrisierung (siehe Tabelle 15 Seite 3-40) und das kommerziell erhältliche COSMOtherm [105] verwendet. Grundlage für die COSMOtherm-Rechnungen bildeten quantenchemische DFT-Rechnungen. Die Gasphasen- und COSMO-Rechnungen wurden mit der TURBOMOLE-Software [106] durchgeführt. Es wurde das BP-Funktional (BP = B88-PW91) [59,60] mit TZVP-Basen [63] und der RI-Option (RI = Resolution of Identity) [64,65] verwendet (vergleiche Kapitel 3.5.1 Seite 3-41). Erste 3-D-Strukturen der Moleküle wurden mit SYBYL 6.5 [100] generiert und mit einer implementierten Kraftfeldmethode voroptimiert.

Noch einmal zur Theorie von Verteilungskoeffizienten: Ist das chemische Potential m eines Stoffes j in zwei fluiden Phasen 1 und 2 bekannt, so berechnet sich der zugehörige logarithmische Verteilungskoeffizient P_{1-2} nach Gleichung 121:

Der Quotient aus den molaren Volumina v_x der zwei fluiden Komponenten korrigiert die Gleichgewichtskonstante vom Konzentrationsmaß Molenbruch zum Konzentrationsmaß Molarität (siehe auch Gleichung 84 bis Gleichung 86 in Kapitel 2.3.3 auf Seite 2-25 ff.).

Wird Gleichung 121 zur Berechnung des Fischöl/Wasser-Verteilungskoeffizienten herangezogen, sind $m_j^{(1)}$ and $m_j^{(2)}$ die chemischen Potentiale der Substanz j in Fischöl und in Wasser. Soll der Fischöl/Wasser-Verteilungskoeffizient mit COSMO-RS abgeschätzt werden, so treten vor allem zwei Probleme auf: Erstens ist Fischöl eine Mixtur aus zahlreichen Stoffen, und zweitens ist die genaue chemische Zusammensetzung des Fischöls nicht exakt definierbar, da das Verhältnis der Einzelkomponenten schwankt. Deshalb muss eine Modellphase für Fischöl definiert werden. Clausen gibt in seiner Dissertationsschrift [107] die Ergebnisse einer chemischen Fischölanalyse (Medizinaldorschlebertran) an. Die zwei Hauptkomponenten der insgesamt 19 Bestandteile sind: 9-Octadecensäuremethylester (Ölsäuremethylester) (17 mol-%) (**1**) und Hexadekansäuremethylester (Palmitinsäuremethylester) (15.5 mol-%) (**2**).

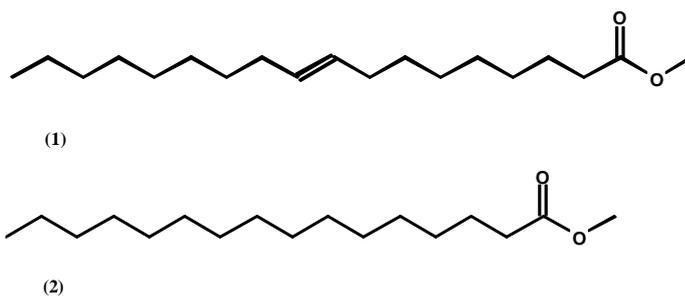


Abbildung 44: Hauptkomponenten einer Fischölprobe nach Clausen.

Mit COSMO-RS ist es prinzipiell möglich, noch weitere Komponenten zu berücksichtigen, aber um die DFT-Rechenzeitanforderungen in Grenzen zu halten wurden nur zwei alternative Ansätze ausprobiert. Als erstes wurde ausschließlich Komponente **1** als Modellsubstanz für die Fischölphase herangezogen ($\rightarrow \log K_{\text{ÖW}}$) und dann alternativ eine äquimolare Mischung der Substanzen **1** und **2** ($\rightarrow \log K_{\text{ÖPW}}$). Erste AM1-Rechnungen identifizierten das E-Isomer von **1** als das stabilere Isomer in der Gasphase, und deshalb wurde dieses für die Rechnungen herangezogen. Prinzipiell können mit COSMO $therm$ auch Isomerenmischungen behandelt werden, doch schien dies unter dem Aspekt, dass der Methylsäureester seinerseits nur als Modellsubstanz für den Dorschlebertran fungiert, als zu aufwendig. Bei zukünftigen genaueren Berechnungen sollten zuerst mehrere Komponenten des Fischöls berücksichtigt werden, bevor die verschiedenen Isomere der einzelnen Komponenten im Detail aufgeschlüsselt werden.

Tabelle 26 zeigt, dass experimentelle Fischöl/Wasser-Verteilungskoeffizienten mit COSMO-RS und den Modellsubstanzen gut vorhergesagt werden können. Die Gesamtstatistik der beiden alternativen Modellphasen ist von ähnlicher Qualität.

Tabelle 26: Experimentelle (Exp) und berechnete Verteilungskoeffizienten. *Mittelwert aus drei Einzelmessungen; **aus der Medchem-Datenbank; ¹⁻³ mit der linearen Beziehung nach Clausen berechnet; ¹ mit experimentellem K_{OW} ; ² K_{OW} mit ALOGP berechnet; ³ K_{OW} mit CLOGP berechnet. Mit Standardabweichungen (SD) und Fehlerquadratsummen (SOR) bzgl. exp. K_{FW} -Daten

	log K_{FW}	K_{OW}	log $K_{ÖW}$	log $K_{ÖPW}$	log $K_{ÖPW}$	log $K_{ÖPW}$	log K_{FW}	log K_{FW}	log K_{FW}
	Exp*	Exp**	DFT	DFT	AM1_2	AM1_1	Clausen ¹	Clausen ²	Clausen ³
Phenol	0.88	1.46	0.67	0.71	0.56	1.05	1.86	1.82	1.88
2-Chlorphenol	1.55	2.15	2.06	2.10	1.96	2.28	2.55	2.52	2.56
2,4-Dichlorphenol	3.28	3.06	2.79	2.83	2.93	3.14	3.46	3.21	3.37
Anilin	1.20	0.90	0.93	0.98	-0.08	-0.15	1.30	1.35	1.32
3-Nitroanisol	2.58	2.16	2.36	2.39	2.05	1.46	2.56	0.68	2.50
Tetrachlormethan	3.56	2.83	2.95	3.00	3.28	3.08	3.23	3.24	3.28
2-Nitrophenol	2.11	1.79	1.89	1.92	1.47	1.23	2.19	0.51	2.25
2,4-Dinitrophenol	-	1.67	2.08	2.11	1.95	0.97	2.07	-0.80	2.22
SOR			1.10	1.08	3.11	5.36	3.45	8.91	3.29
SD			0.40	0.39	0.67	0.88	0.70	1.13	0.69

Der Vergleich mit den Mittelwerten aus den drei Einzelexperimenten (ohne das nicht messbare 2,4-Dinitrophenol) zeigt Fehlerquadratsummen von 1.10 und 1.08 für log $K_{ÖW}$ bzw. $K_{ÖPW}$ (berechnet mit COSMO-RS/DFT). COSMO-RS/AM1 kann die Qualität der DFT-Parametrisierung nicht erreichen, allerdings macht sich bei den Nitroverbindungen der Core-Ladungsausgleich (AM1_2 vergleiche Kapitel 3.4.2. Seite 3-28 ff.) positiv bemerkbar. Der log K_{FW} von Phenol wird dagegen (trotz OH-Ladungsausgleich) mit der AM1_1 Parametrisierung besser wiedergegeben. Eine Anwendung der linearen Korrelation nach Clausen (Gleichung 120 Seite 4-17) zeigt bei Verwendung experimenteller Oktanol/Wasser-Verteilungskoeffizienten ähnliche Qualität wie COSMO-RS/AM1. Wird die Clausen-Beziehung allerdings mit berechneten K_{OW} -Werten (ALOGP [108]) verwendet, können die experimentellen Daten der sieben Verbindungen nur unbefriedigend wiedergegeben werden (SOR = 8.91). Eine bessere Korrelation ergibt sich auf Basis der mit CLOGP berechneten Oktanol/Wasser-Verteilungskoeffizienten (SOR = 3.29).

Abbildung 45 zeigt die Fehler von COSMO-RS/DFT, COSMO-RS/AM1 und der nach Clausen umgerechneten CLOGP-Werte.

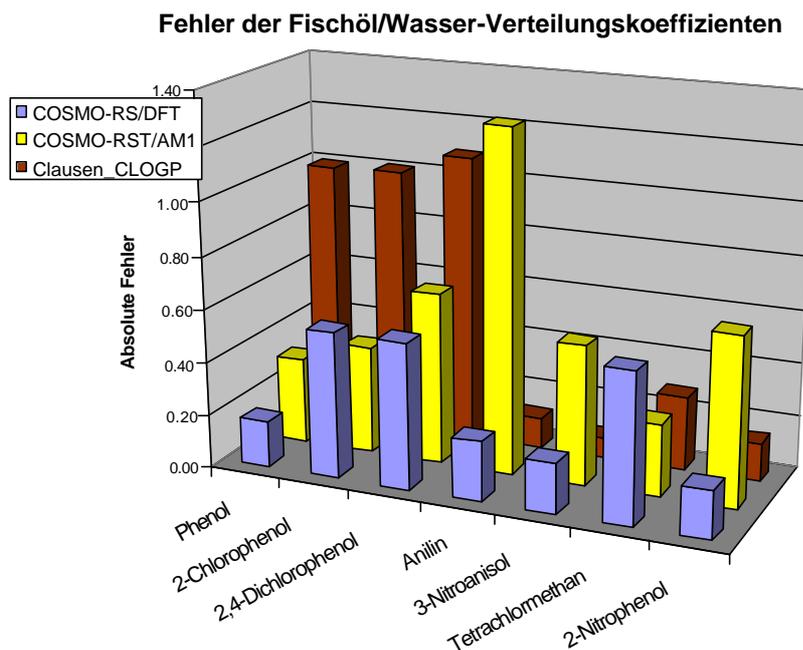


Abbildung 45: Vergleich der Fehler dreier Rechenmodelle zur Berechnung von $\log K_{FW}$ -Werten. Neben den Fehlern der COSMO-RS-Verfahren sind auch die Fehler der mit der linearen Beziehung nach Clausen umgerechneten CLOGP-Daten angegeben

Die Modellierung des Fischöl/Wasser-Verteilungskoeffizienten durch den $K_{ÖW}$ und den $K_{ÖPW}$ gelingt mit COSMO-RS/DFT für die experimentell untersuchten Substanzen besser als mit der Berechnung nach der Clausen-Korrelation und experimentellen Oktanol/Wasser-Verteilungskoeffizienten. Diese ist wie anfangs gezeigt nur für die spezielle Stoffgruppe chlorierter und fluoriertes Kohlenwasserstoffe mit hoher Vorhersagekraft anwendbar. Die AM1-Parametrisierung mit den Core-Ausgleichsladungen ist der DFT-Parametrisierung unterlegen, doch besitzt sie immer noch höhere Vorhersagekraft als die mit K_{OW} -Werten angewendete lineare Korrelation nach Clausen.

Es zeigt sich, dass Oktadecensäure und eine 50%-Mischung mit Palmitinsäuremethylester gute COSMO-RS-Modellphasen für die Fischölphase sind.

4.4 Abschätzung des Blut/Hirnschranken-Übergangskoeffizienten $\log P_{BB}$

Die Befähigung eines Wirkstoffs, die Blut/Hirnschranke (*engl.: BBB blood brain barriere*) zu überwinden ist eine wichtige pharmakokinetische Eigenschaft bei der Erforschung CNS-aktiver Arzneien (CNS = Central Nervous System). Die experimentelle *in vivo* Bestimmung

des $\log P_{BB}$ [= $\log(c_{\text{brain}}/c_{\text{blood}})$] ist naturgemäß schwierig und zeitaufwendig und kann deshalb nicht in high-throughput-Messverfahren integriert werden. Der Bedarf an effizienten theoretischen CADD-Modellen (CADD = Computer Aided Drug Design) ist deshalb Gegenstand intensiver Forschung. Ältere QSAR-Ansätze bedienten sich verschiedener physikochemischer Stoffeigenschaften (z. B. des Oktanol/Wasser-Verteilungskoeffizienten [109]), molekularer Größendeskriptoren [110,111] oder Solvatochromer Parameter [112,113]. Auch an der Auswertung molekularer Deskriptoren mit modernen statistischen Methoden wurde gearbeitet [114,115].

In den letzten Jahren wurde gezeigt, dass der $\log P_{BB}$ sehr gut mit den polaren Anteilen der Moleküloberfläche (engl. PSA = Polar Surface Area) korreliert. Die PSA ist hier als Summe der Stickstoff-, Sauerstoff- und polaren Wasserstoffanteile der Moleküloberfläche definiert. Kansy et al. ermittelten eine lineare Abhängigkeit des $\log P_{BB}$ von der PSA in Kombination mit dem molekularen Volumen ($r^2 = 0.71$, $SE = 0.45$) [116]. Eine genauere Analyse der dynamischen PSA (= mittlere PSA aller energetisch tief-liegenden Konformere einer Komponente) durch Kelder et al. führte zu einer noch besseren Korrelation ($r^2 = 0.83$) [117]. Potenzielle Wirkstoffe haben nicht selten hohe Molmassen und beinhalten oft verschiedene funktionelle Gruppen. Die COSMO-RS/AM1-Methode scheint deshalb zur Abschätzung des $\log P_{BB}$ grundsätzlich geeignet zu sein. Der Blut/Hirn-Verteilungskoeffizient kann auf zwei Arten mit COSMO-RS bestimmt werden. Einerseits liefert COSMO-RS zwanglos molekulare Deskriptoren wie die Moleküloberfläche, das Molekülvolumen, die Oberflächenladungsdichten auf der SAS, die \mathbf{s} -Momente und Wasserstoffbrückenbindungsdonor- bzw. Akzeptormomente, mit denen QSAR Studien durchgeführt werden können. Andererseits kann der $\log P_{BB}$ anstatt mit dem Oktanol/Wasser-Verteilungskoeffizient auch mit anderen Verteilungskoeffizienten korreliert werden. Im Folgenden werden Abschätzungsversuche nach den zwei alternativen Ansätzen anhand eines Datensatzes von 60 Verbindungen [118] vorgestellt. Keserü et al. berechneten hier die freien Hydratationsenthalpien mit einem selbstproduzierten C-code (GSOLV) nach dem *Generalized Born (GB)* Ansatz und korrelierten diese mit den experimentellen Daten. Nach Ausschluss der fünf größten Ausreißer ermittelten die Autoren eine lineare Beziehung:

$$\log P_{BB} = 0.035G_{sol}^{GB} + 0.2592 \quad \text{Gleichung 122}$$

mit G_{sol}^{GB} in kJ/mol. Der quadratischen Regressionskoeffizient ist $r^2 = 0.72$ und der Standardfehler ist $SE = 0.37$. Die mit COSMO-RS ermittelten Ergebnisse werden mit diesen statistischen Parametern verglichen.

Nachfolgende Abbildung zeigt die Lewis-Strukturformeln der 60 Substanzen.

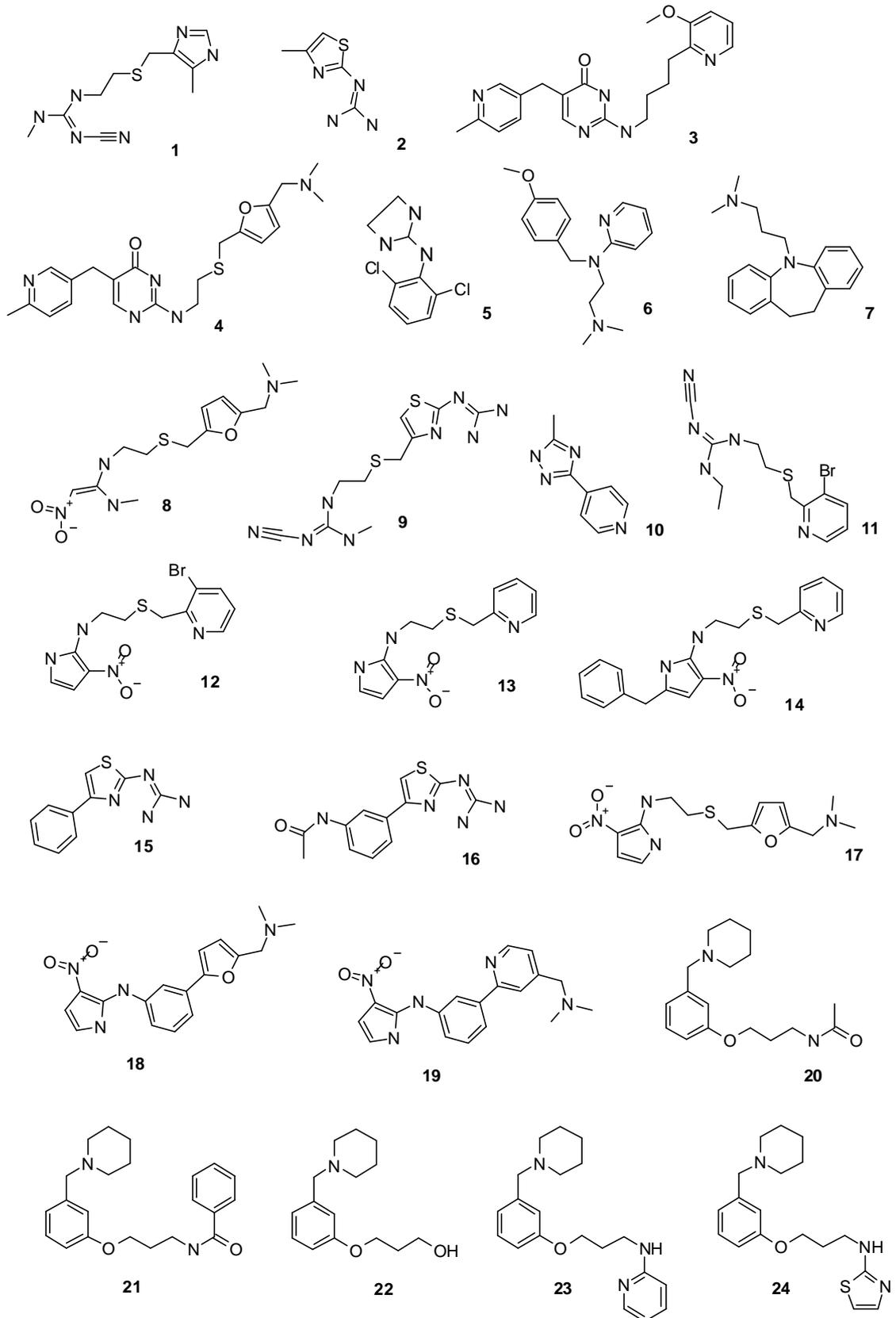
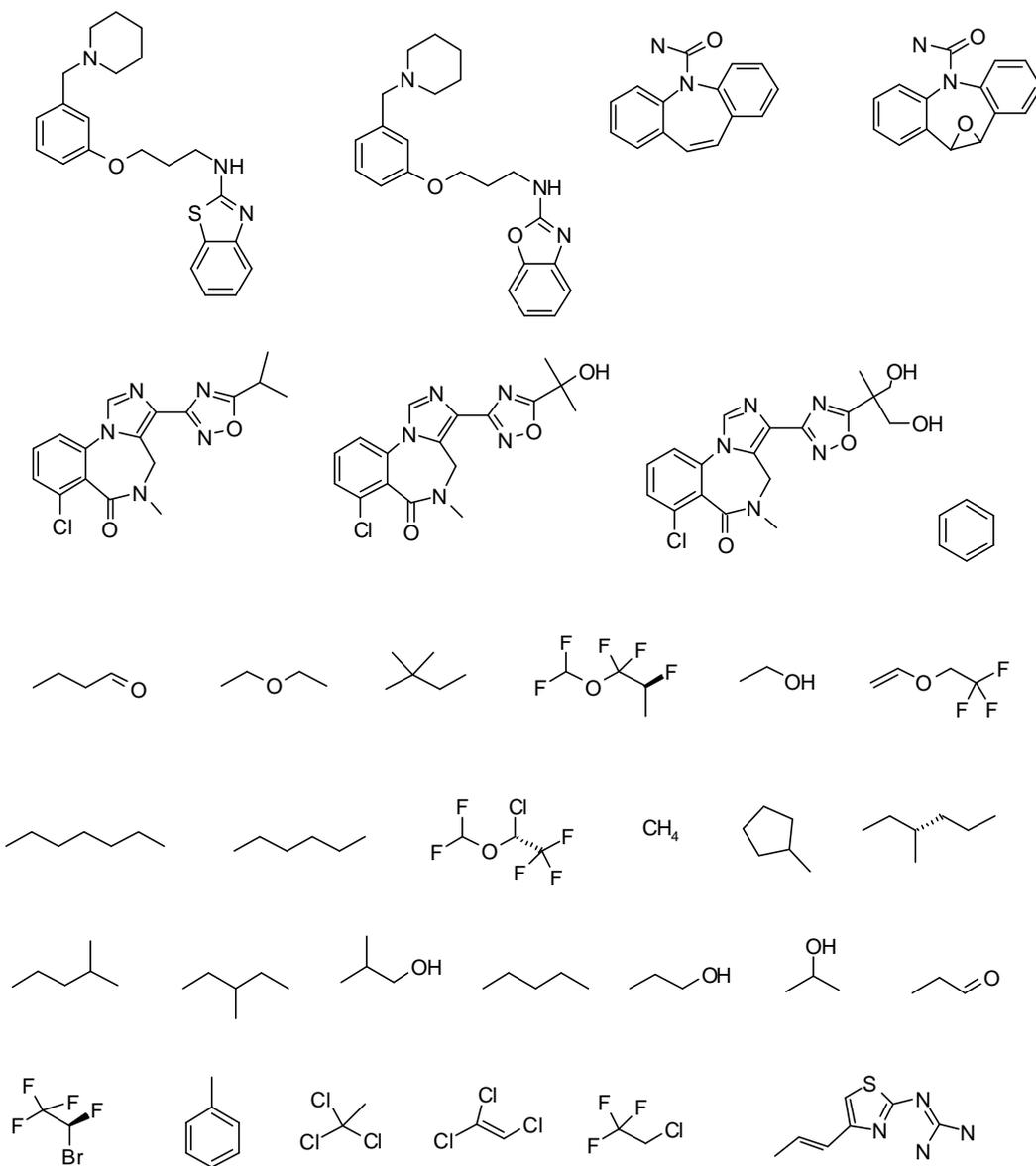


Abbildung 46: Moleküle des Datensatzes (Fortsetzung nächste Seite).



realen Fall der Überschreitung der Blut/Hirnschranke muss ein Stoff aus dem Blutkreislauf durch eine Zellmembran ins Hirngewebe diffundieren. Es liegt nahe, die beiden Phasen (Blut und Zellmembran) für die COSMO-RS-Modellierung wiederum in einer sehr einfachen Näherung durch geeignete Modellsubstanzen zu ersetzen. Da eine Zellmembran aus vielen Komponenten bestehen kann, in die verschiedene Proteine eingelagert sind, wurde eine typische Zellmembrankomponente aus der Gruppe der Phospholipide: das Phosphatidylcholin (in der nicht-zwitterionischen Form) als Modellkomponente ausgewählt. Als Modellkomponente für Blut wurde Wasser verwendet.

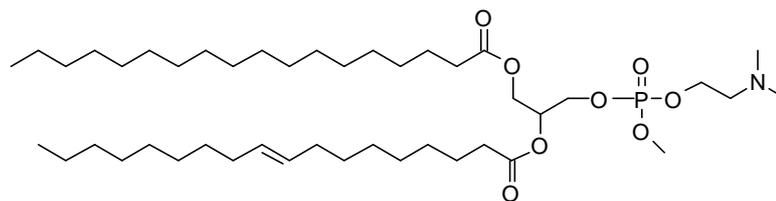


Abbildung 48: Lecithin $C_{44}H_{86}NO_8P$ (Phosphatidylcholin).

Die COSMO-RS/AM1-Rechnungen wurden mit der neuen Parametrisierung (siehe Tabelle 15 Seite 3-40) durchgeführt.

Abbildung 49 zeigt, dass sich der Blut/Hirn-Verteilungskoeffizient weder qualitativ geschweige denn quantitativ mit den gewählten Modellsubstanzen reproduzieren lässt.

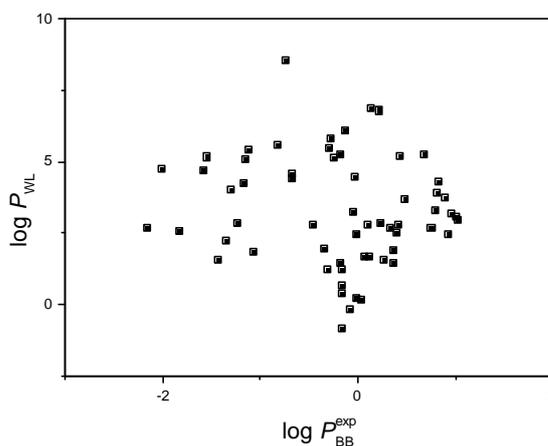


Abbildung 49: Korrelation des Wasser/Lecithin-Verteilungskoeffizienten ($\log P_{WL}$) mit dem $\log P_{BB}$.

Vor allem zwei Aspekte wurden von COSMO-RS mit den gewählten Modellsubstanzen nicht berücksichtigt. Einerseits die erwähnte zwitterionische Form der hydrophilen Kopfgruppe des Lipids und andererseits wurde vernachlässigt, dass die Membran eine *semi-liquide* Phase ist, wohingegen in COSMO-RS alle Komponenten als flüssig behandelt werden. Deshalb werden

wahrscheinlich vor allem entropische Anteile der freien Enthalpien fehlerhaft quantifiziert. Hauptgrund ist die bei der Berechnung der Misfitenergie vorgenommene Segmentpaarung *aller* Oberflächensegmente. Bei Feststoffen und semi-liquiden Phasen sind die sterischen Restriktionen allerdings größer als in der flüssigen Phase, und deshalb ist der durch die Segmentpaarungen entstehende Fehler (vgl. Kapitel 2.2.3 Abschnitt: *Entkoppelte Segmente* Seite 2-12 unten) hier wohl noch größer.

Analog der Arbeit von Keserü et al. wurde der $\log P_{\text{BB}}$ dann mit der freien Solvatationsenthalpie korreliert. Physikalisch ähnelt der Luft/Wasser-Übergang zwar weniger der Blut/Hirnschranke, aber die freie COSMO-RS-Hydratationsenthalpie beinhaltet explizit molekulare Größeninformationen (deren Korrelation mit dem $\log P_{\text{BB}}$ wurde wie eingangs erwähnt von einigen Forschungsgruppen bewiesen): In das zur Berechnung der freien Hydratationsenthalpie verwendete COSMO-RS-Gasphasenpotenzial gehen explizit die elementspezifischen Oberflächenanteile ein (siehe Gleichung 59 Seite 2-19). Dagegen ist im COSMO-RS-Bild bei Verteilungskonstanten zwischen fluiden Phasen die Größe der SAS des gelösten Moleküls nicht mehr explizit enthalten. Die Größe der SAS des gelösten Moleküls geht nur noch indirekt bei der Berechnung des Misfitenergie (durch Größe und Anzahl der gepaarten Segmente) in das Potenzial in fluider Phase ein. Dagegen wird die SAS-Größe der Lösungsmittelmoleküle explizit (im kombinatorischen Faktor vgl. Gleichung 52 Seite 2-17) im Flüssigpotenzial berücksichtigt.

Abbildung 50 auf Seite 4-27 zeigt die Korrelation von ΔG_{hydr} und $\log P_{\text{BB}}$. Mit linearer Regression erhält man die in der Abbildung gezeigte Beziehung. Nach Auslassung der fünf größten Ausreißer (4, 9, 11, 14, 23) erhält man:

$\log P_{\text{BB}} = 10.68 \cdot \Delta G_{\text{hydr}} [\text{kcal/mol}] - 6.88$. ($r^2=0.72$, $SD = 0.48$). Der Regressionskoeffizient entspricht genau dem der Beziehung von Keserü wohingegen die Standardabweichung größer ist. Abbildung 50 zeigt aber noch weitere interessante Aspekte. Eine Gruppe von Molekülen (in der Abbildung blau eingefasst) separiert sich deutlich. Hierbei handelt es sich um die Moleküle 32 bis 56.

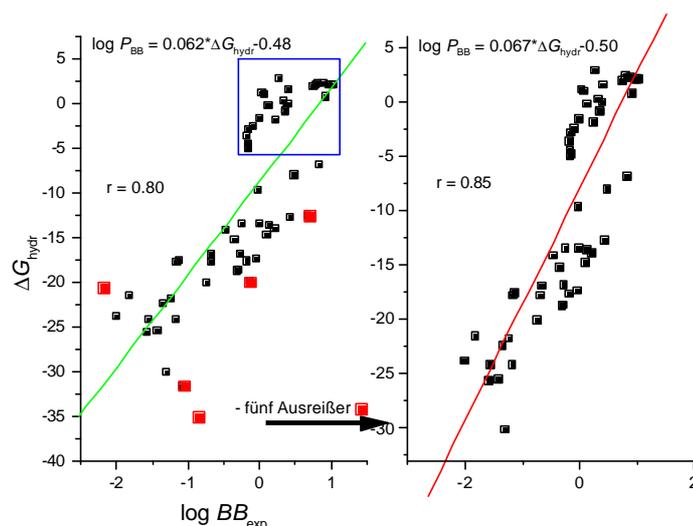


Abbildung 50: Korrelation der freien Hydratationsenthalpie [kJ/mol] mit dem Blut/Hirn-Verteilungskoeffizienten.

Diese Molekülgruppe unterscheidet sich durch zwei Eigenschaften vom Rest. Erstens ist das Molgewicht dieser Verbindungen deutlich kleiner ($MG < 200 \text{ g/mol}$) als von den übrigen Molekülen ($MG \gg 200 \text{ g/mol}$), zweitens beinhaltet die Molekülgruppe keine Stickstoffatome. Ein besseres statistisches Ergebnis kann bei separater Behandlung der beiden Gruppen erzielt werden. Der Standardfehler der Gruppe mit hohem Molgewicht sinkt dann auf $SD = 0.38$.

Mit Blick auf die Verwendbarkeit bei high-throughput-Verfahren ist es aber wünschenswert, ein Vorhersagemodell für alle Substanzen zu ermitteln, und deshalb wurden noch lineare QSPR-Beziehungen mit Hilfe der von COSMO-RS offerierten Parameter aufgestellt. Die \mathbf{s} -Momente M_i^X eines Moleküls können aus den Abschirmladungsdichten \mathbf{s} und dem \mathbf{s} -Profil (vergleiche Abbildung 5 Seite 2-13) berechnet werden:

$$M_i^X = \int p^X(\mathbf{s}) \mathbf{s}^i d\mathbf{s} \quad \text{Gleichung 123}$$

Daneben lassen sich die Wasserstoffbrückenbindung-Momente, die ein Maß für die Wasserstoffbrückendonator bzw. Akzeptoreigenschaften eines Moleküls sind, formulieren:

$$M_{hba/hbd}^X = -/+ \sum (|\mathbf{s}| - |\mathbf{s}_{hb}|) \quad \text{wenn } |\mathbf{s}| > |\mathbf{s}_{hb}| \quad \text{Gleichung 124}$$

Ein logarithmischer Verteilungskoeffizient kann bei einer bestimmten Temperatur dann als Linearkombination der \mathbf{s} -Momente und der Wasserstoffbrückenbindungsmomente (und weiterer Deskriptoren) des gelösten Moleküls aufgefasst werden [119]:

$$\log P_{BB} = c_{B,B} + c_{B,B}^{hba} M_{hba}^X + c_{B,B}^{hbd} M_{hbd}^X + \sum_{i=0}^m c_{B,B}^i M_i^X + \dots \quad \text{Gleichung 125}$$

Das σ -Moment erster Ordnung ist vom Betrag gleich der Gesamtladung des Moleküls (also Null für neutrale Moleküle) und das Moment nullter Ordnung ist gleich der Moleküloberfläche (SAS). Das zweite und dritte Moment sind Maße für die Polarität des Moleküls und der Asymmetrie des s -Profils (vgl. Abbildung 5 auf Seite 2-13). Die Einflüsse der Lösungsmittelphasen sind in den Koeffizienten $c_{B,B}$ zusammengefasst. Neben den angegebenen COSMO-RS-Momenten können natürlich weitere beliebige Deskriptoren berücksichtigt werden. Im Folgenden werden aber nur die oben aufgeführten COSMO-RS-Parameter verwendet.

Tabelle 27 zeigt eine Übersicht der verwendeten Parameter und die zugehörigen Regressionskoeffizienten und Standardabweichungen der multilinenen Regressionen. Bei Verwendung aller Parameter wird ein r^2 -Wert von 0.80 erzielt.

Tabelle 27: Statistische Ergebnisse (SD = Standardabweichung, r^2 = quadratischer Regressionskoeffizient) der multilinenen Regression des $\log P_{BB}$ mit Hilfe verschiedener Deskriptoren. „+“-Zeichen markieren in der Regression verwendete Parameter.

SAS	Vol	MG	M ₂	M ₃	M ₄	M ₅	M _{hba}	M _{hbd}	r ²	SD
+	+	+	+	+	+	+	+	+	0.80	0.39
		+	+	+	+	+	+	+	0.76	0.42
	+		+	+	+	+	+	+	0.78	0.41
+			+	+	+	+	+	+	0.79	0.39
+			+	+	+		+	+	0.79	0.39
+			+	+			+	+	0.79	0.39
+			+				+	+	0.77	0.40
+			+	+			+		0.79	0.38
+			+	+				+	0.77	0.40
+			+				+		0.73	0.43

Allerdings sind neun Variablen für einem Datensatz von 60 Molekülen sehr viel. Mit insgesamt drei Deskriptoren für die Molekülgröße (SAS, Volumen und Molgewicht) sind redundante Informationen enthalten. Es wurden Regressionen mit nur jeweils einem der drei Parameter durchgeführt. Der am besten geeignete Größendeskriptor ist die SAS ($r^2 = 0.79$, SD = 0.39). Weiter zeigte sich, dass auch die s -Momente vierter und fünfter Ordnung und der Wasserstoffbrückenbindungsdonor-Deskriptor vernachlässigt werden können, ohne dass eine deutliche Verschlechterung der Statistik auftritt. Schließlich erhält man die lineare Regressionsgleichung:

$$\log P_{BB} = 0.00789 + 0.00721 \cdot M_{hba}^X - 0.03134 \cdot SAS - 0.01807 \cdot M_2^X + 15.53901 \cdot M_3^X \quad \text{Gleichung 126}$$

mit $r^2 = 0.79$ und $SD = 0.38$.

Die Mehrheit der Substanzen ist strukturell anspruchsvoll, und dennoch kann COSMO-RS mit Hilfe der freien Hydratationsenthalpien oder mit multilinenaren Regressionsmodellen den $\log P_{BB}$ abschätzen. Anzumerken ist, dass einige der untersuchten Verbindungen Schwefelatom, deren COSMO-RS-Parameter nur abgeschätzt sind. Darüber hinaus wurde auf umfangreiche Konformationsanalysen der Moleküle verzichtet. Wegen der hohen Recheneffizienz und Vielseitigkeit der neuen COSMO-RS/AM1-Variante könnte die Arznei- und Wirkstoffforschung ein attraktives Anwendungsfeld für COSMO-RS/AM1 sein.

5 Zusammenfassung und Ausblick

Kernstück der COSMO-RS-Theorie ist die statistisch-physikalische Auswertung quantenchemisch berechneter Oberflächenabschirmungsladungsdichten von Molekülen in einem idealen Leiter. Zur Berechnung der Oberflächenladungsdichten wird das quantenchemische Kontinuummodell COSMO verwendet. In dieser Arbeit wurde nun erstmalig eine Parametrisierung der COSMO-RS-Methode für den semiempirischen AM1-Hamiltonoperator vorgestellt. Zur Anpassung der COSMO- und COSMO-RS-Parameter wurde ein Moleküldatensatz mit experimentell bekannten physikochemischen Stoffeigenschaften (Henry-Konstante, Dampfdruck und Oktanol/Wasser-Verteilungskoeffizient) zusammengestellt. Dieser wurde in einen 749 Substanzen umfassenden Trainingsdatensatz und einen 203 Verbindungen enthaltenden Vorhersagedatensatz aufgeteilt. Mit Hilfe einer Optimierungsroutine wurden die adjustierbaren COSMO-RS-Parameter so angepasst, dass die experimentellen Größen des Trainingsdatensatzes möglichst genau reproduziert werden. In zwei Teilschritten wurde zuerst eine Parametrisierung für CHO-Verbindungen ermittelt, und später wurden die Datensätze um Stickstoff-, Fluor-, und Chlorkomponenten erweitert.

Zur Verbesserung der statistischen Qualität der COSMO-RS/AM1-Vorhersagen wurden dann zwei alternative Ansätze verfolgt. Einerseits wurde die COSMO-RS-Theorie modifiziert und erweitert, und andererseits wurden die zugrundeliegenden semiempirischen COSMO/AM1-Rechnungen verändert.

Ergebnisse

Die Optimierung für den CHO-Teildatensatz offenbarte eine systematische Abweichung von Verbindungen mit Hydroxygruppen. Da die Abweichung hauptsächlich bei der Henry-Konstante und dem Dampfdruck zu beobachten war, wurde zuerst ein linearer Korrekturterm in das Gasphasenpotenzial eingeführt. Eine elegantere Verbesserung wurde jedoch durch die nachträgliche Skalierung der Oberflächenladungsdichten der Hydroxysauerstoffe im Wasserstoffbrückenbindungsenergieterm erreicht. Diese Modifikation in Verbindung mit einem signifikant vergrößerten Sauerstoffradius bei den COSMO-Rechnungen konnte die systematische Abweichung ebenfalls korrigieren.

Bei der Optimierung mit dem um N-, Cl- und F-Verbindungen erweiterten Datensatz wurde dann versucht, *mit lokalen Grenzorbitaldeskriptoren* die Wasserstoffbrückenbindungsenergie um eine nicht-elektrostatische, eher der Polarisierbarkeit verwandte Komponente zu

erweitern. Es wurde gezeigt, dass Atome im Molekül sich durch die Wertebereiche der LFODs auf natürliche Weise ihren verschiedenen funktionellen Gruppen zuordnen lassen. Trotzdem konnten die LFODs nicht zu einer verbesserten Beschreibung der COSMO-RS-Wasserstoffbrückenenergie und damit zu einer genaueren Vorhersage der Stoffeigenschaften beitragen.

Neben den Vorhersagen für die Hydroxyverbindungen erwiesen sich auch die Vorhersagen der Stoffeigenschaften für Substanzen mit Stickstoff-Sauerstoff-Bindungen als stark fehlerbehaftet. Dipolmomentrechnungen zeigten, dass der Grund hierfür die ungenaue Beschreibung der Ladungsverteilung ist. Die mit AM1 berechneten Dipolmomente der Nitro- und Nitratverbindungen weichen sowohl stark von experimentellen Gasphasendipolmomenten als auch von DFT- und MP2-Dipolmomenten ab.

Zur Korrektur wurden bei den COSMO/AM1-Rechnungen Ausgleichsladungen auf die Atomzentren der Sauerstoffe und Stickstoffe der Nitrogruppe aufgebracht. Mit Hilfe dieser Ausgleichsladungen konnten einerseits experimentelle Dipolmomente besser reproduziert werden, andererseits verbesserte sich auch die Vorhersage der Stoffeigenschaften. Nach der erfolgreichen Anwendung für Nitro- und Nitratverbindungen wurde das Konzept der Core-Ausgleichsladungen ebenfalls für Hydroxygruppen und für nicht an Heteroatome gebundene sp^3 -hybridisierte Kohlenstoffatome angewendet.

Statistische Validierung von COSMO-RS/AM1

Zur Bewertung der vorgestellten COSMO-RS/AM1-Parametrisierung wurden umfangreiche Vergleiche mit alternativen Rechenmethoden für den Trainings- und den Testdatensatz durchgeführt. Sowohl weitere quantenchemische (semiempirische und auch ab initio-) Kontinuumssolvatationsmodelle als auch Gruppenbeitragsmethoden fanden hierbei Anwendung.

Für die Henry-Konstante konnten sehr gute statistische Ergebnisse erzielt werden. COSMO-RS/AM1 mit der neuen Parametrisierung ist *allen* anderen Rechenmodellen überlegen. Vor allem die übrigen quantenchemischen Kontinuumsansätze können die Qualität von COSMO-RS/AM1 nicht erreichen. Lediglich die DFT-Variante COSMO-RS/BP liefert ebenfalls gute statistische Vorhersagen. Bei den Vergleichsrechnungen zeigte sich aber auch, dass die COSMO-RS/B3LYP-Parametrisierung mit SVP-Basissätzen erhebliche Defizite aufweist. Vor allem für Verbindungen, die Stickstoff- und Sauerstoffatome enthalten, sind höchstens qualitative Vorhersagen möglich.

Ähnlich gute Ergebnisse können für Dampfdruckvorhersagen erzielt werden. Ein Vergleich mit anderen Kontinuumsmodellen ist hier allerdings nicht möglich. COSMO-RS ist bislang das einzige quantenchemische Modell, das eine Berechnung des Dampfdrucks aus der Molekülstruktur gestattet. Wie schon bei den Vorhersagen der Gaslöslichkeit, ist auch hier die statistische Qualität der COSMO-RS/AM1-Parametrisierung der DFT-Parametrisierung ebenbürtig. Defizite zeigen sich allerdings gerade bei einfachen Stoffgruppen. Dies wird teilweise auch in den vorgestellten Anwendungsbeispielen deutlich. Hier sollte die überlegene DFT-Parametrisierung verwendet werden. Im Hinblick auf die zukünftige Verwendung der AM1-Methode für komplexe Arzneistoffe oder Biomoleküle scheint dies allerdings wenig nachteilig zu sein.

Obwohl die berechneten Oktanol/Wasser-Verteilungskoeffizienten kleinere statistische Fehler als die Henry-Konstanten aufweisen, ist COSMO-RS/AM1 der etablierten, auf mehreren Hundert optimierten Parametern basierenden CLOGP-Methode unterlegen. Deren Vorhersagequalität kann wohl nur mit großem Aufwand und vielen stoffgruppenspezifischen Korrekturen erreicht werden. Durch eine exklusive Anpassung an den Oktanol/Wasser-Verteilungskoeffizient und die Integration eines molekülgrößenabhängigen Dispersionsterms konnte ad hoc eine signifikante Verbesserung erzielt werden, doch die hohe Genauigkeit der CLOGP-Methode kann auch so nicht erreicht werden. Im Falle der K_{OW} -Vorhersagen liegen die Vorteile der COSMO-RS-Methode im derzeitigen Entwicklungsstadium ausschließlich in der prinzipiellen universellen Anwendbarkeit und der Möglichkeit der temperaturabhängigen Berechnung. Darüber hinaus können detaillierte vergleichende Untersuchungen (z.B. zwischen Isomeren) oder energetische Untersuchungen durchgeführt werden.

Anwendungen

In Kapitel 4 wurden einige Anwendungsbeispiele aufgezeigt, die einige der typischen Vorteile des COSMO-RS-Verfahrens veranschaulichen:

1. COSMO-RS ist derzeit das einzige Verfahren, mit dem Henry-Konstanten temperaturabhängig direkt aus der Molekülstruktur ohne zusätzliche Informationen berechnet werden können. Dadurch ist es möglich, aus dem Temperaturverlauf der freien Hydratationsenthalpie die Lösungsenthalpie und die Lösungsentropie mittels linearer Regression zu bestimmen. Für die untersuchte Stoffgruppe aromatischer Kohlenwasserstoffe gelingt dies mit hoher Genauigkeit.
2. Durch Betrachtung homologer Reihen können experimentell ermittelte Henry-Konstanten bewertet werden.

3. Eine Aufspaltung der freien Hydratationsenthalpie in elektrostatische und nicht-elektrostatische Anteile zeigt, dass die elektrostatische Stabilisierung der Solvation bei Vergrößerung des aromatischen Systems zunimmt und die nicht-elektrostatische Destabilisierung innerhalb der untersuchten homologen Reihe abnimmt.
4. Qualitativ können kleine Differenzen der Stoffeigenschaften von Strukturisomeren aufgelöst werden.
5. Bei der zweiten vergleichenden temperaturabhängigen Untersuchung zeigt COSMO-RS/DFT die kleinsten statistischen Fehler aller verwendeten Methoden. Allerdings ist anzumerken, dass COSMO/RS-DFT auch die höchsten Rechenzeitanforderungen stellt.
6. Exzessenergien können für Mischungsdiagramme bestimmt werden. Hierbei zeigt COSMO-RS keine Schwierigkeiten bei der Beschreibung von Verbindungen in unendlicher Verdünnung.
7. Mit COSMO-RS kann der Ölsäuremethylester/Wasser-Verteilungskoeffizient berechnet werden, obwohl die Methode nicht für diese Größe parametrisiert ist. Der $K_{\text{ÖW}}$ korreliert gut mit experimentellen Fischöl/Wasser-Verteilungskoeffizienten. Für die untersuchten Verbindungen gelingt dies mit COSMO-RS/BP besser als mit COSMO-RS/AM1.
8. COSMO-RS/AM1 mit der neuen Parametrisierung kann zur Abschätzung des Blut/Hirn-Verteilungskoeffizienten verwendet werden. Neben einer befriedigenden Korrelation der freien Hydratationsenthalpie mit dem $\log P_{\text{BB}}$ können auch mit linearen Regressionsmodellen, welche sich verschiedener COSMO-RS-Deskriptoren bedienen, gute Übereinstimmungen mit experimentellen Größen erzielt werden.

Fazit

Die Beispiele zeigen den Hauptvorteil von COSMO-RS: *die breite Anwendbarkeit*. Wohingegen mit den meisten alternativen Rechenmodellen nur jeweils *eine* physikalische Größe berechnet werden kann, ist COSMO-RS ein sehr robustes, streng physikalisches Modell zur Berechnung von *verschiedenen* physikochemischen Stoffeigenschaften. Mit COSMO-RS können prinzipiell alle Arten von Stoffen und Mischungen studiert werden. Darüber hinaus gewähren COSMO-RS-Analysen modelltheoretische energetische Einsichten in Lösungs- und Verteilungsvorgänge.

Die COSMO-RS/AM1-Parametrisierungen offenbarten phänomenologisch mehrere Defizite der AM1-Methode durch fehlerhafte Vorhersagen der physikochemischen Stoffeigenschaften. Beispielsweise überschätzen AM1-Rechnungen die Dipolmomente polarer Verbindungen mit Stickstoff-Sauerstoffbindungen, und andererseits werden auch stabilisierende Wassertoffbrück-

kenbindungsenergieanteile der freien Hydratationsenthalpien von Hydroxyverbindungen relativ zu anderen wasserstoffbrückenbindungsaktiven Substanzen systematisch unterschätzt. Obwohl die physikalischen Ursachen verschiedener Natur sind, konnten *beide* Defizite mit *einer* sehr einfachen Methodik (der Einführung von zusätzlichen Ladungen auf den Atomzentren) ausgeglichen werden.

Ausblick

Die nächsten Schritte zur weiteren Verbesserung der COSMO-RS/AM1-Methode liegen auf der Hand: Zum einen sollte mit weiteren Anwendungen geprüft werden, inwieweit die vorgestellte Parametrisierung auch zur Vorhersage anderer Stoffeigenschaften geeignet ist und welche weiteren Modifikationen eventuell zur Verbesserung vorgenommen werden können. Denkbar ist z. B. die Berücksichtigung von Hexan/Wasser-Verteilungskoeffizienten und von Wasserlöslichkeitsdaten bei der Optimierung.

Da die hier vorgestellte Parametrisierung mit experimentellen Daten bei 25°C ermittelt wurden, kann die Methode mit Hinblick auf die Bestimmung temperaturabhängiger Größen noch verbessert werden.

In jedem Fall bietet die hier vorgestellte Parametrisierung die Möglichkeit einer Anwendung der COSMO-RS-Theorie auf semiempirischem Niveau und öffnet damit den Zugang zur direkten Berechnung thermodynamischer Stoffdaten und ihrer Temperaturabhängigkeit aus der Molekülstruktur auch für größere Verbindungen.

Die AM1-Parametrisierung kann ein Ausgangspunkt für die Anpassung der COSMO-RS-Parameter an andere semiempirische Verfahren (z.B. PM3) sein. Vergleiche zwischen semiempirischen Parametrisierungen könnten Defizite und Stärken der verschiedenen Methoden aufdecken.

Wegen der im Vergleich zur DFT-Variante geringen Rechenzeitanforderung ist der Bereich der Lebenswissenschaften ein attraktives Anwendungsfeld für COSMO-RS/AM1. Um jedoch auch anspruchsvolle Moleküle wie Membrankomponenten, Proteine oder DNS-Fragmente studieren zu können, muss die Parametrisierung auf weitere Elemente (Phosphor, Schwefel, Brom und Iod) ausgedehnt, und die Anwendbarkeit für geladene Moleküle geprüft werden.

6 Literaturverzeichnis

- [1] R. P. Schwarzenbach, P. M. Gschwend, D. M. Imboden: Environmental organic chemistry, John Wiley, New York **1992**.
- [2] Jahresbericht **1999** Umweltbundesamt.
- [3] Jahresbericht **2000** Verband der Chemischen Industrie e.V. Karlsstr. 21, 60329 Frankfurt am Main.
- [4] D. Mackay, W. Y. Shiu: A critical review of Henry's law constants for chemicals of environmental interest *J. Phys. Chem. Rev. Data* **1981** *10*, 1175-1199.
- [5] L. P. Burkhard, D. E. Armstrong, A. W. Andren: Henry's law constants for polychlorinated biphenyls *Environ. Sci. Technol.* **1985** *19*, 590-596.
- [6] J. Staudinger, P. V. Roberts: A critical review of Henry's law constants for environmental applications *Crit. Rev. Environ. Sci. Technol.* **1996** *26*, 205-297.
- [7] J. Hine, P. K. Mookerjee: The intrinsic hydrophilic character of organic compounds. Correlations in terms of structural contribution. *J. Org. Chem.* **1975**, *40*:292-298.
- [8] D. Mackay: Multimedia environmental models, the fugacity approach, Lewis Publishers Inc., Boca Raton **1991**.
- [9] A. Leo, C. Hansch, D. Elkins: Partition coefficients and their uses *Chem. Rev.* **1971**, *71*, 524-616.
- [10] C. Hansch, A. Leo: Exploring QSAR: Fundamentals and applications in chemistry and biology; American Chemical Society: Washington, DC, **1995**.
- [11] A. Leo: ClogP; Daylight Chemical Information Systems: Irvine, CA, **1991**.
- [12] G. Klopman, J-Y. Li, S. Wang, M. Diamayuga: Computer automated logP calculations based on an extended group contribution approach *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 752.
- [13] R. Kühne, R-U. Ebert, G. Schüürmann: Estimation of vapour pressures for hydrocarbons and halogenated hydrocarbons from chemical structure by a neural network *Chemosphere* **1996**, *34*, 671-686.
- [14] B. Beck, A. Breindl, T. Clark: QM/NN QSPR models with error estimation: vapor pressure and logP *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1046-1051.
- [15] C. J. Cramer, D. G. Truhlar: Continuum solvation models: classical and quantum mechanical implementations *Reviews in Computational Chemistry*, Volume VI, Seite 1-72, Editoren: K. B. Lipkowitz, D. B. Boyd VCH Publishers, Inc. New York, **1995**.

- [16] G. Schüürmann, M. Cossi, V. Barone, J. Tomasi: Prediction of the pK_a of carboxylic acids using the ab initio continuum-solvation model PCM-UAHF *J. Phys. Chem. A*. **1998**, *102*, 6706-6712.
- [17] G. Schüürmann: Gas-phase and solution-phase proton transfer to H₂O analyzed by high-level ab initio quantum chemistry including complete basis set and Gaussian theory schemes *Chem. Phys. Lett.* **1999**, *302*, 471-479.
- [18] G. Schüürmann: Prediction of Henry's law constant of benzene derivatives using quantum chemical continuum solvation models *J. Comp. Chem.* **2000**, *21*, 17-34.
- [19] P. H. Howard, W. M. Meylan: Prediction of physical properties, transport, and degeneration for environmental fate and exposure assessment *Quant. Struc.-Act. Rel. In Env. Sci.*, VII, 185-204.
- [20] W. J. Leyman, W. F. Reehl, D. H. Rosenblatt: *Handbook of Chemical property estimation Methods*, ACS, Washington **1990**, 1.10-1.38.
- [21] A. Klamt, G. Schüürmann: COSMO: a new approach to dielectric screening in solvents with explicit expression for the screening energy and its gradient *J. Chem. Soc. Perkin Trans. 2* **1993**, 799-805.
- [22] J. Tomasi, M. Persico: Molecular interactions in solution: an overview of methods based on continuous distribution of the solvent *Chem. Rev.* **1994**, *94*, 2027-2094.
- [23] C. J. Cramer, D. G. Truhlar: Development and biological applications of quantum mechanical continuum solvation models *Quantitative Treatments of Solute/Solvent Interactions* Theoretical and Computational Chemistry Vol.1 (9-54) Editoren: P. Politzer, J. S. Murray, Elsevier Science B.V. **1994**.
- [24] C. J. Cramer, D. G. Truhlar: Implicit solvation models: equilibria, structure, spectra, and dynamics *Chem. Rev.* **1999**, *99*, 2161-2200.
- [25] M. Orozco, F. J. Luque: Theoretical methods for the description of the solvent effect in biomolecular Systems *Chem. Rev.* **2000**, *100*, 4187-4225.
- [26] Abbildung von J. Labanowski: <http://server.ccl.net/cca/documents/molecular-modelling/>
- [27] A. Klamt, V. Jonas: Treatment of the outlying charge in continuum solvation models *J. Chem. Phys.* **1996**, *105*, 9972-9981.
- [28] B. Mennucci, J. Tomasi: Continuum solvation models: a new approach to the problem of solute's charge distribution and cavity boundaries *J. Chem. Phys.* **1997**, *106*, 5151-5158.
- [29] D. M. Chipman: Charge penetration in dielectric models of solvation *J. Chem. Phys.* **1997**, *106*, 10194-10206.
- [30] J. D. Jackson: *Classical Electrodynamics* (Seite 56 ff.), Wiley, New York **1975**.
- [31] A. Schäfer, A. Klamt, D. Sattel, J. C. W. Lohrenz, F. Eckert: COSMO Implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187-2193.
- [32] A. Klamt: Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena *J. Phys. Chem.* **1995**, *99*, 2224-2235.

- [33] A. Klamt, V. Jonas, T. Bürger, J. C. W. Lohrenz: Refinement and parameterization of COSMO-RS *J. Phys. Chem. A* **1998** 102, 5074-5085.
- [34] A. Klamt, F. Eckert: COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids *Fluid Phase. Equi.* **2000**, 172, 43-72.
- [35] W. J. Leyman, W. F. Reehl, D. H. Rosenblatt: *Handbook of Chemical property estimation Methods*, ACS, Washington **1990**, 1.47-1.49.
- [36] MOPAC 98 (non commercial test version) Fujitsu Ltd., 9-3, Nakase 1-Chome, Mihama-ku, Chiba-city, Chiba 261, Japan; Stewart Computational Chemistry, 15210 Paddington Circle, Colorado Springs, CO 80921, 1998.
- [37] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart: AM1: A new general purpose quantum mechanical molecular model *J. Am. Chem. Soc.* **1985**, 107, 3902-3909.
- [38] J. J. P. Stewart: Semiempirical molecular orbital methods *Reviews in Computational Chemistry* (Seite 45-81) Editoren: K. B. Lipkowitz und D. B. Boyd, VCH Weinheim **1990**.
- [39] J. J. P. Stewart: Optimization of parameters for semiempirical methods. I. method *J. Comput. Chem.* **1989**, 10, 209-220.
- [40] J.J. P. Stewart: MOPAC 93 Manual (Revision Number 2) Fujitsu Limited **1994** (Seite 222)
- [41] J. J. Dannenberg: Hydrogen bonds: a comparison of semiempirical and ab initio treatments *J. Molec. Struct.* **1997**, 401, 279-286.
- [42] C. Breitkopf, R. Kühne, G. Schüürmann: Dependence of multimedia level III partitioning and residence times of compounds on physicochemical properties and system parameters of water-rich and water-poor environments *Environ. Toxicol. Chem.* **2000**, 19, 1430-1440
- [43] D. A. Leo: MedChem99 database. Daylight Chemical Information Systems. Irvine (CA,USA) **1999**.
- [44] A. Bondi: van der Waals volumes and radii *J. Phy. Chem.* **1964**, 68, 441-451.
- [45] S. S. Batsanov: Van der Waals radii of elements from the data of structural inorganic chemistry *Russ. Chem. Bl.* **1995**, 44, 18-23.
- [46] S. C. Nyburg, C. H. Faerman: A revision of van der Waals atomic radii for molecular crystals: N, O, F, S, Cl, Se, Br, and I bonded to carbon *Acta Crystallogr. Sect. B* **1985** 41, 274-279.
- [47] R. Chauvin: Explicit periodic trend of van der Waals radii *J. Phys. Chem.* **1992** 96 9194-9197.
- [48] P. Buchwald: Modeling liquid properties, solvation, and hydrophobicity: A molecular-size-based perspective *Perspectives in Drug Discovery and Design* **2000**, 19, 19-45.
- [49] J. Tolls, J. van Dijk, E. J. M. Verbruggen, J. L. M. Hermens, B. Loeprecht, G. Schüürmann: Aqueous solubility – molecular size relationships – a case study using C₁₀ to C₁₉-alkanes *submitted to Environmental Science and Technology*.
- [50] D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman in: Chemometrics: a textbook. Elsevier, Amsterdam-Oxford-New York-Tokio **1988** (271-291).

- [51] A. Klamt: Estimation of gas-phase hydroxyl radical rate constants of organic compounds from molecular orbital calculations *Chemosphere* **1993**, *26*, 1273-1289.
- [52] A. Klamt: Estimation of gas-phase hydroxyl radical rate constants of oxygenated compounds based on molecular orbital calculations *Chemosphere* **1995**, *32*, 717-726.
- [53] K. Fukui, T. Yonezawa, C. Nagata: Theory of substitution in conjugated molecules *Bull. Chem. Soc. Jap.* **1954**, *27*, 423-427.
- [54] K. Fukui, H. Kato, T. Yonezawa: A new quantum-mechanical reactivity index for saturated compounds *Bull. Chem. Soc. Jap.* **1961**, *34*, 1111-1115.
- [55] W. Kohn, L. J. Sham: Self-consistent equations including exchange and correlation effects *Phys. Rev. A* **1965**, *140*, 1133-1138.
- [56] P. Hohenberg, W. Sham: Inhomogeneous electron gas *Phys. Rev. B.* **1964**, *136*, 864-871.
- [57] B. J. Delley: An all-electron numerical method for solving the local density functional for polyatomic molecules *J. Chem. Phys.* **1990**, *92*, 508-517. DMol³ is available from MSI in the Cerius² program-package.
- [58] Cerius2, version 3.8; Molecular Simulations Inc.: San Diego, CA, **1998**.
- [59] A. D. Becke: Density-functional exchange-energy approximation with correct asymptotic behavior *Phys. Rev. A.* **1988**, *38*, 3098-3100.
- [60] J. P. Perdew: Density-functional approximation for correlation energy of the inhomogeneous electron gas *Phys. Rev. B* **1986**, *33*, 8822-8824.
- [61] S. H. Vosko, L. Wilk, M. Nusair: Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis **1980** *Can. J. Phys.* *58*: 1200.
- [62] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel: Electronic structure calculations: the program system TURBOMOLE *Chem. Phys. Lett.* **1989**, *162*, 165-169.
- [63] A. Schäfer, C. Huber, R. Ahlrichs: Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829-5835.
- [64] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs: Auxiliary basis sets to approximate coulomb potentials *Chem. Phys. Lett.* **1995**, *240*, 283-290.
- [65] K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs: Auxiliary basis sets for main row atoms and transition metals and their use to approximate coulomb potentials *Theor. Chem. Acc.* **1997**, *97*, 119-124.
- [66] A. D. Becke: Density-functional thermochemistry. III. The role of exact exchange *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- [67] A. Schäfer, H. Horn, R. Ahlrichs: Fully optimized contracted Gaussian basis sets for Atoms Li to Kr *J. Chem. Phys.* **1992**, *97*, 2571-2577.
- [68] C. Curutchet, M. Orozco, F. J. Luque: Solvation in octanole: parameterization of the continuum MST model *J. Comp. Chem.* **2001**, *22*, 1180-1193.

- [69] S. Miertus, E. Scrocco, J. Tomasi: Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects *Chem. Phys.* **1981**, 55, 117.
- [70] S. Miertus, J. Tomasi: Approximate evaluation of the electrostatic free energy and internal energy changes in solution processes *Chem. Phys.* **1982**, 65, 239.
- [71] M. Cossi, V. Barone, R. Cammi, J. Tomasi: Ab initio study of solvated molecules: a new implementation of the polarizable continuum model *Chem. Phys. Lett.* **1996**, 255, 327-335.
- [72] V. Barone, M. Cossi, J. Tomasi: A new definition of cavities for the computation of solvation free energies by the polarizable continuum model *J. Phys. Chem.* **1997**, 107, 3210-3221.
- [73] GAUSSIAN 98, Revision A.7, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, **1998**.
- [74] C. J. Cramer, D. G. Trular: An SCF solvation model for the hydrophobic effect and absolute free energies of solvation *Science* **1992**, 256, 213-217.
- [75] C. J. Cramer, D. G. Trular: A general parameterization for including aqueous solvation effects in the PM3 molecular orbital model *J. Comput. Chem.* **1992**, 13, 1089-1097.
- [76] AMSOL 4.0, QCPE 606, Bloomington, IN 47405, **1993** (Autoren siehe Referenz [81]).
- [77] C. C. Chambers, G. D. Hawkins, C. J. Cramer, D. G. Trular: A model for aqueous solvation based on class IV atomic charges and first-solvation-shell effects *J. Phys. Chem.* **1996**, 100, 16385-16398.
- [78] J. W. Storer, D. J. Giesen, C. J. Cramer, D. G. Trular: Class IV charge models: a new semiempirical approach in quantum chemistry *J. Comp. Aid. Mol. Design* **1995** 9, 87-110.
- [79] J. Li, T. Zhu, C. J. Cramer, D. G. Trular: A new class IV charge model for extracting accurate partial charges from wave functions *J. Phys. Chem. A* **1998** 102, 1820-1831.
- [80] G. D. Hawkins, C. J. Cramer, D. G. Truhlar: Parameterized model for aqueous free energies of solvation using geometry-dependent atomic surface tensions with implicit electrostatics *J. Phys. Chem. B* **1997**, 101, 7147-7157.
- [81] AMSOL 6.5.3. by G. D. Hawkins, D. J. Giesen, G. C. Lynch, C. C. Chambers, I. Rossi, J. W. Storer, J. Li, D. Rinaldi, D. A. Liotard, C. J. Cramer and D. G. Trular, University of Minnesota, Minneapolis, **1999**, based in part on AMPAC-version 2.1 by D. A. Liotard, E. F. Healy, J. M. Ruiz and M. J. S. Dewar.
- [82] D. Tiegs, J. Gmehling, P. Rasmussen, A. Fredenslund: Vapor-liquid equilibria by group contribution. 4. Revision and extension *Ind. Chem. Eng. Res.* **1987**, 26, 159-161.

- [83] W. J. Leyman, W. F. Reehl, D. H. Rosenblatt: *Handbook of Chemical property estimation Methods, ACS, Washington* **1990**, Kapitel 14.
- [84] D. S. Mishra, S. H. Yalkowski: Estimation of vapor pressure of some organic compounds *Ind. Chem. Res.* **1991**, *30*, 1609-1612.
- [85] S. Banerjee, P. H. Howard: Improved estimation of solubility and partitioning through correction of UNIFAC-derived activity coefficients *Environ. Sci. Technol.* **1988**, *22*, 839-841.
- [86] W. B. Arbuckle: Estimating activity coefficients for use in calculating environmental parameters. *Environ. Sci. Technol.* **1983**, *17*, 537-542.
- [87] D. Mackay: Comment on "Partition coefficient and bioaccumulation of selected organic chemicals" *Environ. Sci. Technol.* **1977**, *11*, 1219.
- [88] G. Nocon, U. Weidlich, J. Gmehling, J. Menke, U. Onken: Prediction of gas phase solubilities by a modified UNIFAC equation *Fluid Phase Equilibria* **1983**, *13*, 381-392.
- [89] G. Wienke, J. Gmehling: Prediction of octanol-water partition coefficients, Henry coefficients and water solubilities using UNIFAC *Toxicol. Environ. Chem.* **1998**, *65*, 57-86.
- [90] H. A. Bamford, D. L. Poster, J. E. Baker: Temperature dependence of Henry's law constant of thirteen polycyclic aromatic hydrocarbons between 4°C and 31°C. *Environ. Toxicol. Chem* **1999**, *18*, 1905-1912.
- [91] W. M. Meylan, P. H. Howard: Bond contribution method for estimating Henry's law constants. *Environ. Toxicol. Chem.* **1991** *10*, 1283-1293.
- [92] R. Kühne, R-U. Ebert, F. Kleint, G. Schmidt, G. Schüürmann: Group contribution methods to estimate water solubilities of organic chemicals. *Chemosphere*, **1995**, *30*, 2061-2077.
- [93] G. Schüürmann, R. Kühne, F. Kleint, R-U. Ebert, C. Rothenbacher, P. Herth A software system for automatic chemical property estimation from molecular structure. in *QSAR in Environmental Sciences VII* (Hrsg.: F. Chen, G. Schüürmann), SETAC Press, Pensacola, FL, USA, **1997**, 93-114.
- [94] B. Marten, K. Kim, C. Cortis, R. A. Friesner, R. B. Murphy, M. N. Ringnalda, D. Sitkoff, B. Honig: New model for calculation of solvation free energies: correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects *J. Phys. Chem.*, **1996** *100*, 11775-11788.
- [95] W. J. Sonnenfeld, W. H. Zoller, W. E. May: Dynamic coupled-column liquid chromatographic determination of ambient temperature vapour pressures of polynuclear aromatic hydrocarbons. *Anal. Chem.* **1983**, *55*, 275-280.
- [96] W. E. May, S. P. Wasik, D. H. Freeman: Determination of the solubility behavior of some polycyclic aromatic hydrocarbons in water. *Anal. Chem.* **1978**, *50*, 997-1000.
- [97] M. Alaei, R. M. Whittall, W. M. J. Strachan: The effect of water temperature and composition on Henry's law constants for various PAHs. *Chmosphere*, **1996**, *32*, 1153-1164.
- [98] R. L. Bohon, W. F. Claussen *J. Am. Chem. Soc.* **1951**, *72*, 1571.
- [99] Owens, J. W.; Wasik, S. P., Devoe, H. Aqueous solubilities and enthalpies of solution of n-alkylbenzenes *J. Chem. Eng. Data* **1986**, *31*, 47.

- [100] SYBYL Molecular Modelling Software 6.5; Tripos Associates Inc.: St. Lois, MO, **1999**.
- [101] MOPAC 93 Fujitsu Ltd. , 9-3, Nakase 1-Chome, Mihama-ku, Chiba-city, Chiba 261, Japan; Stewart Computational Chemistry, 15210 Paddington Circle, Colorado Springs, CO 80921, **1998**.
- [102] COSMO-RS 98.31, A. Klamt, COSMOlogic GmbH & Co. KG, Burscheider Str. 515, 51381 Leverkusen, Germany.
- [103] B. I. Lee, M. G. Kesler: A generalized thermodynamic correlation based on three-parameter corresponding states. *AIChE Journal* **1975**, *21*, 510-527.
- [104] W. J. Leyman, W. F. Reehl, D. H. Rosenblatt: *Handbook of Chemical property estimation Methods*, ACS, Washington 1990, 14.4.
- [105] A. Klamt, F. Eckert, COSMOtherm, version C1.1, Release 05.00; COSMOlogic GmbH & Co. KG, Leverkusen, Germany, **2000**.
- [106] TURBOMOLE Version 5.2, (1999), Electronic Structure Calculations on Workstation Computers: The Program System TURBOMOLE. R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel; *Chem. Phys. Letters* **162**, (1998) 165-169.
- [107] Iven Clausen: A priory Berechnung thermodynamischer Daten, Dissertationsschrift eingereicht am Fachbereich 6 Verfahrenstechnik der Technischen Universität Berlin.
- [108] A. K. Ghose, V. N. Viswanadhan, J.J. Wendoloski: Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: An analysis of ALOGP and CLOGP methods. *J. Phys. Chem. A* **1998**, *102*, 3762-3772.
- [109] C. Hansch, J. P. Bjorkroth, A. Leo: Hydrophobicity and central nervous system agents: on the principle of minimal hydrophobicity in drug design *J. Pharm. Sci.* **1987**, *76*, 663-687.
- [110] J. B. M. M. van Bree, A. G. De Boer, M. Danhof, L. A. Ginsel, D. D. Breimer: Characterization of an "in vitro" blood-brain barrier: effects of molecular size and lipophilicity on cerebrovascular endothelial transport rates of drugs. *J. Pharamcol. Exp. Ther.* **1988**, *247*, 1233-1239.
- [111] R. Kaliszan, M. Markuszewski: Blood/brain distribution described by a combination of partition coefficients and molecular mass *Int. J. Pharm.* **1996**, *45*, 9-16.
- [112] M. H. Abraham, H. S. Chadha R. C. Mitchel: Hydrogen bonding 33. Factors that influence the distribution of solutes between blood and brain *J. Pharm. Sci.* **1994**, *83*, 1257-1268.
- [113] M. H. Abraham, H. S. Chadha R. C. Mitchel: Hydrogen bonding 36. Determination of blood brain distribution using octanol-water partition coefficients *Drug. Des. Discuss.* **1995**, *13*, 123-131.
- [114] U. Norinder, P. Sjoberg, T. Ostenberg: Theoretical calculation and prediction of brain-blood partitioning of organic solutes using Molsurf parametrization and PLS statistics *J. Pharm. Sci.* **1998**, *87*, 952-959.
- [115] J. M. Luco: Prediction of the brain-blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares (PLS) modelling *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 396-404.

-
- [116] M. Kansy, H. Waterbemd: Hydrogen bonding capacity and brain penetration *Chimia* **1992**, *46*, 299-303.
- [117] J. Kelder, P. D. J. Grootenhuis, D. M. Bayada, L. P. C. Delbressine, J. P. Ploemen: Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharm. Res.* **1999**, *16*, 1514-1519.
- [118] G. M. Keserü, L. Molnár: High-throughput prediction of blood-brain partitioning: a thermodynamic approach *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 120-128.
- [119] A. Klamt, F. Eckert, M. Hornig: COSMO-RS: A novel view to physiological solvation and partition questions *J. Comp. Aid. Molec. Des.* **2001**, *15*, 355-365.

A Anhang

A.1 Moleküldatensätze

Tabelle A-1 zeigt den Trainingsdatensatz mit experimentellen Stoffdaten. $\log H'$ -Werte (Henrykonstanten) mit zwei Literaturangaben sind aus (experimentellen) Dampfdrücken (P_V [Pa]) und Wasserlöslichkeiten berechnet. K_{OW} = Oktanol/Wasser-Verteilungskoeffizient. Die Literaturangaben (Ref) der experimentellen Größen sind in Kapitel A.2 Seite A-21 ff. gelistet.

Nr	CAS-NR	Name	$\log H'$	Ref	$\log P_V$	Ref	$\log K_{OW}$	Ref
T 1	000074-98-6	PROPANE	1.46	L66	5.98	L14	2.36	L14
T 2	000106-97-8	N-BUTANE	1.58	L66	5.39	L14	2.89	L14
T 3	000109-66-0	N-PENTANE	1.69	L56	4.84	L3	3.39	L14
T 4	000110-54-3	N-HEXANE	1.82	L47	4.30	L3	3.90	L14
T 5	000142-82-5	N-HEPTANE	1.91	L56	3.79	L14	4.66	L14
T 6	000111-65-9	N-OCTANE	2.10	L56	3.26	L3	5.18	L14
T 7	000111-84-2	N-NONANE	2.26	L14 / L16	2.77	L14		
T 8	000124-18-5	N-DECANE	2.24	L56	2.28	L14		
T 9	000112-40-3	DODECANE	2.54	L14 / L16	1.26	L14	6.10	L14
T 10	000629-59-4	TETRADECANE	2.76	L14 / L18	0.19	L14		
T 11	000075-28-5	ISOBUTANE	1.68	L66	5.54	L14	2.76	L14
T 12	000078-78-4	ISOPENTANE	1.75	L56	4.96	L14		
T 13	000107-83-5	2-METHYLPENTANE	1.85	L56	4.45	L14		
T 14	000096-14-0	3-METHYLPENTANE	1.84	L56	4.40	L3		
T 15	000079-29-8	2,3-DIMETHYLBUTANE	1.76	L56	4.50	L14	3.42	L14
T 16	000591-76-4	2-METHYLHEXANE	2.15	L56	3.94	L14		
T 17	000565-59-3	2,3-DIMETHYLPENTANE	1.85	L56	3.96	L14		
T 18	000108-08-7	2,4-DIMETHYLPENTANE	2.11	L66	4.02	L14		
T 19	000589-81-1	3-METHYLHEPTANE	2.18	L56	3.42	L14		
T 20	003221-61-2	2-METHYLOCTANE	2.68	L14 / L14	2.92	L14		
T 21	002216-34-4	4-METHYLOCTANE	2.61	L56	2.96	L14		
T 22	000075-83-2	2,2-DIMETHYLBUTANE	1.90	L66	4.63	L12		
T 23	000562-49-2	3,3-DIMETHYLPENTANE	1.88	L33	4.04	L14		
T 24	001067-20-5	3,3-DIETHYLPENTANE	1.63	L47	2.99	L14		
T 25	000564-02-3	2,2,3-TRIMETHYLPENTANE	1.91	L14 / L37	3.63	L14		
T 26	000540-84-1	2,2,4-TRIMETHYLPENTANE	2.12	L56	3.82	L14		
T 27	003522-94-9	2,2,5-TRIMETHYLHEXANE	2.33	L56	3.34	L14		
T 28	000075-19-4	CYCLOPROPANE	0.55	L47	5.86	L14	1.72	L14
T 29	000287-92-3	CYCLOPENTANE	0.88	L47	4.62	L32	3.00	L14
T 30	000110-82-7	CYCLOHEXANE	0.90	L47	4.11	L3	3.44	L14
T 31	000291-64-5	CYCLOHEPTANE	0.63	L14 / L78	3.46	L14	4.00	L14
T 32	000292-64-8	CYCLOOCTANE	0.62	L6				
T 33	000108-87-2	METHYLCYCLOHEXANE	1.19	L56	3.79	L14	3.61	L14
T 34	006876-23-9	1,2-DIMETHYLCYCLOHEXANE	1.16	L56	3.43	L32		
T 35	003741-00-2	PENTYLCYCLOPENTANE	1.87	L56				
T 36	000099-82-1	P-MENTHANE	1.86	L14 / L14	2.55	L14	5.56	L14
T 37	004516-69-2	1,1,3-TRIMETHYLCYCLOPENTANE	1.81	L56				

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 38	000074-85-1	ETHYLENE	0.92	L35			1.13	L14
T 39	000115-07-1	PROPENE	0.87	L35	6.06	L14	1.77	L14
T 40	000106-98-9	1-BUTENE	0.98	L66	5.40	L14	2.40	L14
T 41	000624-64-6	CIS-2-BUTENE	0.98	L14	5.33	L14	2.33	L14
T 42	000590-18-1	TRANS-2-BUTENE	0.96	L14	5.37	L14	2.31	L14
T 43	000592-41-6	1-HEXENE	1.16	L47	4.35	L3	3.39	L14
T 44	000592-76-7	1-HEPTENE	1.22	L47	3.90	L14	3.99	L14
T 45	014686-13-6	TRANS-2-HEPTENE	1.22	L66	3.81	L14		
T 46	000111-66-0	1-OCTENE	1.41	L47	3.37	L14	4.57	L14
T 47	000124-11-8	1-NONENE	1.51	L47	2.86	L14	5.15	L14
T 48	000115-11-7	ISOBUTENE	0.95	L14			2.34	L14
T 49	000563-46-2	2-METHYL-1-BUTENE	1.25	L14 / L16	4.91	L14		
T 50	000513-35-9	TRIMETHYLETHYLENE	0.98	L66	4.85	L3		
T 51	000142-29-0	CYCLOPENTENE	0.41	L33	4.71	L14		
T 52	000110-83-8	CYCLOHEXENE	0.27	L33	4.06	L3	2.86	L14
T 53	000591-49-1	1-METHYLCYCLOHEXENE	0.49	L33	3.61	L14		
T 54	000591-93-5	1,4-PENTADIENE	0.69	L66	5.00	L14	2.47	L97
T 55	000592-42-7	1,5-HEXADIENE	0.74	L33	4.47	L14	2.87	L14
T 56	000078-79-5	2-METHYL-1,3-BUTADIENE	0.50	L33	4.87	L14		
T 57	000542-92-7	CYCLOPENTADIENE	0.36	L14 / L16	4.76	L14		
T 58	000100-40-3	4-VINYLCYCLOHEXENE	0.30	L3 / L37	3.36	L3	3.93	L14
T 59	000586-62-9	TERPINOLENE	0.40	L15 / L15	1.90	L15	4.47	L14
T 60	000123-35-3	MYRCENE	1.10	L92 / L14	2.43	L14	4.17	L14
T 61	000544-25-2	CYCLOHEPTATRIENE	-0.73	L33	3.50	L14	2.63	L14
T 62	000074-86-2	ACETYLENE	-0.01	L47	5.84	L14	0.37	L14
T 63	000074-99-7	PROPYNE	-0.35	L47	5.76	L14	0.94	L14
T 64	000693-02-7	1-HEXYNE	0.21	L47	4.25	L14	2.73	L14
T 65	000628-71-7	1-HEPTYNE	0.44	L47	3.85	L14		
T 66	000629-05-0	1-OCTYNE	0.52	L47	3.26	L14		
T 67	003452-09-3	1-NONYNE	0.77	L47	2.92	L14		
T 68	000689-97-4	1-BUTEN-3-YNE	0.07	L14	5.23	L3		
T 69	000071-43-2	BENZENE	-0.67	L48	4.10	L3	2.13	L14
T 70	000108-88-3	TOLUENE	-0.65	L26	3.58	L3	2.73	L14
T 71	000095-47-6	O-XYLENE	-0.75	L56	2.94	L14	3.12	L14
T 72	000108-38-3	M-XYLENE	-0.57	L48	3.04	L3	3.20	L14
T 73	000106-42-3	P-XYLENE	-0.59	L33	3.07	L14	3.15	L14
T 74	000103-65-1	N-PROPYLBENZENE	-0.39	L33	2.66	L14	3.72	L97
T 75	000611-14-3	1-METHYL-2-ETHYLBENZENE	-0.64	L66	2.54	L14	3.53	L14
T 76	000622-96-8	1-ETHYL-4-METHYLBENZENE	-0.68	L14 / L20	2.60	L14		
T 77	000526-73-8	1,2,3-TRIMETHYLBENZENE	-0.89	L33	2.35	L14	3.59	L97
T 78	000095-63-6	1,2,4-TRIMETHYLBENZENE	-0.63	L33	2.45	L14	3.63	L97
T 79	000108-67-8	1,3,5-TRIMETHYLBENZENE	-0.51	L48	2.58	L3	3.42	L97
T 80	000104-51-8	BUTYLBENZENE	-0.29	L33	2.18	L62	4.38	L97
T 81	000135-01-3	O-DIETHYLBENZENE	-0.93	L3 / L8	2.18	L3	3.72	L97
T 82	000141-93-5	M-DIETHYLBENZENE	-0.47	L14 / L14	2.18	L14		
T 83	000105-05-5	1,4-DIETHYLBENZENE	-0.41	L3 / L21	2.15	L14		
T 84	000538-68-1	PENTYLBENZENE	-0.17	L47	1.77	L14	4.90	L97
T 85	000700-12-9	PENTAMETHYLBENZENE	-1.74	L14 / L14	0.67	L14	4.56	L97
T 86	000087-85-4	HEXAMETHYLBENZENE	-1.49	L14 / L14	-0.94	L14	4.61	L97
T 87	001078-71-3	N-HEPTYLBENZENE	-0.17	L14 / L14	0.62	L14		
T 88	002189-60-8	N-OCTYLBENZENE	0.24	L14 / L14	0.18	L14	6.30	L14

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 89	000135-98-8	2-BUTYLBENZENE	-0.13	L56	2.37	L14		
T 90	000527-84-4	1-METHYL-2-ISOPROPYLBENZENE	-0.33	L14	2.30	L14		
T 91	000099-87-6	P-ISOPROPYLTOLUEN	-0.34	L14 / L14	2.29	L14	4.10	L14
T 92	000496-11-7	INDANE	-1.07	L33			3.18	L14
T 93	000100-42-5	STYRENE	-0.91	L33	2.93	L14	2.95	L14
T 94	000100-80-1	M-METHYLSTYRENE	-0.63	L3 / L37	2.64	L3		
T 95	000622-97-9	P-METHYLSTYRENE	-0.89	L14 / L37	2.38	L14		
T 96	000536-74-3	ETHYNYL BENZENE	-1.60	L14 / L14	2.44	L14	2.53	L14
T 97	000092-52-4	BIPHENYL	-1.90	L46	0.08	L14	4.01	L97
T 98	000086-73-7	9H-FLUORENE	-2.39	L14	-1.05	L46	4.18	L97
T 99	000091-20-3	NAPHTHALENE	-1.76	L47	1.05	L14	3.30	L97
T 100	000090-12-0	1-METHYLNAPHTHALENE	-1.80	L56	1.00	L76	3.87	L97
T 101	000091-57-6	2-METHYLNAPHTHALENE	-1.73	L51	0.86	L14	3.86	L97
T 102	001127-76-0	1-ETHYLNAPHTHALENE	-1.72	L14	0.53	L14	4.39	L97
T 103	000939-27-5	2-ETHYLNAPHTHALENE	-1.66	L66	0.62	L14	4.38	L14
T 104	000575-41-7	1,3-DIMETHYLNAPHTHALENE	-1.81	L33	0.45	L76	4.42	L14
T 105	000571-58-4	1,4-DIMETHYLNAPHTHALENE	-1.88	L76 / L21	0.41	L76	4.37	L14
T 106	000571-61-9	1,5-DIMETHYLNAPHTHALENE	-1.82	L66	-0.16	L76	4.38	L14
T 107	000581-40-8	2,3-DIMETHYLNAPHTHALENE	-2.04	L33			4.40	L14
T 108	002131-41-1	1,4,5-TRIMETHYLNAPHTHALENE	-2.02	L56			4.90	L14
T 109	000083-32-9	ACENAPHTHENE	-2.31	L33	-0.48	L14	3.92	L14
T 110	000207-08-9	BENZO[K]FLUORANTHENE	-4.74	L66	-7.28	L46		
T 111	000205-99-2	BENZO[B]FLUORANTHENE	-4.68	L66			5.78	L14
T 112	000218-01-9	CHRYSENE	-3.74	L56	-6.08	L14	5.73	L97
T 113	000217-59-4	TRIPHENYLENE	-3.29	L14 / L14	-3.64	L14	5.49	L14
T 114	000208-96-8	ACENAPHTHYLENE	-2.46	L46 / L14	-0.05	L46		
T 115	000129-00-0	PYRENE	-3.42	L51	-3.22	L46	4.88	L14
T 116	000139-39-5	INDENO-[1,2,3-CD]-PYRENE	-4.92	L66	-7.88	L14		
T 117	000050-32-8	BENZO(A)PYRENE	-4.52	L66	-6.15	L46	5.97	L14
T 118	000192-97-2	BENZO(E)PYRENE	-4.00	L66			6.44	L14
T 119	000198-55-0	PERYLENE	-3.82	L14 / L14	-6.23	L14	5.82	L97
T 120	000191-24-2	BENZO[GHI]PERYLENE	-4.85	L27			6.63	L97
T 121	007732-18-5	WATER	-4.61	L91	3.50	L91	-1.38	L91
T 122	000630-08-0	CARBON MONOXIDE	1.63	L91				
T 123	000124-38-9	CARBON DIOXIDE	0.08	L91	6.18	L91		
T 124	001333-74-0	HYDROGEN	1.71	L91				
T 125	007722-84-1	HYDROGEN PEROXIDE			3.42	L91	-1.18	L91
T 126	000067-56-1	METHANOL	-3.71	L48	4.23	L14	-0.77	L14
T 127	000064-17-5	ETHANOL	-3.59	L65	3.90	L14	-0.31	L14
T 128	000071-23-8	1-PROPANOL	-3.50	L65	3.45	L14	0.25	L14
T 129	000071-36-3	1-BUTANOL	-3.44	L61	2.95	L14	0.88	L14
T 130	000071-41-0	1-PENTANOL	-3.27	L14	2.47	L14	1.56	L97
T 131	000111-70-6	1-HEPTANOL	-3.09	L47	1.46	L14	2.72	L97
T 132	000111-87-5	1-OCTANOL	-3.00	L61	1.02	L14	3.00	L97
T 133	000112-30-1	1-DECANOL	-2.67	L47	0.05	L14	4.57	L14
T 134	000112-53-8	DODECANOL	-2.44	L14 / L8	-0.95	L14	5.13	L97
T 135	000112-92-5	OCTADECANOL	-1.44	L14 / L14	-3.44	L14		
T 136	000078-83-1	ISOBUTYL ALCOHOL	-3.30	L47	3.14	L14	0.76	L97
T 137	000105-30-6	2-METHYL-1-PENTANOL	-2.97	L15 / L8	2.41	L14		
T 138	054206-54-1	2,3-DIMETHYLBUTANOL	-2.87	L65				
T 139	000075-84-3	2,2-DIMETHYL-1-PROPANOL	-2.66	L14 / L13	3.43	L3	1.31	L97

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 140	000067-63-0	ISOPROPANOL	-3.36	L66	3.78	L14	0.05	L97
T 141	006032-29-7	2-PENTANOL	-3.22	L14	2.88	L3	1.19	L97
T 142	000543-49-7	2-HEPTANOL	-2.63	L14 / L8	2.21	L14	2.31	L97
T 143	000589-82-2	3-HEPTANOL	-2.94	L14 / L14	1.99	L14	2.24	L97
T 144	000628-99-9	2-NONANOL	-2.70	L14 / L14	0.95	L14		
T 145	000598-75-4	3-METHYL-2-BUTANOL	-3.13	L14 / L16	3.08	L14	1.28	L97
T 146	000108-11-2	4-METHYL-2-PENTANOL	-2.74	L14	2.85	L14		
T 147	000108-82-7	2,6-DIMETHYL-4-HEPTANOL	-2.54	L14 / L37	1.70	L14		
T 148	000096-41-3	CYCLOPENTANOL	-4.03	L47	2.30	L3		
T 149	000108-93-0	CYCLOHEXANOL	-3.94	L43	2.03	L14	1.23	L97
T 150	000502-41-0	CYCLOHEPTANOL	-4.02	L47				
T 151	000583-59-5	2-METHYLCYCLOHEXANOL	-3.51	L66			1.84	L97
T 152	000591-23-1	3-METHYLCYCLOHEXANOL	-3.82	L66				
T 153	000589-91-3	4-METHYLCYCLOHEXANOL	-3.31	L15 / L36	2.20	L15	1.79	L97
T 154	000075-65-0	T-BUTANOL	-3.23	L14			0.35	L97
T 155	000594-60-5	2,3-DIMETHYL-2-BUTANOL	-3.35	L14 / L8	2.63	L14		
T 156	000115-18-4	2-METHYL-3-BUTENE-2-OL	-3.07	L66				
T 157	000098-55-5	ALPHA-TERPINEOL	-3.45	L14 / L57	0.60	L14		
T 158	000115-19-5	2-METHYL-3-BUTYN-2-OL	-3.80	L66	3.33	L14	0.28	L97
T 159	000077-75-8	METHYL PENTYNOL	-3.66	L14 / L37	2.85	L14		
T 160	000107-21-1	ETHYLENE GLYCOL	-5.61	L14	1.09	L14	-1.36	L97
T 161	000094-96-2	2-ETHYL-1,3-HEXANEDIOL	-2.73	L14 / L14	3.12	L14		
T 162	000115-77-5	PENTAERYTHRITOL	-3.65	L92 / L3	2.30	L92		
T 163	000100-51-6	PHENYLMETHANOL	-4.86	L33	0.90	L3	1.10	L97
T 164	000060-12-8	2-PHENYLETHANOL	-4.62	L14 / L17	1.06	L14	1.36	L97
T 165	000108-95-2	PHENOL	-4.60	L47	1.67	L14	1.47	L97
T 166	000106-44-5	P-CRESOL	-4.50	L47	1.17	L14	1.94	L97
T 167	000526-75-0	2,3-DIMETHYLPHENOL	-4.52	L33			2.00	L97
T 168	000095-87-4	2,5-DIMETHYLPHENOL	-4.34	L33			2.33	L14
T 169	000576-26-1	2,6-DIMETHYLPHENOL	-3.86	L33			2.36	L14
T 170	000095-65-8	3,4-DIMETHYLPHENOL	-4.77	L33			2.23	L14
T 171	000108-68-9	3,5-DIMETHYLPHENOL	-4.60	L33			2.35	L14
T 172	000123-07-9	P-ETHYLPHENOL	-4.50	L33	0.70	L14	2.47	L97
T 173	000645-56-7	P-PROPYLPHENOL	-4.33	L33			3.00	L97
T 174	000098-54-4	P-T-BUTYLPHENOL	-4.31	L14	0.60	L3	3.31	L97
T 175	000090-15-3	1-NAPHTHOL	-5.63	L33			2.84	L97
T 176	000135-19-3	2-NAPHTHOL	-5.95	L33	0.80	L3	2.70	L97
T 177	000108-46-3	1,3-BENZENEDIOL	-7.36	L3 / L13	0.15	L3	0.80	L97
T 178	000115-10-6	DIMETHYL ETHER	-1.40	L33	5.78	L14	0.10	L97
T 179	000540-67-0	METHYL ETHYL ETHER	-1.54	L33	5.30	L14		
T 180	000060-29-7	DIETHYL ETHER	-1.17	L47	4.85	L3	0.89	L14
T 181	000598-53-8	METHYL ISOPROPYL ETHER	-1.43	L14 / L14	4.91	L14		
T 182	000628-28-4	METHYL N-BUTYL ETHER	-1.13	L14 / L14	4.27	L14	1.66	L14
T 183	006795-87-5	SEC-BUTYL METHYL ETHER	-1.22	L14 / L14	4.44	L14		
T 184	000108-20-3	DIISOPROPYL ETHER	-0.91	L14 / L42	4.30	L14	1.52	L14
T 185	000142-96-1	DI-N-BUTYL ETHER	-0.74	L48	2.90	L14	3.21	L14
T 186	001634-04-4	METHYL T-BUTYL ETHER	-1.34	L14 / L3	4.52	L14	0.94	L14
T 187	000625-44-5	METHYL ISOBUTYL ETHER	-1.04	L14 / L14	4.45	L14		
T 188	000919-94-8	ETHYL T-AMYL ETHER	-1.11	L14 / L14	3.82	L14		
T 189	000107-25-5	METHYL VINYL ETHER	-0.80	L14 / L3	5.24	L14		
T 190	000111-34-2	BUTYL VINYL ETHER	-1.03	L3 / L37	3.81	L14		

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 191	000109-53-5	ISOBUTYL VINYL ETHER	-0.82	L3 / L14	3.90	L14		
T 192	000109-93-3	DIVINYL ETHER	-0.48	L14 / L16	4.95	L14		
T 193	000109-87-5	DIMETHOXYMETHANE	-2.15	L14	4.72	L14	0.18	L97
T 194	000112-48-1	1,2-DIBUTOXYETHANE	-3.25	L15 / L36	1.20	L15	2.47	L97
T 195	000112-73-2	DIETHYLENE GLYCOL DIBUTYL ETHER	-3.99	L14 / L14	0.54	L14	1.92	L14
T 196	000100-66-3	METHOXYBENZENE	-1.80	L47	2.67	L14	2.11	L14
T 197	000101-84-8	DIPHENYL ETHER	-2.00	L14 / L17	0.48	L14	4.21	L14
T 198	000075-21-8	ETHYLENE OXIDE	-2.31	L19	5.24	L14	-0.30	L14
T 199	000109-99-9	TETRAHYDROFURAN	-2.54	L14	4.34	L14	0.47	L97
T 200	000106-88-7	1,2-BUTYLENEOXIDE	-2.13	L14 / L13	4.38	L14		
T 201	000096-47-9	2-METHYLTETRAHYDROFURAN	-2.42	L14	4.11	L14		
T 202	026093-63-0	3-METHYLTETRAHYDROPYRAN	-2.09	L3 / L42	3.60	L3		
T 203	000470-82-6	1,8-CINEOLE	-2.29	L14 / L13	2.40	L14	2.50	L14
T 204	000123-91-1	1,4-DIOXANE	-3.71	L14	3.71	L14	-0.27	L14
T 205	000110-88-3	1,3,5-TRIOXANE	-3.47	L3 / L13	3.20	L3	-0.43	L14
T 206	000123-63-7	PARALDEHYDE	-2.72	L15 / L3	3.17	L14	0.67	L14
T 207	000110-00-9	FURANE	-0.65	L3 / L14	4.92	L3	1.34	L14
T 208	000534-22-5	2-METHYLFURAN	-0.61	L3 / L36	4.35	L3	1.85	L14
T 209	000096-09-3	STYRENE OXIDE	-3.09	L14 / L14	1.70	L14	1.61	L14
T 210	000132-64-9	DIBENZOFURAN	-2.63	L14 / L13	-0.46	L14	4.12	L14
T 211	000122-60-1	(PHENOXYMETHYL)-OXIRANE	-4.48	L14 / L37	0.12	L14		
T 212	087818-31-3	CINMETHYLIN	-4.75	L14 / L37	-2.00	L14		
T 213	063837-33-2	DIOFENOLAN	-5.56	L99 / L99	-3.96	L99	4.35	L14
T 214	000110-05-4	BIS(1,1-DIMETHYLETHYL)PEROXIDE	-0.05	L15 / L94	3.50	L15		
T 215	000075-91-2	TERT-BUTYLHYDROPEROXIDE	-3.44	L14 / L3	2.86	L14		
T 216	000050-00-0	FORMALDEHYDE	-2.02	L47			0.35	L14
T 217	000075-07-0	ACETALDEHYDE	-2.57	L61	5.08	L14		
T 218	000123-38-6	PROPIONALDEHYDE	-2.47	L66	4.62	L14	0.59	L14
T 219	000111-71-7	HEPTANAL	-1.96	L61	2.60	L15		
T 220	000124-13-0	OCTANAL	-1.72	L48	2.28	L81		
T 221	000124-19-6	NONANAL	-1.53	L65	1.69	L14		
T 222	000097-96-1	2-ETHYLBUTYRALDEHYDE	-1.67	L3 / L44	3.35	L3		
T 223	000123-15-9	2-METHYLVALERALDEHYDE	-1.67	L3 / L44	3.40	L3		
T 224	000123-05-7	2-ETHYLHEXANEALDEHYDE	-1.45	L14 / L44	2.42	L14		
T 225	000107-02-8	ACROLEIN	-2.90	L5	4.56	L14	-0.01	L14
T 226	000123-73-9	TRANS-CROTONALDEHYDE	-2.80	L5	3.70	L14		
T 227	000505-57-7	TRANS-2-HEXENAL	-2.69	L65				
T 228	000078-85-3	ALPHA-METHYLACROLEIN	-2.02	L59	4.30	L14		
T 229	000645-62-5	2-ETHYL-2-HEXENAL	-1.93	L3 / L15	2.20	L3		
T 230	000142-83-6	2,4-HEXADIENAL	-3.40	L65				
T 231	000100-52-7	BENZALDEHYDE	-3.04	L43	2.10	L3	1.47	L97
T 232	000067-64-1	ACETONE	-2.79	L14	4.49	L14	-0.24	L97
T 233	000107-87-9	2-PENTANONE	-2.49	L66	3.67	L14	0.91	L97
T 234	000096-22-0	DIETHYL KETONE	-2.50	L47	3.67	L15		
T 235	000591-78-6	METHYL BUTYL KETONE	-2.41	L47	3.19	L14	1.38	L97
T 236	000589-38-8	3-HEXANONE	-2.29	L14	3.27	L14		
T 237	000110-43-0	2-HEPTANONE	-2.23	L61	2.71	L14	1.98	L97
T 238	000106-35-4	3-HEPTANONE	-2.34	L14 / L42	2.65	L14		
T 239	000111-13-7	2-OCTANONE	-2.11	L61	2.15	L3	2.37	L97
T 240	000106-68-3	3-OCTANONE	-2.06	L97	2.43	L14		
T 241	000502-56-7	DIBUTYL KETONE	-2.01	L3 / L1	1.80	L3		

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 242	000821-55-6	2-NONANONE	-2.01	L48	1.92	L14	3.14	L97
T 243	000693-54-9	2-DECANONE	-1.72	L47			3.73	L97
T 244	000563-80-4	METHYL ISOPROPYL KETONE	-2.38	L47	3.84	L14	0.84	L97
T 245	000108-10-1	METHYL ISOBUTYL KETONE	-2.24	L47	3.42	L3	1.31	L97
T 246	000565-84-0	DIISOPROPYL KETONE	-2.01	L33	3.30	L3	1.86	L14
T 247	000541-85-5	5-METHYL-3-HEPTANONE	-2.09	L3 / L42	2.43	L3		
T 248	000108-83-8	2,6-DIMETHYL-4-HEPTANONE	-1.87	L3 / L37	2.79	L3		
T 249	000108-94-1	CYCLOHEXANONE	-3.32	L69	2.76	L14	0.81	L14
T 250	000563-60-8	2-METHYLCYCLOHEXANONE	-3.00	L3 / L42	2.60	L3		
T 251	000823-76-7	METHYL CYCLOHEXYL KETONE	-2.86	L33	2.15	L3		
T 252	000078-94-4	METHYL VINYL KETONE	-2.72	L59	4.05	L15		
T 253	000078-59-1	ISOPHORONE	-3.61	L3 / L14	1.72	L3		
T 254	000431-03-8	2,3-BUTANEDIONE	-3.26	L68	3.88	L14	-1.34	L14
T 255	000123-54-6	ACETYLACETONE	-3.92	L14 / L42	2.70	L14	0.40	L14
T 256	000098-86-2	ACETOPHENONE	-3.37	L59	1.72	L14	1.58	L14
T 257	000122-00-9	4-METHYLACETOPHENONE	-3.45	L33			2.10	L14
T 258	000084-65-1	ANTHRAQUINONE	-5.10	L99 / L99	-5.10	L99	3.39	L14
T 259	000082-66-6	DIPHACINONE	-8.20	L3 / L15	-7.86	L3		
T 260	000106-51-4	1,4-BENZOQUINONE	-4.39	L96 / L13	1.10	L96	0.20	L14
T 261	000078-98-8	METHYL GLYOXAL	-4.96	L72				
T 262	000064-18-6	FORMIC ACID	-4.91	L47	3.75	L14	-0.54	L14
T 263	000064-19-7	ACETIC ACID	-5.13	L43	3.32	L14	-0.17	L14
T 264	000079-09-4	PROPIONIC ACID	-5.14	L43	2.67	L14	0.33	L14
T 265	000107-92-6	BUTYRIC ACID	-5.01	L43	2.10	L14	0.79	L14
T 266	000142-62-1	N-HEXANOIC ACID	-4.56	L47	0.76	L14	1.92	L14
T 267	000111-14-8	N-HEPTANOIC ACID	-4.52	L47	-0.13	L14		
T 268	000124-07-2	CAPRYLIC ACID	-4.44	L47	-0.34	L14	3.05	L14
T 269	000079-31-2	ISOBUTYRIC ACID	-4.39	L14 / L3	2.38	L14		
T 270	000503-74-2	ISOVALERIC ACID	-4.47	L47	1.77	L14	1.16	L14
T 271	000116-53-0	2-METHYLBUTANOIC ACID	-4.22	L14 / L14	1.82	L14	1.18	L14
T 272	000088-09-5	2-ETHYLBUTYRIC ACID	-4.18	L14 / L36	1.40	L14	1.68	L14
T 273	000144-62-7	OXALIC ACID	-8.23	L66	-0.50	L15		
T 274	000065-85-0	BENZOIC ACID	-5.54	L19			1.87	L14
T 275	000079-21-0	PEROXYACETIC ACID	-4.07	L66	3.29	L14		
T 276	000107-31-3	METHYLFORMATE	-1.98	L6	4.89	L14	0.03	L14
T 277	000079-20-9	METHYL ACETATE	-2.28	L65	4.46	L14	0.18	L14
T 278	000141-78-6	ETHYL ACETATE	-2.16	L47	4.10	L14	0.73	L14
T 279	000625-55-8	ISOPROPYL FORMATE	-1.48	L47	4.26	L14		
T 280	000109-60-4	N-PROPYL ACETATE	-2.05	L14	3.65	L14	1.24	L14
T 281	000108-21-4	ISOPROPYL ACETATE	-1.94	L47	3.90	L14		
T 282	000592-84-7	BUTYL FORMATE	-1.68	L3 / L13	3.58	L3		
T 283	000105-37-3	ETHYL PROPIONATE	-2.05	L65	3.67	L3	1.21	L14
T 284	000542-55-2	ISOBUTYL FORMATE	-1.62	L65	3.73	L14		
T 285	000624-24-8	METHYL VALERATE	-1.89	L61	3.41	L14	1.96	L14
T 286	000123-86-4	N-BUTYL ACETATE	-1.94	L14	3.18	L14	1.78	L14
T 287	000105-54-4	ETHYL BUTYRATE	-1.84	L66	3.35	L14		
T 288	000105-46-4	SEC-BUTYL ACETATE	-1.67	L14 / L13	3.45	L14	1.72	L14
T 289	000598-98-1	METHYL TRIMETHYLACETATE	-1.76	L33			1.83	L14
T 290	000106-70-7	METHYL HEXANOATE	-1.82	L61	2.70	L3		
T 291	000590-01-2	N-BUTYL PROPIONATE	-1.70	L3 / L13	2.75	L3		
T 292	000626-38-0	2-PENTANOL ACETATE	-1.48	L15 / L15	3.10	L15		

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 293	000105-66-8	N-PROPYL BUTYRATE	-1.67	L47	2.90	L14		
T 294	000539-82-2	ETHYL VALERATE	-1.83	L47	2.81	L14		
T 295	000142-92-7	N-HEXYL ACETATE	-1.73	L3 / L13	2.20	L3		
T 296	000109-21-7	BUTYL BUTYRATE	-1.55	L14 / L37	2.38	L14		
T 297	000624-54-4	AMYL PROPIONATE	-1.55	L47	2.59	L3		
T 298	000111-11-5	METHYL OCTANOATE	-1.50	L61	-3.39	L14		
T 299	000110-45-2	ISOAMYL FORMATE	-1.56	L14	3.30	L3		
T 300	000540-88-5	TERT-BUTYL ACETATE	-1.54	L3 / L37	3.80	L14	1.76	L14
T 301	000123-92-2	ISOAMYL ACETATE	-1.62	L14	2.87	L14		
T 302	000108-64-5	3-METHYLBUTANOIC ACID ETHYL ESTER	-1.69	L14 / L37	3.05	L14		
T 303	000097-85-8	ISOBUTYL ISOBUTYRATE	-1.47	L14 / L14	2.76	L14		
T 304	000108-84-9	4-METHYL-2-PENTYL ACETATE	-1.50	L3 / L37	2.85	L3		
T 305	000057-57-8	BETA-PROPIOLACTONE	-4.44	L15 / L14	2.66	L15		
T 306	002868-37-3	METHYL CYCLOPROPYLCARBOXYLATE	-3.01	L33				
T 307	004630-82-4	METHYL CYCLOHEXYLCARBOXYLATE	-2.42	L33				
T 308	000108-05-4	VINYL ACETATE	-1.68	L14 / L14	4.08	L14	0.73	L14
T 309	000080-62-6	METHYL METHACRYLATE	-1.88	L14 / L13	3.71	L14	1.38	L14
T 310	000103-11-7	2-ETHYLHEXYL ACRYLATE	-1.75	L14 / L14	1.38	L14		
T 311	000097-86-9	ISOBUTYL METHACRYLATE	-1.65	L3 / L36	2.75	L3	2.66	L14
T 312	041096-46-2	HYDROPRENE	-2.20	L99 / L37	-1.50	L99		
T 313	054406-48-3	EMPENTHRIN	-1.85	L99 / L99	-1.85	L99		
T 314	000553-90-2	METHYL OXALATE	-3.92	L14 / L14	2.18	L14	-0.17	L14
T 315	000106-65-0	DIMETHYL SUCCINATE	-4.57	L92 / L42	1.75	L92	0.35	L14
T 316	000123-25-1	ETHYL SUCCINATE	-4.66	L14 / L1	0.77	L14	1.19	L97
T 317	002050-20-6	DIETHYL PIMELATE	-4.73	L14 / L14	-0.38	L14		
T 318	000093-58-3	METHYL BENZOATE	-2.88	L33	1.70	L14	2.12	L97
T 319	000093-89-0	ETHYL BENZOATE	-2.67	L47	1.55	L14	2.64	L97
T 320	000084-66-2	DIETHYL PHTHALATE	-4.88	L14 / L94	-0.66	L14	2.47	L97
T 321	000131-16-8	DIPROPYL PHTHALATE	-4.78	L14 / L13	-1.75	L14		
T 322	000084-74-2	DIBUTYL PHTHALATE	-4.13	L14	-2.01	L14	4.72	L97
T 323	000084-75-3	DIHEXYL PHTHALATE	-3.60	L14 / L14	-3.33	L14		
T 324	000117-84-0	DIOCTYL PHTHALATE	-4.00	93 / L18	-4.90	93		
T 325	000105-58-8	DIETHYL CARBONATE	-2.43	L3 / L13	3.16	L3	1.21	L97
T 326	000108-32-7	PROPYLENE CARBONATE	-5.54	L47	0.75	L92	-0.41	L97
T 327	000108-24-7	ACETIC ANHYDRIDE	-3.68	L3 / L3	2.83	L3		
T 328	000108-31-6	MALEIC ANHYDRIDE	-5.69	L92 / L92	1.45	L92		
T 329	000094-36-0	BENZOYL PEROXIDE	-4.17	L94 / L36	2.20	L94		
T 330	000109-86-4	2-METHOXYETHANOL	-4.96	L33	3.10	L14	-0.77	L14
T 331	000109-59-1	2-PROPOXYETHANOL	-4.70	L33				
T 332	000111-76-2	2-BUTOXYETHANOL	-4.59	L33	2.07	L14	0.83	L14
T 333	000112-25-4	2-HEXYLOXYETHANOL	-4.39	L14 / L36	0.83	L14	1.86	L14
T 334	000122-99-6	2-PHENOXYETHANOL	-5.71	L14 / L17	-0.03	L14	1.16	L14
T 335	000150-19-6	3-METHOXYPHENOL	-5.62	L33			1.58	L97
T 336	000097-53-0	EUGENOL	-4.37	L15 / L13	0.20	L15		
T 337	000091-10-1	2,6-DIMETHOXYPHENOL	-5.09	L71	-0.35	L77	1.15	L97
T 338	000141-46-8	GLYCOLALDEHYDE	-6.01	L72				
T 339	000090-02-8	SALICYLALDEHYDE	-2.89	L14 / L44	1.90	L14	1.81	L97
T 340	000123-08-0	4-HYDROXYBENZALDEHYDE	-6.48	L33			1.35	L97
T 341	000129-43-1	1-HYDROXYANTHRAQUINONE	-6.53	L14 / L14	-4.55	L14	3.53	L97
T 342	000605-32-3	2-HYDROXYANTHRAQUINONE	-9.12	L14 / L14	-8.04	L14		
T 343	000069-72-7	SALICYLIC ACID	-6.53	L14 / L13	-1.96	L14	2.26	L97

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 344	000138-22-7	BUTYL LACTATE	-3.97	L15 / L37	1.85	L15		
T 345	002314-09-2	N-BUTYL-9-HYDROXYFLUORENE-(9)-CARB.	-5.01	L3 / L13	-0.51	L3		
T 346	095266-40-3	TRINEXAPAC-ETHYL	-7.96	L99 / L99	-2.67	L99		
T 347	000098-01-1	FURFURAL	-3.82	L14 / L13	2.47	L14	0.41	L97
T 348	000100-06-1	4-METHOXYACETOPHENONE	-3.23	L33	0.12	L92	1.74	L97
T 349	000108-65-6	PROPYLENE GLYCOL METHYL ETHER	-3.75	L14 / L14	2.72	L14	0.56	L14
T 350	040596-69-8	METHOPRENE	-3.55	L3 / L37	-2.50	L3	5.50	L14
T 351	000124-17-4	2-(2-BUTOXYETHOXY)ETHANOL ACETATE	-5.46	L15 / L13	0.20	L15		
T 352	026002-80-2	PHENOTHRIN	-5.54	L97	-4.72	L14		
T 353	000127-17-3	PYRUVIC ACID	-6.90	L45	2.24	L14		
T 354	000105-45-3	METHYL ACETOACETATE	-3.95	L3 / L14	2.90	L3		
T 355	000141-97-9	ETHYL ACETOACETATE	-3.84	L3 / L13	2.48	L3	0.24	L97
T 356	000074-89-5	METHYLAMINE	-3.34	L14	5.55	L3	-0.57	L97
T 357	000075-04-7	ETHYLAMINE	-3.38	L65	5.16	L3	-0.13	L97
T 358	000107-10-8	PROPYLAMINE	-3.30	L65	4.61	L3	0.47	L97
T 359	000109-73-9	N-BUTYLAMINE	-3.21	L65	4.10	L3	0.97	L97
T 360	000110-58-7	N-PENTYLAMINE	-3.00	L14	3.49	L3	1.49	L97
T 361	000111-26-2	N-HEXYLAMINE	-2.96	L14	3.08	L14	2.06	L97
T 362	000111-68-2	HEPTYLAMINE	-2.78	L47	2.56	L14	2.57	L97
T 363	000111-86-4	OCTYLAMINE	-2.68	L47	2.11	L14	2.90	L14
T 364	002869-34-3	TRIDECYLAMINE	-2.35	L66				
T 365	000108-91-8	CYCLOHEXYLAMINE	-3.77	L66	3.21	L3	1.49	L97
T 366	000107-15-3	1,2-DIAMINOETHANE	-7.15	L14	3.21	L3	-2.04	L97
T 367	000124-40-3	DIMETHYLAMINE	-3.14	L47	5.32	L3	-0.38	L97
T 368	000109-89-7	DIETHYLAMINE	-2.98	L14	4.51	L3	0.58	L97
T 369	000142-84-7	DIPROPYLAMINE	-2.68	L14	3.43	L14	1.67	L97
T 370	000108-18-9	DIISOPROPYLAMINE	-2.36	L47	4.07	L3	1.40	L97
T 371	000123-75-1	PYRROLIDINE	-4.01	L14	3.98	L3	0.47	L97
T 372	000110-89-4	PIPERIDINE	-3.74	L14	3.64	L3	0.84	L97
T 373	000111-49-9	HEXAMETHYLENIMINE	-3.60	L14	3.03	L14		
T 374	000075-50-3	TRIMETHYLAMINE	-2.35	L47	5.34	L3	0.16	L97
T 375	000121-44-8	TRIETHYLAMINE	-2.22	L14	3.88	L14	1.45	L97
T 376	000120-94-5	1-METHYL-PYRROLIDINE	-2.91	L14			0.92	L97
T 377	000626-67-5	N-METHYLPYRROLIDINE	-2.85	L14			1.30	L97
T 378	000098-94-2	N,N-DIMETHYLCYCLOHEXYLAMINE	-3.02	L66	2.61	L3		
T 379	000062-53-3	ANILINE	-4.03	L47	1.79	L14	0.90	L97
T 380	000108-44-1	M-METHYLANILINE	-4.17	L66	1.56	L90	1.40	L97
T 381	000106-49-0	4-METHYLANILINE	-4.04	L41			1.39	L97
T 382	000095-53-4	O-METHYLANILINE	-4.06	L33	1.54	L14	1.32	L97
T 383	000087-62-7	2,6-DIMETHYLANILINE	-3.82	L33	1.24	L14	1.84	L97
T 384	000095-64-7	3,4-DIMETHYLANILINE	-4.12	L95				
T 385	000137-17-7	2,4,5-TRIMETHYLANILINE	-3.98	L41				
T 386	000134-32-7	1-NAPHTHYLAMINE	-5.34	L33			2.25	L97
T 387	000091-59-8	2-NAPHTHYLAMINE	-5.48	L33			2.28	L97
T 388	000103-69-5	N-ETHYLANILINE	-3.22	L5	1.70	L94	2.16	L97
T 389	000121-69-7	N,N-DIMETHYLANILINE	-2.53	L47	1.97	L14	2.31	L97
T 390	000074-90-8	HYDROCYANIC ACID	-2.26	L14	5.00	L3	-0.25	L97
T 391	000075-05-8	ACETONITRILE	-2.85	L14	4.09	L3	-0.34	L97
T 392	000107-12-0	PROPIONITRILE	-2.67	L69	3.80	L14	0.16	L97
T 393	000109-74-0	BUTYRONITRILE	-2.56	L69	3.41	L14	0.53	L97
T 394	000078-82-0	ISOBUTYRONITRILE	-2.43	L48	3.64	L14	0.46	L14

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 395	000107-13-1	ACRYLONITRILE	-2.35	L19	4.16	L3	0.25	L97
T 396	000100-47-0	BENZONITRILE	-3.09	L33	2.01	L14	1.56	L97
T 397	000109-97-7	PYRROLE	-3.14	L69	3.05	L14	0.75	L97
T 398	000110-86-1	PYRIDINE	-3.44	L100	3.44	L14	0.65	L97
T 399	000109-06-8	2-METHYLPYRIDINE	-3.39	L14	3.15	L3	1.11	L97
T 400	000108-99-6	3-METHYLPYRIDINE	-3.50	L14			1.20	L97
T 401	000108-89-4	4-METHYLPYRIDINE	-3.62	L14	2.85	L3	1.22	L97
T 402	000100-71-0	2-ETHYLPYRIDINE	-3.17	L14			1.69	L97
T 403	000536-78-7	3-ETHYLPYRIDINE	-3.37	L14			1.66	L97
T 404	000536-75-4	4-ETHYLPYRIDINE	-3.47	L100			1.65	L97
T 405	000583-61-9	2,3-DIMETHYLPYRIDINE	-3.54	L14				
T 406	000108-47-4	2,4-DIMETHYLPYRIDINE	-3.39	L66				
T 407	000583-58-4	3,4-DIMETHYLPYRIDINE	-3.83	L14				
T 408	000591-22-0	3,5-DIMETHYLPYRIDINE	-3.55	L100			1.78	L97
T 409	000589-93-5	2,5-DIMETHYLPYRIDINE	-3.46	L14				
T 410	000108-48-5	2,6-LUTIDINE	-3.37	L14			1.68	L97
T 411	003978-81-2	4-T-BUTYLPYRIDINE	-3.27	L47				
T 412	000091-22-5	QUINOLINE	-4.97	L4	0.90	L14	2.03	L97
T 413	000109-08-0	2-METHYLPYRAZINE	-4.05	L33			0.23	L97
T 414	013925-00-3	2-ETHYL PYRAZINE	-4.00	L33			0.69	L97
T 415	029460-93-3	2-ISOBUTYLPYRAZINE	-3.70	L33				
T 416	000100-54-9	3-CYANOPYRIDINE	-4.95	L33			0.23	L97
T 417	000100-48-1	4-CYANOPYRIDINE	-4.42	L33			0.47	L97
T 418	000141-43-5	ETHANOLAMINE	-8.17	L88	1.80	L3	-1.31	L97
T 419	000873-62-1	3-CYANOPHENOL	-7.08	L33			1.70	L97
T 420	000767-00-0	4-CYANOPHENOL	-7.46	L47			1.60	L97
T 421	000090-04-0	2-METHOXYANILINE	-4.24	L94	0.91	L3	1.18	L97
T 422	000536-90-3	3-METHOXYANILINE	-5.35	L33	1.00	L14	0.93	L97
T 423	000104-94-9	4-METHOXYANILINE	-5.49	L33			0.95	L97
T 424	000110-91-8	MORPHOLINE	-5.26	L33	3.12	L3	-0.86	L97
T 425	025680-58-4	2-ETHYL-3-METHOXPYRAZINE	-3.22	L65			1.80	L97
T 426	024683-00-9	2-ISOBUTYL-3-METHOXPYRAZINE	-2.70	L66				
T 427	000500-22-1	3-FORMYLPYRIDINE	-5.21	L33			0.29	L14
T 428	000872-85-5	4-FORMYLPYRIDINE	-5.14	L33			0.43	L97
T 429	000350-03-8	3-ACETYLPYRIDINE	-6.06	L33			0.43	L97
T 430	001122-54-9	4-ACETYLPYRIDINE	-5.59	L33			0.48	L14
T 431	012771-68-5	ANCYMIDOL	-9.78	L97			1.91	L14
T 432	000068-12-2	N,N'-DIMETHYLFORMAMIDE	-5.73	L33	2.75	L3	-1.01	L97
T 433	000127-19-5	N,N'-DIMETHYLACETAMIDE	-6.27	L14	2.42	L14	-0.77	L97
T 434	001119-49-9	N-BUTYLACETAMIDE	-6.83	L33				
T 435	004030-18-6	N-ACETYLPYRROLIDINE	-7.19	L33				
T 436	000055-21-0	BENZAMIDE	-8.07	L33			0.64	L97
T 437	024691-80-3	FENFURAM	-7.79	L97				
T 438	014214-32-5	DIFENOXYURON	-11.15	L97				
T 439	149877-41-8	BIFENAZATE	-6.40	L89			3.40	L98
T 440	000062-75-9	N-NITROSODIMETHYLAMINE	-4.15	L14	1.80	L3	-0.57	L97
T 441	000601-77-4	DI-I-PROPYLNITROSOAMINE	-3.51	L14			1.38	L97
T 442	000621-64-7	N-NITROSODIPROPYLAMINE	-3.67	L14	1.20	L15	1.36	L97
T 443	000924-16-3	N-NITROSODIBUTYLAMINE	-3.29	L14	0.60	L14	2.63	L97
T 444	000930-55-2	N-NITROSOPYRROLIDINE	-5.70	L14			-0.19	L97
T 445	000100-75-4	N-NITROSOPIPERIDINE	-4.48	L14			0.36	L97

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 446	000075-52-5	NITROMETHANE	-2.93	L14	3.68	L14	-0.35	L97
T 447	000108-03-2	1-NITROPROPANE	-2.45	L14	3.13	L14	0.87	L14
T 448	000079-46-9	2-NITROPROPANE	-2.30	L47	3.36	L14	0.80	L97
T 449	000627-05-4	1-NITROBUTANE	-2.27	L47			1.47	L97
T 450	000628-05-7	1-NITROPENTANE	-2.07	L47			2.01	L97
T 451	000098-95-3	NITROBENZENE	-3.01	L14	1.61	L14	1.85	L97
T 452	000099-08-1	3-NITROTOLUENE	-2.84	L48	1.40	L3	2.42	L97
T 453	000099-99-0	4-NITROTOLUENE	-2.68	L4			2.37	L97
T 454	000086-57-7	1-NITRONAPHTHALENE	-4.14	L66			3.19	L97
T 455	000099-65-0	1,3-DINITROBENZENE	-4.96	L4			1.49	L97
T 456	000100-25-4	1,4-DINITROBENZENE	-5.00	L5			1.47	L97
T 457	000619-15-8	2,5-DINITROTOLUENE	-5.07	L94				
T 458	000606-20-2	2,6-DINITROTOLUENE	-5.05	L4			2.10	L97
T 459	000099-35-4	1,3,5-TRINITROBENZENE	-5.42	L4			1.18	L97
T 460	000118-96-7	2,4,6-TRINITROTOLUENE	-5.96	L4			1.60	L97
T 461	000121-82-4	CYCLONITE	-5.70	L4			0.87	L14
T 462	002691-41-0	1,3,5,7-TETRANITRO-1,3,5,7-TETRAZACYCLOOCTANE	-9.58	L4				
T 463	000598-58-3	METHYL NITRATE	-1.75	L29				
T 464	000625-58-1	ETHYL NITRATE	-1.72	L29				
T 465	000627-13-4	1-PROPYL NITRATE	-1.57	L29	3.46	L84		
T 466	000928-45-0	1-BUTYL NITRATE	-1.48	L65	3.04	L84	2.15	L97
T 467	000543-29-3	2-BUTYL NITRATE	-1.25	L65	3.24	L84		
T 468	001002-16-0	1-PENTYL NITRATE	-1.46	L29	2.83	L84		
T 469	021981-48-6	2-PENTYL NITRATE	-1.11	L29	2.84	L84		
T 470	000543-29-3	2-METHYL-PROPYL-1-NITRATE	-1.36	L29				
T 471	000628-96-6	1,2-ETHANE DINITRATE	-3.29	L55	2.29	L55	1.16	L14
T 472	003457-90-7	1,3-PROPYL DINITRATE	-3.51	L55	1.73	L55	1.68	L58
T 473	006423-43-4	1,2-PROPYLENEGLYCOL DINITRATE	-2.90	L52	2.18	L55	1.83	L58
T 474		1,2-BUTYL DINITRATE	-2.71	L55	1.80	L55	2.36	L58
T 475		1,3-BUTYL DINITRATE	-3.15	L55	1.59	L55	2.15	L58
T 476	003457-91-8	1,4-BUTYL DINITRATE	-3.59	L55	1.13	L55	2.17	L58
T 477	006423-45-6	2,3-BUTYL DINITRATE	-2.48	L55	2.03	L55	2.36	L58
T 478		1,2-PENTYL DINITRATE	-2.52	L55	1.38	L55	2.94	L58
T 479		1,4-PENTYL DINITRATE	-2.98	L55	0.97	L55	2.88	L58
T 480		1,5-PENTYL DINITRATE	-3.46	L55	0.74	L55	2.64	L58
T 481		CIS-2,4-PENTYL DINITRATE	-2.73	L55	1.45	L55	2.67	L58
T 482		TR-2,4-PENTYL DINITRATE	-2.55	L55	1.54	L55	2.75	L58
T 483		1,2-HEXYL DINITRATE	-2.38	L55	0.91	L55	3.51	L58
T 484		1,5-HEXYL DINITRATE	-2.83	L55	0.72	L55	2.64	L58
T 485		2,5-HEXYL DINITRATE	-2.89	L55	0.84	L55	3.09	L58
T 486		1,2-OCTYL DINITRATE	-2.11	L55	0.00	L55	4.63	L58
T 487		1,2-DECYL DINITRATE	-1.69	L55	-0.96	L55	5.88	L58
T 488		1,10-DECYL DINITRATE	-3.02	L55	-2.00	L55	5.56	L58
T 489		CIS-1,2-CYCLOHEXYL DINITRATE	-3.50	L55	0.74	L55	2.97	L58
T 490		CIS-1,3-CYCLOHEXYL DINITRATE	-3.93	L55	0.25	L55	2.66	L58
T 491		TR-1,2-CYCLOHEXYL DINITRATE	-3.11	L55	0.75	L55	2.62	L58
T 492		TR-1,3-CYCLOHEXYL DINITRATE	-3.23	L55	0.71	L55	2.89	L58
T 493		TR-1,2-CYCLOHEPTYL DINITRATE	-3.34	L55	0.05	L55	3.41	L58
T 494	000055-63-0	NITROGLYCEROL	-4.70	L4	-0.50	L3	1.62	L97
T 495	000078-11-5	PENTAERYTHRITETRANITRATE	-7.35	L4				
T 496	000088-74-4	O-NITROANILINE	-5.62	L66			1.85	L97

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 497	000099-09-2	M-NITROANILINE	-6.49	L33			1.37	L97
T 498	000100-01-6	P-NITROANILINE	-7.33	L66			1.39	L97
T 499	019406-51-0	4-AMINO-2,6-DINITROTOLUENE	-7.26	L4				
T 500	035572-78-2	2-AMINO-4,6-DINITROTOLUENE	-6.79	L4				
T 501	040487-42-1	PENDIMETHALIN	-4.46	L52			5.18	L97
T 502	000479-45-8	N-METHYL-N,2,4,6-TETRANITROANILINE	-5.57	L4				
T 503	000088-75-5	2-NITROPHENOL	-3.36	L33			1.79	L97
T 504	000554-84-7	3-NITROPHENOL	-7.06	L47			2.00	L97
T 505	000100-02-7	4-NITROPHENOL	-7.77	L66			1.91	L97
T 506	000534-52-1	4,6-DINITRO-O-CRESOL	-4.75	L19			2.13	L97
T 507	016051-48-2	2-NITROOXY-ETHANOL	-5.98	L29				
T 508		2-NITROOXY-1-BUTANOL	-5.15	L50				
T 509	020266-65-3	1-NITROOXY-2-PROPANOL	-5.40	L50				
T 510		1-NITROOXY-2-BUTANOL	-5.15	L50				
T 511		2-NITROOXY-3-BUTANOL	-5.40	L50				
T 512	000059-89-2	N-NITROSOMORPHOLINE	-6.00	L14			-0.44	L97
T 513	001456-28-6	2,6-DIMETHYL-N-NITROSOMORPHOLINE	-4.96	L97			0.32	L97
T 514	000091-23-6	O-NITROANISOLE	-3.79	L94	0.55	L94	1.73	L97
T 515	000100-17-4	P-NITROANISOLE	-4.09	L94			2.03	L97
T 516	000593-53-3	FLUOROMETHANE	-0.16	L87	6.60	L11	0.51	L97
T 517	000075-10-5	DIFLUOROMETHANE	-0.23	L35	6.23	L14	0.20	L97
T 518	000075-46-7	TRIFLUOROMETHANE	0.51	L35	6.70	L3	0.64	L97
T 519	000075-73-0	TETRAFLUOROMETHANE	2.29	L33			1.18	L97
T 520	000359-35-3	1,1,1,2-TETRAFLUROETHANE	0.39	L35				
T 521	000354-33-6	PENTAFLUROETHANE	1.05	L35				
T 522	000076-16-6	HEXAFLUROETHANE	2.92	L66	6.32	L3	2.00	L97
T 523	000431-89-0	HEPTAFLUROPROPANE	1.46	L35				
T 524	000076-19-7	PERFLUROPROPANE	3.14	L97	5.95	L14		
T 525	000074-87-3	CHLOROMETHANE	-0.44	L66	5.76	L3	0.91	L97
T 526	000075-00-3	CHLOROETHANE	-0.34	L35	5.21	L3	1.43	L97
T 527	000540-54-5	1-CHLOROPROPANE	-0.26	L66	4.65	L3	2.04	L97
T 528	000075-29-6	2-CHLOROPROPANE	-0.18	L47	4.85	L3	1.90	L97
T 529	000109-69-3	1-CHLOROBUTANE	-0.17	L14	4.13	L3	2.64	L97
T 530	000078-86-4	2-CHLOROBUTANE	0.00	L47	4.28	L3	2.33	L97
T 531	000543-59-9	1-CHLOROPENTANE	-0.05	L47	3.53	L3		
T 532	000616-20-6	3-CHLOROPENTANE	0.05	L97	3.79	L14		
T 533	000544-10-5	1-CHLOROHEXANE	0.00	L14	3.07	L3		
T 534	000629-06-1	1-CHLOROHEPTANE	0.21	L47			4.15	L97
T 535	000075-09-2	DICHLOROMETHANE	-0.96	L83	4.76	L14	1.25	L97
T 536	000075-34-3	1,1-DICHLOROETHANE	-0.70	L82	4.48	L14	1.79	L97
T 537	000107-06-2	1,2-DICHLOROETHANE	-1.24	L34	4.02	L14	1.47	L97
T 538	000078-87-5	1,2-DICHLOROPROPANE	-0.95	L85	3.85	L3	1.99	L97
T 539	000142-28-9	1,3-DICHLOROPROPANE	-1.39	L47	3.42	L3	2.00	L97
T 540	000110-56-5	1,4-DICHLOROBUTANE	-1.58	L13	2.74	L14		
T 541	000628-76-2	1,5-DICHLOROPENTANE	-1.10	L13	2.18	L14		
T 542	002162-98-3	1,10-DICHLORODECANE	-0.69	L60				
T 543	003922-28-9	1,12-DICHLORODODECANE	-0.58	L60				
T 544	000067-66-3	CHLOROFORM	-0.84	L4	4.42	L3	1.97	L97
T 545	000079-00-5	1,1,2-TRICHLOROETHANE	-1.47	L14	3.49	L14	2.07	L97
T 546	000071-55-6	1,1,1-TRICHLOROETHANE	-0.24	L83	4.21	L3	2.49	L97
T 547	000056-23-5	CARBON TETRACHLORIDE	0.02	L26	4.19	L14	2.83	L97

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 548	000630-20-6	1,1,1,2-TETRACHLOROETHANE	-1.00	L40	3.20	L14	2.62	L97
T 549	000079-34-5	1,1,2,2-TETRACHLOROETHANE	-1.82	L14	2.93	L3	2.62	L97
T 550		1,2,9,10-TETRACHLORODECANE	-2.14	L60				
T 551		1,2,10,11-TETRACHLOROUNDECANE	-2.59	L60				
T 552	000076-01-7	PENTACHLOROETHANE	-1.02	L47	2.78	L3	3.22	L97
T 553	000067-72-1	HEXACHLOROETHANE	-0.47	L34			4.14	L97
T 554	000593-70-4	CHLOROFLUOROMETHANE	-0.57	L66				
T 555	000075-45-6	CHLORODIFLUOROMETHANE	0.06	L35	6.02	L3	1.08	L97
T 556	000075-68-3	1-CHLORO-1,1-DIFLUOROETHANE	-0.19	L14	5.53	L3		
T 557	000075-72-9	CHLOROTRIFLUOROMETHANE	1.66	L35	6.55	L3	1.65	L97
T 558	000075-88-7	1,1,1-TRIFLUORO-2-CHLOROETHANE	0.04	L66				
T 559	000421-04-5	1-CHLORO-1,1,2-TRIFLUOROETHANE	0.04	L35				
T 560	002837-89-0	1-CHLORO-1,2,2,2-TETRAFLUROETHANE	0.59	L35				
T 561	000076-15-3	CHLOROPENTAFLUROETHANE	2.11	L35	5.95	L3		
T 562	001717-00-6	1,1-DICHLORO-1-FLUROETHANE	0.15	L35	5.53	L14		
T 563	000076-14-2	1,2-DICHLOROTETRAFLUROETHANE	1.65	L35	5.34	L3	2.82	L97
T 564	000075-69-4	TRICHLOROFLUOROMETHANE	0.60	L35	5.03	L3	2.53	L97
T 565	000076-12-0	1,1,2,2-TETRACHLORODIFLUOROETHANE	0.60	L97	3.85	L3		
T 566	000058-89-9	GAMMA-HEXACHLOROCYCLOHEXANE	-3.68	L66			3.72	L97
T 567	002385-85-5	MIREX	-1.67	L14			5.28	L97
T 568	000116-14-3	TETRAFLUROETHYLENE	1.41	L87	6.51	L14		
T 569	000116-15-4	PERFLUROPROPENE	1.77	L35	5.87	L3		
T 570	000075-01-4	VINYLCBLORIDE	-0.05	L47	5.59	L3		
T 571	000107-05-1	3-CHLOROPROPYLENE	-0.42	L47	4.69	L3		
T 572	000075-35-4	1,1-DICHLOROETHENE	0.03	L34	4.90	L3	2.13	L97
T 573	000156-59-2	CIS-1,2-DICHLOROETHENE	-0.51	L47	4.44	L3	1.86	L97
T 574	000156-60-5	TRANS-1,2-DICHLOROETHENE	-0.41	L34	4.63	L3	1.86	L97
T 575	010061-01-5	CIS-1,3-DICHLOROPROPENE	-0.95	L14	3.78	L3	1.82	L97
T 576	000079-01-6	TRICHLOROETHENE	-0.33	L63	3.99	L3	2.61	L97
T 577	000127-18-4	TETRACHLOROETHENE	-0.05	L83	3.39	L3	3.40	L97
T 578	000359-10-4	1-CHLORO-2,2-DIFLUOROETHENE	0.39	L35	5.68	L14		
T 579	000057-74-9	CHLORDANE	-2.66	L74			6.00	L97
T 580	000087-68-3	HEXACHLOROBUTADIENE	-0.76	L13	1.47	L14	4.78	L97
T 581	000309-00-2	ALDRIN	-1.69	L14			6.50	L97
T 582	000076-44-8	HEPTACHLOR	-1.92	L66				
T 583	000462-06-6	FLUROBENZENE	-0.47	L6	4.01	L3	2.27	L97
T 584	000098-08-8	α,A,A-TRIFLUOROTOLUENE	-0.18	L33	3.72	L3	3.01	L97
T 585	000108-90-7	CHLOROBENZENE	-0.80	L48	3.20	L14	2.89	L97
T 586	000100-44-7	A-XHAOPOTOAYENE	-1.46	L48	2.22	L3		
T 587	000095-50-1	1,2-DICHLOROBENZENE	-1.07	L48	2.26	L14	3.43	L97
T 588	000106-46-7	1,4-DICHLOROBENZENE	-0.89	L34			3.44	L97
T 589	000541-73-1	1,3-DICHLOROBENZENE	-0.93	L34	2.51	L3	3.53	L97
T 590	000120-82-1	1,2,4-TRICHLOROBENZENE	-1.11	L34	1.73	L3	4.05	L97
T 591	000108-70-3	1,3,5-TRICHLOROBENZENE	-1.10	L66			4.19	L97
T 592	002077-46-5	2,3,6-TRICHLOROTOLUENE	-1.21	L66				
T 593	023749-65-7	2,4,6-TRICHLOROTOLUENE	-1.21	L66				
T 594	000634-66-2	1,2,3,4-TETRACHLOROBENZENE	-1.59	L66			4.64	L97
T 595	000634-90-2	1,2,3,5-TETRACHLOROBENZENE	-1.39	L27			4.66	L97
T 596	000095-94-3	1,2,4,5-TETRACHLOROBENZENE	-1.39	L66			4.60	L97
T 597	000608-93-5	PENTACHLOROBENZENE	-1.62	L66			5.18	L97
T 598	000877-11-2	2,3,4,5,6-PENTACHLOROTOLUENE	-1.50	L66			5.62	L97

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 599	000118-74-1	HEXACHLOROBENZENE	-2.02	L66			5.73	L97
T 600	029082-74-4	OCTACHLOROSTYRENE	-2.27	L66				
T 601	000050-29-3	P,P'-DDT	-3.28	L74			6.91	L97
T 602	000072-55-9	P,P'-DDE	-2.59	L14 / L13			6.96	L97
T 603	002051-60-7	2-CHLOROBIPHENYL	-1.52	L7			4.53	L97
T 604	002051-62-9	4-CHLOROBIPHENYL	-2.01	L43			4.61	L97
T 605	013029-08-8	2,2'-DICHLOROBIPHENYL	-1.85	L52			4.97	L97
T 606	016605-91-7	2,3-DICHLOROBIPHENYL	-2.03	L43				
T 607	033284-50-3	2,4-DICHLOROBIPHENYL	-1.85	L2	-0.77	L10	5.30	L97
T 608	034883-43-7	2,4'-DICHLOROBIPHENYL	-1.89	L2			5.10	L97
T 609	034883-39-1	2,5-DICHLOROBIPHENYL	-1.80	L9	-0.74	L14	5.16	L97
T 610	033146-45-1	2,6-DICHLOROBIPHENYL	-1.90	L2			4.98	L97
T 611	002050-67-1	3,3'-DICHLOROBIPHENYL	-2.02	L9			5.30	L97
T 612	002974-92-7	3,4-DICHLOROBIPHENYL	-2.24	L23			5.29	L97
T 613	002050-68-2	4,4'-DICHLOROBIPHENYL	-2.09	L52			5.58	L97
T 614	038444-86-9	2',3,4-TRICHLOROBIPHENYL	-1.80	L7			5.87	L97
T 615	038444-78-9	2,2',3-TRICHLOROBIPHENYL	-2.09	L43			5.60	L97
T 616	037680-65-2	2,2',5-TRICHLOROBIPHENYL	-1.82	L2			5.60	L97
T 617	038444-73-4	2,2',6-TRICHLOROBIPHENYL	-2.03	L23			5.47	L97
T 618	038444-84-7	2,3,3'-TRICHLOROBIPHENYL	-2.18	L43			5.57	L97
T 619	055712-37-3	2,3',4-TRICHLOROBIPHENYL	-1.84	L2				
T 620	038444-76-7	2,3',6-TRICHLOROBIPHENYL	-1.83	L2				
T 621	038444-85-8	2,3,4'-TRICHLOROBIPHENYL	-2.00	L2			5.42	L97
T 622	058702-45-9	2,3,6-TRICHLOROBIPHENYL	-2.05	L23			5.67	L97
T 623	015862-07-4	2,4,5-TRICHLOROBIPHENYL	-2.09	L23			5.90	L97
T 624	016606-02-3	2,4',5-TRICHLOROBIPHENYL	-2.11	L23			5.79	L97
T 625	035693-92-6	2,4,6-TRICHLOROBIPHENYL	-1.58	L13			5.71	L97
T 626	007012-37-5	2,4,4'-TRICHLOROBIPHENYL	-1.89	L2			5.62	L97
T 627	038444-87-0	3,3',5-TRICHLOROBIPHENYL	-2.16	L23				
T 628	038444-90-5	3,4,4'-TRICHLOROBIPHENYL	-2.39	L23			5.90	L97
T 629	052663-59-9	2,2',3,4-TETRACHLOROBIPHENYL	-2.04	L2			6.11	L97
T 630	041464-39-5	2,2',3,5'-TETRACHLOROBIPHENYL	-2.02	L2			5.81	L97
T 631	070362-45-7	2,2',3,6-TETRACHLOROBIPHENYL	-1.86	L2				
T 632	041464-40-8	2,2',4,5'-TETRACHLOROBIPHENYL	-2.07	L43			6.36	L97
T 633	062796-65-8	2,2',4,6-TETRACHLOROBIPHENYL	-1.51	L7				
T 634	058194-04-7	2,2',4,6'-TETRACHLOROBIPHENYL	-1.51	L43				
T 635	035693-99-3	2,2',5,5'-TETRACHLOROBIPHENYL	-1.99	L13			6.26	L97
T 636	041464-41-9	2,2',5,6'-TETRACHLOROBIPHENYL	-1.91	L7				
T 637	054230-23-7	2,3,4,6-TETRACHLOROBIPHENYL	-2.07	L23				
T 638	033284-54-7	2,3,5,6-TETRACHLOROBIPHENYL	-1.97	L2			5.94	L97
T 639	032690-93-0	2,4,4',5-TETRACHLOROBIPHENYL	-2.39	L23			6.67	L97
T 640	032598-13-3	3,3',4,4'-TETRACHLOROBIPHENYL	-2.47	L2			6.63	L97
T 641	033284-52-5	3,3',5,5'-TETRACHLOROBIPHENYL	-1.99	L2			6.85	L97
T 642	038444-93-8	2,2',3,3'-TETRACHLOROBIPHENYL	-2.39	L23			6.18	L97
T 643	070362-48-0	2,3',4',5-TETRACHLOROBIPHENYL	-2.39	L23			6.18	L97
T 644	052663-62-4	2,2',3,3',4-PENTACHLOROBIPHENYL	-2.09	L7				
T 645	065510-45-4	2,2',3,4,4'-PENTACHLOROBIPHENYL	-2.54	L23			6.61	L97
T 646	041464-51-1	2,2',3',4,5-PENTACHLOROBIPHENYL	-2.57	L23			6.67	L97
T 647	038380-01-7	2,2',4,4',5-PENTACHLOROBIPHENYL	-1.99	L39			7.21	L97
T 648	037680-73-2	2,2',4,5,5'-PENTACHLOROBIPHENYL	-2.34	L7			6.85	L97
T 649	068194-06-9	2,2',4,5,6'-PENTACHLOROBIPHENYL	-2.43	L23				

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 650	056558-16-8	2,2',4,6,6'-PENTACHLOROBIPHENYL	-1.44	L9				
T 651	031508-00-6	2,3',4,4',5-PENTACHLOROBIPHENYL	-2.50	L23			7.12	L97
T 652	038380-02-8	2,2',3,4,5'-PENTACHLOROBIPHENYL	-2.52	L43			6.85	L97
T 653	038380-07-3	2,2',3,3',4,4'-HEXACHLOROBIPHENYL	-2.91	L9			7.32	L97
T 654	052687-58-8	2,2',3,3',4,5'-HEXACHLOROBIPHENYL	-2.82	L23			7.39	L97
T 655	038380-05-1	2,2',3,3',4,6'-HEXACHLOROBIPHENYL	-2.75	L23				
T 656	052704-70-8	2,2',3,3',5,6-HEXACHLOROBIPHENYL	-2.70	L23				
T 657	052744-13-5	2,2',3,3',5,6'-HEXACHLOROBIPHENYL	-2.64	L23			7.15	L97
T 658	038411-22-2	2,2',3,3',6,6'-HEXACHLOROBIPHENYL	-2.44	L23			7.12	L97
T 659	038380-04-0	2,2',3,4',5',6'-HEXACHLOROBIPHENYL	-1.91	L7				
T 660	068194-14-9	2,2',3,4,5',6-HEXACHLOROBIPHENYL	-1.61	L7			6.79	L97
T 661	052712-04-6	2,2',3,4,5,5'-HEXACHLOROBIPHENYL	-3.03	L23			7.19	L97
T 662	051908-16-8	2,2',3,4',5,5'-HEXACHLOROBIPHENYL	-2.99	L23			7.12	L97
T 663	068194-13-8	2,2',3,4',5,6-HEXACHLOROBIPHENYL	-2.68	L23			6.93	L97
T 664	035065-27-1	2,2',4,4',5,5'-HEXACHLOROBIPHENYL	-2.30	L13			7.44	L97
T 665	033979-03-2	2,2',4,4',6,6'-HEXACHLOROBIPHENYL	-2.33	L13			7.29	L97
T 666	055215-18-4	2,2',3,3',4,5-HEXACHLOROBIPHENYL	-2.93	L23			7.32	L97
T 667	052663-63-5	2,2',3,5,5',6-HEXACHLOROBIPHENYL	-2.62	L23			6.85	L97
T 668	035065-30-6	2,2',3,3',4,4',5-HEPTACHLOROBIPHENYL	-3.43	L23				
T 669	052663-74-8	2,2',3,3',4,5,5'-HEPTACHLOROBIPHENYL	-3.27	L23				
T 670	068194-16-1	2,2',3,3',4,5,6-HEPTACHLOROBIPHENYL	-3.24	L23				
T 671	038411-25-5	2,2',3,3',4,5,6'-HEPTACHLOROBIPHENYL	-3.24	L23				
T 672	052663-67-9	2,2',3,3',5,5',6-HEPTACHLOROBIPHENYL	-3.03	L23				
T 673	052663-64-6	2,2',3,3',5,6,6'-HEPTACHLOROBIPHENYL	-3.01	L23				
T 674	035065-29-3	2,2',3,4,4',5,5'-HEPTACHLOROBIPHENYL	-3.39	L23				
T 675	052663-69-1	2,2',3,4,4',5',6-HEPTACHLOROBIPHENYL	-2.52	L7				
T 676	052712-05-6	2,2',3,4,5,5',6-HEPTACHLOROBIPHENYL	-3.18	L23			7.93	L97
T 677	035694-08-7	2,2',3,3',4,4',5,5'-OCTACHLOROBIPHENYL	-3.39	L23			8.68	L97
T 678	052663-75-9	2,2',3,3',4,5,5',6'-OCTACHLOROBIPHENYL	-3.39	L23				
T 679	040186-71-8	2,2',3,3',4,5',6,6'-OCTACHLOROBIPHENYL	-3.16	L23				
T 680	002136-99-4	2,2',3,3',5,5',6,6'-OCTACHLOROBIPHENYL	-3.13	L23			7.73	L97
T 681	042740-50-1	2,2',3,3',4,4',5,6'-OCTACHLOROBIPHENYL	-3.39	L23				
T 682	068194-17-1	2,2',3,3',4,5,5',6-OCTACHLOROBIPHENYL	-3.24	L23				
T 683	000090-13-1	1-CHLORONAPHTHALENE	-1.84	L14	0.59	L14	4.10	L97
T 684	000091-58-7	2-CHLORONAPHTHALENE	-1.88	L14			4.14	L97
T 685	000075-89-8	2,2,2-TRIFLUOROETHANOL	-3.16	L14	3.90	L3	0.41	L97
T 686	000076-37-9	2,2,3,3-TETRAFLUOROPROPANOL	-3.59	L14				
T 687	000422-05-9	PENTAFLUORO-1-PROPANOL	-3.04	L14			1.23	L97
T 688	000920-66-1	1,1,1,3,3,3-HEXAFLUOROPROPAN-2-OL	-2.76	L14	4.35	L3	1.66	L97
T 689	000302-17-0	CHLORAL HYDRATE	-6.93	L66			0.99	L97
T 690	000367-12-4	2-FLUOROPHENOL	-3.88	L33			1.71	L97
T 691	000095-57-8	2-CHLOROPHENOL	-3.34	L33	2.49	L14	2.15	L97
T 692	000108-43-0	3-CHLOROPHENOL	-4.85	L33			2.50	L97
T 693	000106-48-9	4-CHLOROPHENOL	-5.16	L33			2.39	L97
T 694	000120-83-2	2,4-DICHLOROPHENOL	-3.74	L5			3.06	L97
T 695	028523-86-6	SEVOFLURANE	0.44	L97	4.42	L3		
T 696	000111-44-4	2,2'-DICHLOROETHYLETHER	-2.29	L66	2.18	L3	1.29	L97
T 697	000108-60-1	DICHLOROISOPROPYLETHER	-2.21	L86	2.00	L89		
T 698	026675-46-7	ISOFLURANE	0.07	L47	4.64	L15	2.06	L97
T 699	056885-28-0	1,1,1,2-TETRAFLUORO-2- CHLOROPERFLUOROMETHOXYMETHANE	0.27	L97				
T 700	033018-78-9	1,1,1-TRIFLUORO-2- CHLOROPERFLUOROMETHOXYMETHANE	1.37	L97				

Tabelle A-1 Trainingsdatensatz

Nr	CAS-NR	Name	log H'	Ref	log P _v	Ref	log K _{ow}	Ref
T 701	000076-38-0	METHOXYFLUORANE	-0.82	L47	3.61	L3	2.21	L97
T 702	000406-90-6	FLUOROXENE	-0.10	L47				
T 703	050375-10-5	2,3,6-TRICHLOROANISOLE	-1.93	L66			3.64	L97
T 704	007005-72-3	4-CHLOROPHENYL PHENYL ETHER	-2.04	L43	-0.44	L14		
T 705	039001-02-0	OCTACHLORODIBENZOFURAN	-4.11	L39				
T 706	033857-26-0	2,7-DICHLORODIBENZO-P-DIOXIN	-2.62	L66			6.38	L97
T 707	039227-58-2	1,2,4-TRICHLORODIBENZO-P-DIOXIN	-2.83	L66			7.47	L97
T 708	030746-58-8	1,2,3,4-TETRACHLORODIBENZO-P-DIOXIN	-3.09	L66			7.18	L97
T 709	067028-18-6	1,2,3,7-TETRACHLORODIBENZO-P-DIOXIN	-3.51	L14				
T 710	040321-76-4	1,2,3,7,8-PENTACHLORODIBENZO-P-DIOXIN	-3.97	L39			6.64	L97
T 711	003268-87-9	OCTACHLORODIBENZO-P-DIOXIN	-3.56	L39				
T 712	000078-95-5	CHLOROACETONE	-3.16	L68	3.35	L3		
T 713	000144-49-0	FLUOROACETIC ACID	-6.30	L54				
T 714	000381-73-7	DIFLUOROACETIC ACID	-5.87	L54				
T 715	000076-05-1	TRIFLUOROACETIC ACID	-5.34	L54	3.10	L3		
T 716	000079-11-8	CHLOROACETIC ACID	-6.42	L54			0.22	L97
T 717	000079-43-6	DICHLOROACETIC ACID	-6.47	L54	1.47	L3	0.92	L97
T 718	000076-03-9	TRICHLOROACETIC ACID	-6.26	L53			1.33	L97
T 719	000076-04-0	CHLORODIFLUOROACETIC ACID	-5.78	L54				
T 720	014437-17-3	CHLORFENPROP-METHYL	-4.04	L97	-1.00	L3		
T 721	027355-22-2	TETRACHLOROPHTHALIDE	-4.66	L67			3.20	L97
T 722	001929-73-3	2,4-D BUTOXYETHYL ESTER	-5.39	L19	-3.22	L14		
T 723	000555-77-1	N,N,N-TRIS(2-CHLOROETHYL)AMINE	-3.15	L4	0.15	L3		
T 724	000095-51-2	O-CHLOROANILINE	-3.60	L33	1.43	L14	1.90	L97
T 725	000108-42-9	M-CHLOROANILINE	-4.27	L33	0.86	L14	1.88	L97
T 726	000106-47-8	P-CHLOROANILINE	-4.33	L33			1.88	L97
T 727	001194-65-6	2,6-DICHLOROBENZONITRILE	-3.19	L64			2.74	L97
T 728	001897-45-6	CHLOROTHANONIL	-5.10	L67			2.90	L97
T 729	000626-60-8	3-CHLOROPYRIDINE	-2.94	L33			1.33	L97
T 730	000109-09-1	2-CHLOROPYRIDINE	-3.22	L33	2.46	L14	1.22	L97
T 731	000122-34-9	SIMAZINE	-7.60	L75			2.18	L97
T 732	050563-36-5	DIMETHACHLOR	-7.00	L97			2.17	L97
T 733	015972-60-8	ALACHLOR	-5.59	L73			3.52	L97
T 734	051218-45-2	METOLACHLOR	-6.47	L49	-2.38	L89	3.13	L97
T 735	023184-66-9	BUTACHLOR	-5.68	L14	-3.62	L89	4.50	L97
T 736	071422-67-8	CHLORFLUAZURON	-10.13	L89			5.80	L97
T 737	027541-88-4	QUINONAMIDE	-6.30	L97				
T 738	052756-25-9	FLAMPROP-METHYL	-6.74	L97				
T 739	000076-06-2	TRICHLORONITROMETHANE	-1.08	L67	3.50	L14	2.09	L97
T 740	000088-73-3	2-CHLORO-1-NITROBENZENE	-2.74	L14			2.52	L97
T 741	000121-73-3	M-CHLORONITROBENZENE	-3.26	L66			2.47	L97
T 742	000100-00-5	P-CHLORONITROBENZENE	-2.65	L14			2.39	L97
T 743	000089-61-2	2,5-DICHLORONITROBENZENE	-3.31	L66			3.09	L97
T 744	000099-54-7	3,4-DICHLORONITROBENZENE	-3.48	L66				
T 745	000082-68-8	QUINTOZENE	-3.82	L67			5.10	L97
T 746	000097-00-7	1-CHLORO-2,4-DINITROBENZENE	-4.08	L4				
T 747	000121-87-9	2-CHLORO-4-NITROANILINE	-6.41	L66				
T 748	001582-09-8	TRIFLURALIN	-2.37	L49			5.07	L97
T 749	042576-02-3	BIFENOX	-5.34	L89			4.47	L97

Tabelle A-1 Trainingsdatensatz / Ende

Tabelle A-1 Trainingsdatensatz

Tabelle A-2 Vorhersagedatensatz mit experimentellen Stoffdaten (H' = Henrykonstante, P_V = Dampfdruck [Pa], K_{OW} = Oktanol/Wasser-Verteilungskoeffizient). Die Literaturangaben (Ref) der experimentellen Größen sind in Kapitel A.2 Seite A-21 ff. gelistet.

Nr	CAS-NR	Name	log H'	Ref_1	log P_V	Ref_2	log K_{OW}	Ref_3
V 1	000074-82-8	METHANE	1.43	L66			1.09	L97
V 2	000074-84-0	ETHANE	1.34	L47			1.81	L97
V 3	000589-34-4	3-METHYLHEXANE	1.99	L33	3.91	L14		
V 4	026635-64-3	2-METHYLHEPTANE	2.18	L6	3.44	L14		
V 5	000565-75-3	2,3,4-TRIMETHYLPENTANE	1.88	L33	3.56	L14		
V 6	002216-33-3	3-METHYLOCTANE	2.70	L14 / L14	2.92	L14		
V 7	000463-82-1	2,2-DIMETHYLPROPANE	1.84	L47			3.11	L97
V 8	000590-35-2	2,2-DIMETHYLPENTANE	2.11	L56	4.14	L14		
V 9	000096-37-7	METHYLCYCLOPENTANE	1.17	L56	4.26	L14	3.37	L97
V 10	001678-91-7	ETHYLCYCLOHEXANE	1.11	L32 / L31	3.23	L14		
V 11	000589-90-2	1,4-DIMETHYLCYCLOHEXANE	1.51	L14 / L37	3.43	L14		
V 12	002070-96-4	PROPYLCYCLOPENTANE	1.56	L56	3.22	L14		
V 13	000091-17-8	DECALIN	0.68	L66	2.02	L14		
V 14	000109-67-1	1-PENTENE	1.21	L56	4.93	L3		
V 15	000109-68-2	TRANS-2-PENTENE	0.98	L66	4.82	L14		
V 16	000563-45-1	3-METHYL-1-BUTENE	1.34	L33	5.08	L14		
V 17	000763-29-1	2-METHYL-1-PENTENE	1.08	L33	4.41	L14		
V 18	000691-37-2	4-METHYL-1-PENTENE	1.40	L66	4.56	L14		
V 19	000106-99-0	1,3-BUTADIENE	0.41	L66			1.99	L97
V 20	000513-81-5	2,3-DIMETHYL-1,3-BUTADIENE	0.32	L14 / L14	4.31	L14		
V 21	000138-86-3	LIMONENE	0.44	L3 / L57	2.44	L3		
V 22	000107-00-6	1-BUTYNE	-0.12	L47				
V 23	000627-19-0	1-PENTYNE	0.01	L47	4.76	L14	1.98	L97
V 24	000100-41-4	ETHYLBENZENE	-0.52	L48	3.11	L14	3.15	L97
V 25	000095-93-2	1,2,4,5-TETRAMETHYLBENZENE	0.01	L56			4.00	L97
V 26	001077-16-3	N-HEXYLBENZENE	-0.10	L56	1.21	L14	5.52	L97
V 27	000098-82-8	CUMENE	-0.28	L66	2.78	L14	3.66	L97
V 28	000538-93-2	ISOBUTYLBENZENE	0.12	L56	2.41	L14		
V 29	000098-06-6	T-BUTYLBENZENE	-0.28	L56	2.45	L79	4.11	L97
V 30	002049-95-8	T-AMYLBEZENE	-0.13	L66				
V 31	000119-64-2	1,2,3,4-TETRAHYDRONAPHTHALENE	-1.12	L66	1.69	L14	3.49	L97
V 32	000098-83-9	A-METHYLBENZENE	-0.79	L3 / L14	2.59	L3	3.47	L97
V 33	000101-81-5	1,1'-METHYLENEBISBENZENE	-1.60	L14 / L38	0.04	L14	4.14	L97
V 34	025640-78-2	2-ISOPROPYLBIPHENYL	-2.05	L14 / L14	-1.18	L14	5.20	L14
V 35	000581-42-0	2,6-DIMETHYLNAPHTHALENE	-1.93	L33			4.31	L97
V 36	002027-17-0	2-ISOPROPYLNAPHTHALENE	-1.26	L14 / L37	-0.16	L14		
V 37	038640-62-9	DIISOPROPYLNAPHTHALENE	-1.28	L14 / L14	-1.18	L14	4.90	L14
V 38	000206-44-0	FLUORANTHENE	-3.57	L66			5.16	L97
V 39	000120-12-7	ANTHRACENE	-3.10	L52	-3.00	L46	4.45	L97
V 40	000085-01-8	PHENANTHRENE	-2.84	L46	-1.50	L3	4.47	L97
V 41	000056-55-3	BENZ(A)ANTHRACENE	-3.73	L24 / L25	-4.55	L24	5.79	L97
V 42	000111-27-3	1-HEXANOL	-3.16	L61	2.09	L14	2.03	L97
V 43	000143-08-8	NONANOL	-2.85	L47	0.48	L14	4.26	L97
V 44	036653-82-4	HEXADECANOL	-2.39	L80 / L13	-3.26	L80		
V 45	000137-32-6	2-METHYL-1-BUTANOL	-3.24	L14	2.62	L14		
V 46	000123-51-3	ISOPENTANOL	-3.24	L14	2.60	L3	1.16	L97

Tabelle A-2 Testdatensatz

Nr	CAS-NR	Name	log H'	Ref_1	log P_v	Ref_2	log K_{ow}	Ref_3
V 47	000097-95-0	2-ETHYL-1-BUTANOL	-3.07	L14 / L16	2.31	L14		
V 48	000104-76-7	2-ETHYL-1-HEXANOL	-2.73	L15 / L8	1.55	L15		
V 49	000107-18-6	ALLYL ALCOHOL	-3.69	L14			0.17	L97
V 50	000078-92-2	2-BUTANOL	-3.39	L47	3.39	L14	0.61	L97
V 51	000584-02-1	3-PENTANOL	-3.19	L47	3.07	L14	1.21	L97
V 52	000626-93-7	2-HEXANOL	-3.00	L14 / L14	2.52	L14	1.76	L97
V 53	000623-37-0	3-HEXANOL	-2.98	L47	2.81	L14	1.65	L97
V 54	000123-96-6	2-OCTANOL	-2.87	L14 / L13	1.51	L14	2.90	L97
V 55	000565-67-3	2-METHYL-3-PENTANOL	-2.85	L47				
V 56	001724-39-6	CYCLODODECANOL	-3.92	L66	0.10	L94	3.75	L97
V 57	000075-85-4	2-METHYL-2-BUTANOL	-3.25	L47	3.32	L3	0.89	L97
V 58	000590-36-3	2-METHYL-2-PENTANOL	-2.88	L47	3.06	L14		
V 59	000078-70-6	LINALOOL	-3.06	L66	1.33	L14	2.97	L14
V 60	000057-55-6	1,2-PROPANEDIOL	-6.30	L65			-0.92	L97
V 61	000122-97-4	3-PHENYLPROPANOL	-5.08	L33	1.80	L92	1.88	L97
V 62	000095-48-7	O-CRESOL	-4.30	L47	1.51	L15	1.95	L97
V 63	000108-39-4	M-CRESOL	-4.46	L66	1.26	L14	1.96	L97
V 64	000105-67-9	2,4-DIMETHYLPHENOL	-4.09	L43	1.11	L14	2.30	L97
V 65	000090-00-6	O-ETHYLPHENOL	-4.14	L14 / L30	1.31	L14	2.47	L97
V 66	000620-17-7	M-ETHYLPHENOL	-4.59	L33			2.40	L97
V 67	000120-80-9	1,2-BENZENEDIOL	-5.55	L3 / L17	1.45	L3	0.88	L97
V 68	000123-31-9	1,4-BENZENEDIOL	-7.87	L3 / L13	-1.62	L3	0.59	L97
V 69	000557-17-5	METHYL PROPYL ETHER	-1.22	L66	4.79	L14	1.21	L97
V 70	000628-32-0	ETHYL PROPYL ETHER	-1.33	L66	4.38	L14		
V 71	000111-43-3	DI(N-PROPYL) ETHER	-1.02	L48	3.92	L14	2.03	L97
V 72	000628-81-9	ETHYL BUTYL ETHER	-1.22	L14 / L42	3.94	L14	2.03	L97
V 73	000637-92-3	ETHYL T-BUTYL ETHER	-1.24	L14 / L14	4.22	L14		
V 74	000109-92-2	ETHYL VINYL ETHER	-0.71	L14 / L13	4.83	L14	1.04	L97
V 75	000105-57-7	1,1-DIETHOXYETHANE	-2.31	L3 / L14	3.65	L3	0.84	L97
V 76	000629-14-1	1,2-DIETHOXYETHANE	-2.59	L14	3.20	L3	0.66	L97
V 77	000103-73-1	ETHOXYBENZENE	-1.63	L47	2.32	L14	2.51	L97
V 78	000103-50-4	DIBENZYL ETHER	-3.57	L14 / L37	-0.87	L14	3.31	L97
V 79	000075-56-9	PROPYLENE OXIDE	-2.38	L3 / L3	4.85	L3	0.03	L97
V 80	000142-68-7	TETRAHYDROPYRAN	-2.38	L14 / L14	3.98	L14	0.95	L97
V 81	001003-38-9	2,5-DIMETHYLTETRAHYDROFURAN	-2.14	L33			1.22	L97
V 82	000646-06-0	1,3-DIOXALANE	-3.00	L14			-0.37	L97
V 83	000108-62-3	METALDEHYDE	-2.67	93 / 93	0.82	93		
V 84	000262-12-4	DIBENZO-P-DIOXINE	-2.31	L14 / L22	-1.26	L14	4.38	L97
V 85	000106-92-3	ALLYL GLYCIDYL ETHER	-3.70	L3 / L37	2.78	L3		
V 86	003031-73-0	METHYLHYDROPEROXIDE	-3.73	L66				
V 87	000123-72-8	BUTYRALDEHYDE	-2.33	L61	4.17	L14	0.88	L97
V 88	000110-62-3	1-PENTANAL	-2.20	L66	3.50	L3		
V 89	000066-25-1	HEXANAL	-2.06	L61	3.18	L14	1.78	L97
V 90	000112-31-2	DECANAL	-1.30	L45	1.14	L14		
V 91	000078-84-2	ISOBUTYRALDEHYDE	-2.10	L47	4.36	L3		
V 92	000590-86-3	3-METHYL-1-BUTANAL	-1.94	L14 / L3	3.82	L14		
V 93	002363-89-5	2-OCTENAL	-1.99	L65				
V 94	000104-87-0	P-METHYLBENZALDEHYDE	-3.06	L14 / L44	1.52	L14		
V 95	000078-93-3	2-BUTANONE	-2.63	L14	4.10	L14	0.29	L97
V 96	000123-19-3	4-HEPTANONE	-2.14	L47	2.95	L15		
V 97	000112-12-9	2-UNDECANONE	-1.59	L61			4.09	L97

Tabelle A-2 Testdatensatz

Nr	CAS-NR	Name	log H'	Ref_1	log P_v	Ref_2	log K_{ow}	Ref_3
V 98	000927-49-1	6-UNDECANONE	-2.04	L14 / L14	0.82	L14		
V 99	000565-61-7	3-METHYLPENTAN-2-ONE	-2.52	L14 / L14	3.19	L14		
V 100	000565-69-5	2-METHYL-3-PENTANONE	-2.20	L14 / L14	3.38	L14		
V 101	000110-12-3	5-METHYL-2-HEXANONE	-2.15	L14 / L3	2.84	L14	1.88	L97
V 102	000075-97-8	3,3-DIMETHYL-2-BUTANONE	-2.28	L33	3.62	L14	1.20	L97
V 103	000120-92-3	CYCLOPENTANONE	-3.32	L69				
V 104	000765-43-5	METHYL CYCLOPROPYL KETONE	-3.38	L33			0.49	L97
V 105	000591-24-2	3-METHYLCYCLOHEXANONE	-3.22	L14 / L3	2.30	L14		
V 106	000076-22-2	CAMPHOR	-2.90	L15 / L13	1.50	L15		
V 107	000141-79-7	MESITYL OXIDE	-2.61	L15 / L13	3.25	L15		
V 108	000093-55-0	PROPIOPHENONE	-3.16	L14 / L14	1.40	L14	2.19	L97
V 109	000109-52-4	N-VALERIC ACID	-4.71	L43	1.52	L14	1.39	L97
V 110	000112-05-0	PELARGONIC ACID	-4.33	L14 / L14	-0.81	L14		
V 111	000149-57-5	2-ETHYLHEXANOIC ACID	-3.93	L14 / L14	0.60	L14	2.64	L97
V 112	000079-10-7	ACRYLIC ACID	-4.88	L19			0.35	L97
V 113	003724-65-0	TRANS-CROTONIC ACID	-4.81	L15 / L3	1.50	L15	0.72	L97
V 114	000079-41-4	METHACRYLIC ACID	-4.29	L14 / L14	2.11	L14	0.93	L97
V 115	000141-82-2	MALONIC ACID	-8.22	L14 / L3	-0.70	L14	-0.81	L97
V 116	000109-94-4	ETHYL FORMATE	-1.88	L47	4.51	L14		
V 117	000554-12-1	METHYL PROPIONATE	-2.16	L65	4.05	L14	0.82	L97
V 118	000110-74-7	PROPYL FORMATE	-1.82	L33	4.03	L3	0.83	L97
V 119	000623-42-7	METHYLBUTYRATE	-2.08	L61	3.61	L3	1.29	L97
V 120	000106-36-5	PROPYL PROPIONATE	-1.80	L65	3.25	L3		
V 121	000637-78-5	ISOPROPYL PROPIONATE	-1.63	L47	3.48	L14		
V 122	000110-19-0	ISOBUTYL ACETATE	-1.73	L47	3.41	L3	1.78	L97
V 123	000628-63-7	N-AMYL ACETATE	-1.84	L47	2.67	L14		
V 124	000123-66-0	ETHYL HEXANOATE	-1.64	L33				
V 125	000106-30-9	ETHYL HEPTANOATE	-1.70	L47	1.96	L14		
V 126	020487-40-5	T-BUTYL PROPIONATE	-1.27	L97				
V 127	000096-33-3	METHYL ACRYLATE	-2.09	L3 / L14	4.06	L3	0.80	L97
V 128	000140-88-5	ETHYL ACRYLATE	-1.86	L14 / L14	3.71	L14	1.32	L97
V 129	000141-32-2	BUTYL ACRYLATE	-1.72	L14 / L14	2.86	L14	2.36	L97
V 130	000097-63-2	ETHYLMETHACRYLATE	-1.78	L3 / L37	3.29	L3	1.94	L97
V 131	000097-88-1	BUTYLMETHACRYLATE	-1.53	L3 / L36	2.61	L3	2.88	L97
V 132	000111-55-7	ETHYLENE GLYCOL DIACETATE	-4.78	L3 / L13	1.70	L3		
V 133	000105-53-3	MALONIC ACID DIETHYLESTER	-3.98	L14 / L42	1.55	L14	0.96	L97
V 134	000624-48-6	METHYL MALEATE	-4.54	L14 / L14	1.60	L14	0.22	L97
V 135	000102-76-1	GLYCERYL TRIACETATE	-6.30	L14 / L14	-0.48	L14	0.25	L97
V 136	000140-11-4	BENZYL ACETATE	-3.34	L14 / L14	1.37	L14	1.96	L97
V 137	000131-11-3	DIMETHYL PHTHALATE	-4.70	L99 / L14	0.00	L99	1.56	L97
V 138	000117-81-7	DI(2-ETHYLHEXYL)PHTHALATE	-3.22	L43	-3.07	L15	7.45	L97
V 139	027554-26-3	DIISOCTYL PHTHALATE	-2.88	L14 / L14	-3.13	L14		
V 140	000084-61-7	DICYCLOHEXYL PHTHALATE	-5.41	L15 / L14	-3.94	L15		
V 141	000085-68-7	BENZYL BUTYL PHTHALATE	-5.86	L15 / L13	-2.96	L15	4.91	L97
V 142	000616-38-6	DIMETHYL CARBONATE	-2.73	L92 / L92	3.85	L92		
V 143	000085-44-9	PHTHALIC ANHYDRIDE	-6.21	L94 / L14	-1.20	L94	1.60	L97
V 144	000110-80-5	2-ETHOXYETHANOL	-4.91	L33			-0.32	L97
V 145	005131-66-8	1-BUTOXY-2-PROPANOL	-3.87	L3 / L15	2.20	L3		
V 146	000090-05-1	2-METHOXYPHENOL	-4.35	L70	1.24	L14	1.32	L97
V 147	000093-51-6	4-METHYL-2-METHOXYPHENOL	-4.28	L71	0.91	L77		
V 148	000100-83-4	3-HYDROXYBENZALDEHYDE	-6.99	L66			1.38	L97

Tabelle A-2 Testdatensatz

Nr	CAS-NR	Name	log H'	Ref_1	log P_v	Ref_2	log K_{ow}	Ref_3
V 149	005836-29-3	ENDOX	-9.43	L3 / L3	-7.90	L3		
V 150	000119-36-8	METHYL SALICYLATE	-3.39	L14 / L13	0.66	L14	2.34	L97
V 151	000107-70-0	4-METHOXY-4-METHYLPENTAN-2-ONE	-4.02	L3 / L15	2.62	L15		
V 152	000111-15-9	2-ETHOXYETHANOL, ACETATE	-4.10	L3 / L37	2.57	L3		
V 153	000112-07-2	(2-BUTOXYETHYL)ACETATE	-3.66	L14 / L94	1.70	L14		
V 154	027138-31-4	DIPROPYLENE GLYCOL DIBENZOATE	-6.25	L14 / L14	-4.21	L14		
V 155	010453-86-8	RESMETHRIN	-5.44	L3 / L99	-6.00	L3	6.14	L97
V 156	001118-84-9	ALLYL ACETOACETATE	-4.49	L92 / L92	1.45	L92		
V 157	000111-92-2	DIBUTYLAMINE	-2.38	L47	2.54	L14	2.83	L97
V 158	000100-61-8	N-METHYLANILINE	-3.44	L33	1.78	L14	1.66	L97
V 159	000110-59-8	VALERONITRILE	-2.55	L48	2.99	L14	1.12	L97
V 160	000109-02-4	N-METHYLMORPHOLINE	-4.64	L33	3.57	L3	-0.33	L97
V 161	000055-18-5	N-NITROSODIETHYLAMINE	-3.83	L66			0.47	L97
V 162	000079-24-3	NITROETHANE	-2.71	L14	3.43	L3	0.18	L97
V 163	000088-72-2	2-NITROTOLUENE	-2.63	L47	1.34	L90	2.30	L97
V 164	000121-14-2	2,4-DINITROTOLUENE	-5.66	L66			1.98	L97
V 165	001712-64-7	2-PROPYL NITRATE	-1.41	L29	3.63	L84		
V 166		1,6-HEXYL DINITRATE	-3.57	L55	0.05	L55	3.19	L58
V 167		1,7-HEPTYL DINITRATE	-3.45	L55	-0.46	L55	3.81	L58
V 168		1,8-OCTYL DINITRATE	-3.29	L55	-1.05	L55	4.40	L58
V 169	000051-28-5	2,4-DINITROPHENOL	-6.19	L4			1.67	L97
V 170	020266-74-4	2-NITROOXY-1-PROPANOL	-5.24	L29				
V 171	006745-71-7	NITROOXYACETONE	-4.39	L29				
V 172	002278-22-0	PEROXYACETYLNITRATE	-2.09	L65				
V 173	000075-37-6	1,1-DIFLUOROETHANE	-0.08	L14	5.77	L3	0.75	L97
V 174	000625-29-6	2-CHLOROPENTANE	0.05	L47	3.81	L14		
V 175	000541-33-3	1,1-DICHLOROBUTANE	-0.51	L66	3.48	L14		
V 176	000096-18-4	1,2,3-TRICHLOROPROPANE	-1.85	L14	2.59	L3		
V 177	000075-71-8	DICHLORODIFLUOROMETHANE	1.13	L35	5.82	L3	2.16	L97
V 178	000076-13-1	1,1,2-TRICHLOROTRIFLUOROETHANE	1.12	L34	4.64	L3	3.16	L97
V 179	000319-84-6	ALPHA-HEXACHLOROCYCLOHEXANE	-3.50	L28			3.72	L97
V 180	010061-02-6	TRANS-1,3-DICHLOROPROPENE	-0.83	L14	3.80	L3	1.82	L97
V 181	000095-49-8	2-CHLOROTOLUENE	-0.84	L14	2.63	L14	3.42	L97
V 182	000087-61-6	1,2,3-TRICHLOROBENZENE	-1.29	L14			4.14	L97
V 183	006639-30-1	2,4,5-TRICHLOROTOLUENE	-1.21	L13			4.56	L97
V 184	000072-54-8	DDD	-3.57	L66			6.22	L97
V 185	025569-80-6	2,3'-DICHLOROBIPHENYL	-1.87	L2	-0.75	L10	5.02	L97
V 186	038444-81-4	2,3',5'-TRICHLOROBIPHENYL	-1.88	L9			5.76	L97
V 187	002437-79-8	2,2',4,4'-TETRACHLOROBIPHENYL	-2.11	L23			6.29	L97
V 188	015968-05-5	2,2',6,6'-TETRACHLOROBIPHENYL	-2.09	L23			5.94	L97
V 189	073575-53-8	2,3',4,5'-TETRACHLOROBIPHENYL	-2.39	L23				
V 190	032598-12-2	2,4,4',6'-TETRACHLOROBIPHENYL	-1.77	L43				
V 191	068194-12-7	2,3',4,5,5'-PENTACHLOROBIPHENYL	-2.64	L23				
V 192	038379-99-6	2,2',3,5',6'-PENTACHLOROBIPHENYL	-2.31	L23			6.55	L97
V 193	035065-28-2	2,2',3,4,4',5'-HEXACHLOROBIPHENYL	-2.27	L39				
V 194	052663-78-2	2,2',3,3',4,4',5,6-OCTACHLOROBIPHENYL	-3.35	L23				
V 195	000374-01-6	1,1,1-TRIFLUORO-2-PROPANOL	-3.05	L14			0.71	L97
V 196	000371-41-5	4-FLUOROPHENOL	-4.54	L33			1.77	L97
V 197	000059-50-7	3-METHYL-4-CHLOROPHENOL	-4.98	L33	0.82	L14	3.10	L97
V 198	000072-43-5	METHOXYCHLOR	-5.08	L66			5.08	L97
V 199	070648-26-9	1,2,3,4,7,8-HEXACHLORODIBENZOFURAN	-3.23	L39				

Tabelle A-2 Testdatensatz

Nr	CAS-NR	Name	log H'	Ref_1	log P_v	Ref_2	log K_{ow}	Ref_3
V 200	035822-46-9	1,2,3,4,6,7,8-HEPTACHLORODIBENZO-P-DIOXIN	-3.29	L39				
V 201	000421-50-1	1,1,1-TRIFLUOROACETONE	-3.53	L68				
V 202	001912-24-9	ATRAZINE	-6.54	L4			2.61	L97
V 203	022212-55-1	BENZOYLPROP-ETHYL	-7.46	L97				
V 204	001746-81-2	MONOLINURON	-6.53	L89			2.30	L97

Tabelle A-2 Testdatensatz / Ende

A.2 Quellen der experimentellen Stoffeigenschaften

- [L1] T. Suzuki Development of an Automatic Estimation System for Both the Partition Coefficient and Aqueous Solubility. *J. Comput.-Aided Mol. Design* **1991**, *5*, 149-166.
- [L2] F. M. Dunnivant, A. W. Eizerman, P. C. Jurs, M. N. Hasan Quantitative Structure-Property Relationships for Aqueous Solubilities and Henry's Law Constants of Polychlorinated Biphenyls. *Environ. Sci. Technol.* **1992**, *26*, 1567-1573.
- [L3] G. Sorbe *Sicherheitstechnische Kenndaten*; Ecomed: Landsberg/Lech (FRG), 1990-1996.
- [L4] G. Rippen, Ed.; *Handbuch Umweltchemikalien*; Ecomed: Landsberg/Lech (FRG), 1989-1994.
- [L5] Beratergremium für umweltrelevante Altstoffe, Ed.; *BUA-Stoffberichte*; VCH Verlagsgesellschaft: Weinheim (FRG), 1985-1989.
- [L6] J. T. Hoff, D. Mackay, R. Gillham, W. Y. Shiu Partitioning of Organic Chemicals At the Air-Water Interface in Environmental Systems. *Environ. Sci. Technol.* **1993**, *27*, 2174-2180.
- [L7] W. Y. Shiu, D. Mackay A Critical Review of Aqueous Solubilities, Vapor Pressures, Henry's Law Constants, and Octanol-Water Partition Coefficients of the Polychlorinated Biphenyls. *J. Phys. Chem. Ref. Data* **1986**, *15*, 911-929.
- [L8] S. H. Yalkowsky, S. C. Valvani Solubility and Partitioning I: Solubility of Nonelectrolytes in Water. *J. Pharm. Sci.* **1980**, *69*, 912-922.
- [L9] F. M. Dunnivant, A. W. Eizerman Aqueous Solubility and Henry's Law Constant Data for PCB Congeners for Evaluation of Quantitative Structure-Property Relationships (QSPRs). *Chemosphere* **1988**, *17*, 525-541.
- [L10] R. C. Fischer, R. Wittlinger, K. Ballschmiter Retention-Index Based Vapor Pressure Estimation for Polychlorobiphenyl (PCB) By Gas Chromatography. *Fresenius J. Anal. Chem.* **1992**, *342*, 421-425.
- [L11] D. R. Stull Vapor Pressure of Pure Substances. *Ind. Eng. Chem.* **1947**, *39*, 517-550.
- [L12] A. I. Majeed (1983): . Ph.D. Dissertation, Oclahoma State University, Stillwater (USA, OK).
- [L13] S. H. Yalkowsky, R.-M. Dannenfelser (1990): AQUASOL dATABASE of aqueous solubility, 5th edition. University of Arizona, Tucson (AZ, USA).
- [L14] J. A. Beauman, P. H. Howard (1995): Physprop database. Syracuse Research Corporation. Syracuse (NY, USA).
- [L15] M. L. Richardson, S. Gangolli, Eds.; *The Dictionary of Substances and their Effects*; Royal Society of Chemistry: Cambridge, 1992-1994.
- [L16] T. M. Nelson, P. C. Jurs Prediction of Aqueous Solubility of Organic Compounds. *J. Chem. Inform. Comput. Sci.* **1994**, *34*, 601-609.
- [L17] P. Isnard, S. Lambert Aqueous Solubility and N-Octanol/Water Partition Coefficient Correlations. *Chemosphere* **1989**, *18*, 1837-1853.
- [L18] P. H. Howard (1995): Chemfate database v. 1.3. Syracuse Research Corporation. Syracuse (NY, USA).
- [L19] P. H. Howard *Handbook of Environmental Fate and Exposure Data for Organic Chemicals*; Lewis Publishers: Chelsea (MI,USA), 1990.
- [L20] D. Mackay, W. Y. Shiu A Critical Review of Henry's Law Constants for Chemicals of Environmental Interest. *J. Phys. Chem. Ref. Data* **1981**, *10*, 1175-1199.
- [L21] S. H. Yalkowsky, S. C. Valvani, D. Mackay Estimation of the Aqueous Solubility of Some Aromatic Compounds. *Residue Rev.* **1983**, *85*, 43-55.
- [L22] W. Y. Shiu, W. Doucette, F. A. P. C. Gobas, A. Andren, D. Mackay Physical-Chemical Properties of Chlorinated Dibenzo-*P*-Dioxins. *Environ. Sci. Technol.* **1988**, *22*, 651-658.

- [L23] S. Brunner, E. Hornung, H. Santl, E. Wolff, O. G. Piringer, J. Altschuh, R. Brüggemann Henry's Law Constants for Polychlorinated Biphenyl: Experimental Determination and Structure-Property-Relationships. *Environ. Sci. Technol.* **1990**, *24*, 1751-1754.
- [L24] W. J. Sonnefeld, W. H. Zoller, W. E. May Dynamic Coupled-Column Liquid Chromatographic Determination of Ambient Temperature Vapor Pressures of Polynuclear Aromatic Hydrocarbons. *Anal. Chem.* **1983**, *55*, 275-280.
- [L25] S. Banerjee, S. H. Yalkowsky, S. C. Valvani Water Solubility and Octanol/Water Partition Coefficients of Organics. Limitations of the Solubility-Partition Coefficient Correlation. *Environ. Sci. Technol.* **1980**, *14*, 1227-1229.
- [L26] J. Dewulf, D. Drijvers, H. van Langenhove Measurement of Henry's Law Constant As Function of Temperature and Salinity for the Low Temperature Range. *Atmos. Environ.* **1995**, *29*, 323-331.
- [L27] T. E. M. ten Hulscher, L. E. van der Velde, W. A. Bruggeman Temperature Dependence of Henry's Law Constants for Selected Chlorobenzenes, Polychlorinated Biphenyls and Polycyclic Aromatic Hydrocarbons. *Environ. Toxicol. Chem.* **1992**, *11*, 1595-1603.
- [L28] J. R. Kucklick, D. A. Hinckley, T. F. Bidleman Determination of Henry's Law Constants for Hexachlorocyclohexanes in Distilled Water and Artificial Sea Water As a Function of Temperature. *Mar. Chem.* **1991**, *34*, 197-209.
- [L29] J. Kames, U. Schurath Alkyl Nitrates and Bifunctional Nitrates of Atmospheric Interest: Henry's Law Constants and Their Temperature Dependence. *J. Atmos. Chem.* **1992**, *15*, 79-95.
- [L30] D. Varhaníková, W.-Y. Shiu, D. Mackay Aqueous Solubilities of Alkylphenols and Methoxyphenols At 25 °C. *J. Chem. Eng. Data* **1995**, *40*, 448-451.
- [L31] P. B. Myrdal, A. M. Manka, S. H. Yalkowsky AQUAFAC 3: Aqueous Functional Group Activity Coefficients; Application to the Estimation of Aqueous Solubility. *Chemosphere* **1995**, *30*, 1619-1637.
- [L32] I. Mokbel, E. Rauzy, H. Loiseleur, C. Berro, J. Jose Vapor Pressures of 12 Alkylcyclohexanes, Cyclopentane, Butylcyclopentane and Trans-Decahydronaphthalene Down to 0.5 Pa. Experimental Results, Correlation and Prediction By an Equation of State. *Fluid Phase Equilibria* **1995**, *108*, 103-120.
- [L33] M. H. Abraham, J. Andonian-Haftvan, G. S. Whiting, A. Leo, R. S. Taft Hydrogen Bonding. Part 34. The Factors That Influence the Solubility of Gases and Vapors in Water At 298 K, and a New Method for Its Determination. *J. Chem. Soc. Perkin Trans.* **1994**, 1777-1791.
- [L34] R. A. Ashworth, G. B. Howe, M. E. Mullins, T. N. Rogers Air-Water Partitioning Coefficients of Organics in Dilute Aqueous Solutions. *J. Hazard. Materials* **1988**, *18*, 25-36.
- [L35] S. Maaßen *Experimentelle Bestimmung und Korrelierung von Verteilungskoeffizienten in verdünnten Lösungen*; Shaker Verlag: Aachen, 1996.
- [L36] W. M. Meylan, P. H. Howard (1994): Upgrade of PCGEMS water solubility estimation method. SRC-TR-94-009, Syracuse Research Corporation Environmental Science Center, Syracuse (NY, USA).
- [L37] W. M. Meylan, P. H. Howard (1994): Validation of water solubility estimation methods using Log K_{OW} for application in PCGEMS & EPI. SRC-TR-94-024, Syracuse Research Corporation Environmental Science Center, Syracuse (NY, USA).
- [L38] J. Nouwen, B. Hansen Correlation Analysis Between Watersolubility, Octanol-Water Partition Coefficient and Melting Point Based On Clustering. *Quant. Struct.-Act. Relat.* **1996**, *15*, 17-30.
- [L39] M. S. McLachlan Bioaccumulation of Hydrophobic Chemicals in Agricultural Food Chains. *Environ. Sci. Technol.* **1996**, *30*, 252-259.
- [L40] G. Tse, H. Orbey, S. I. Sandler Infinite Dilution Activity Coefficients and Henry's Law Coefficients. *Environ. Sci. Technol.* **1992**, *92*, 2017-2022.
- [L41] D. S. Jayasinghe, B. J. Brownawell, H. Chen, J. C. Westall Determination of Henry's Constants of Organic Compounds of Low Volatility: Methylanilines in Methanol-Water. *Environ. Sci. Technol.* **1992**, *26*, 2275-2281.
- [L42] R. M. Stephenson Mutual Solubilities: Water-Ketones, Water-Ethers and Water-Gasoline-Alcohols. *J. Chem. Eng. Data* **1992**, *37*, 80-95.

- [L43] D. Mackay, W. Y. Shiu, K.-C. Ma *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*; Lewis Publishers: Chelsea, MI, 1992-1995.
- [L44] R. M. Stephenson Mutual Solubility of Water and Aldehydes. *J. Chem. Eng. Data* **1993**, 38, 630-633.
- [L45] J. Staudinger, P. V. Roberts A Critical Review of Henry's Law Constant for Environmental Applications. *Crit. Rev. Environ. Sci. Technol.* **1996**, 26, 205-297.
- [L46] W.-Y. Shiu, D. Mackay Henry's Law Constants of Selected Aromatic Hydrocarbons, Alcohols, and Ketones. *J. Chem. Eng. Data* **1997**, 42, 27-30.
- [L47] M. H. Abraham, G. S. Whiting, R. Fuchs, E. J. Chambers Thermodynamics of Solute Transfer from Water to Hexadecane. *J. Chem. Soc. Perkin Trans.* **1990**, 2, 291-300.
- [L48] J. Li, P. W. Carr Measurement of Water-Hexadecane Partition Coefficients By Headspace Gas Chromatography and Calculation of Limiting Activity Coefficients in Water. *Anal. Chem.* **1993**, 65, 1443-1450.
- [L49] C. P. Rice, S. M. Chernyak, L. L. McConnell Henry's Law Constants for Pesticides Measured As a Function of Temperature and Salinity. *J. Agric. Food Chem.* **1997**, 45, 2291-2298.
- [L50] P. B. Shepson, E. Mackay, K. Muthuramu Henry's Law Constants and Removal Processes for Several Atmospheric β -Hydroxy Alkyl Nitrates. *Environ. Sci. Technol.* **1996**, 30, 3618-3623.
- [L51] P. G.-J. de Maagd, D. T. E. M. ten Hulscher, H. van den Heuvel, A. Opperhuizen, D. T. H. M. Sum Physicochemical Properties of Polycyclic Aromatic Hydrocarbons: Aqueous Solubilities, *N*-Octanol/Water Partition Coefficients, and Henry's Law Constants. *Environ. Toxicol. Chem.* **1998**, 17, 251-257.
- [L52] N. J. Fendinger, D. E. Glotfelty Henry's Law Constants for Selected Pesticides, PAHs and PCBs. *Environ. Toxicol. Chem.* **1990**, 9, 731-735.
- [L53] D. J. Bowden, S. L. Clegg, P. Brimblecombe The Henry's Law Constant of Trichloroacetic Acid. *Water. Air. Soil. Pollut.* **1998**, 101, 197-215.
- [L54] D. J. Bowden, S. L. Clegg, P. Brimblecombe The Henry's Law Constants of the Haloacetic Acids. *J. Atmos. Chem.* **1998**, 29, 85-107.
- [L55] R. G. Fischer, K. Ballschmiter Prediction of the Environmental Distribution of Alkyl Dinitrates - Chromatographic Determination of Vapor Pressure P^0 , Water Solubility S_{H_2O} , Gas-Water Partition Coefficient K_{GW} (Henry's Law Constant) and Octanol-Water Partition Coefficient K_{OW} . *Fresenius J. Anal. Chem.* **1998**, 360, 769-776.
- [L56] L. Eastcott, W. Y. Shiu, D. Mackay Environmentally Relevant Physical-Chemical Properties of Hydrocarbons: A Review of Data and Development of Simple Correlations. *Oil & Chemical Pollution* **1988**, 4, 191-216.
- [L57] J. Li, E. M. Perdue, S. G. Pavlostathis, R. Araujo Physicochemical Properties of Selected Monoterpenes. *Environ. Int.* **1998**, 24, 353-358.
- [L58] R. G. Fischer, K. Ballschmiter Determination of Vapor Pressure, Water Solubility, Gas-Water Partition Coefficient P_{GW} , Henry's Law Constant, and Octanol-Water Partition Coefficient P_{OW} of 26 Alkyl Dinitrates. *Chemosphere* **1998**, 36, 2891-2901.
- [L59] J. M. Allen, W. X. Balcavage, B. R. Ramachandran, A. L. Shroul Determination of Henry's Law Constants By Equilibrium Partitioning in a Closed System Using a New in Situ Optical Absorbance Method. *Environ. Toxicol. Chem.* **1998**, 17, 1216-1221.
- [L60] K. G. Drouillard, G. T. Tomy, D. C. G. Muir, K. J. Friesen Volatility of Chlorinated *N*-Alkanes (C-10-C₁₂): Vapor Pressures and Henry's Law Constants. *Environ. Toxicol. Chem.* **1998**, 17, 1252-1260.
- [L61] R. G. Buttery, L. C. Ling, D. G. Guadagni Volatilities of Aldehydes, Ketones and Esters in Dilute Water Solution. *J. Agric. Food Chem.* **1969**, 17, 385-389.
- [L62] I. Mokbel, E. Rauzy, J. P. Meille, J. Jose Low Vapor Pressures of 12 Aromatic Hydrocarbons. Experimental and Calculated Data Using a Group Contribution Method. *Fluid Phase Equilibria* **1998**, 147, 271-284.
- [L63] L. H. Turner, Y. C. Chiew, R. C. Ahlert, D. S. Kosson Measuring Vapor-Liquid Equilibrium for Aqueous-Organic Systems: Review and a New Technique. *AIChE Journal* **1996**, 42, 1772-1788.

- [L64] K. Schoene, J. Steinhanses Determination of Henry's Law Constants By Headspace Gas Chromatography. *Fresenius Z. Anal. Chem.* **1985**, 321, 538-543.
- [L65] E. A. Betterton Henry's Law Constants of Soluble and Moderately Soluble Organic Gases: Effects On Aqueous Phase Chemistry. In *Gaseous Pollutants. Characterization and Cycling*; J. O. Nriagu, Ed.; John Wiley & Sons: New York, 1992.
- [L66] S. Sixt (1998): Methoden zur Abschätzung umweltrelevanter physikalisch-chemischer und ökotoxikologischer Eigenschaften organischer Substanzen aus der Molekülstruktur. Dr. Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg.
- [L67] K. Kawamoto, K. Urano Parametrs for Predicting Fate of Organochlorine Pesticides in the Environment (1). Octanol-Water and Air-Water Partition Coefficients. *Chemosphere* **1989**, 18, 1987-1996.
- [L68] E. A. Betterton The Partioning of Ketones Between the Gas and Aqueous Phases. *Atmos. Environ.* **1991**, 25A, 1473-1477.
- [L69] S. B. Hawthorne, R. E. Sievers, R. M. Barkley Organic Emissions from Shale Oil Wastewaters and Their Implications for Air Quality. *Environ. Sci. Technol.* **1985**, 19, 992-997.
- [L70] J. C. Sagebiel, J. N. Seiber, J. E. Woodrow Comparison of Headspace and Gas-Stripping Methods for Determining the Henry's Law Constant (H) for Organic Compounds of Low to Intermediate H. *Chemosphere* **1992**, 25, 1763-1768.
- [L71] K.-U. Goss Predicting the Enrichment of Organic Compounds in Fog Caused By Adsorption On the Water Surface. *Atmos. Environ.* **1994**, 28, 3513-3517.
- [L72] E. A. Betterton, M. R. Hoffmann Henry's Law Constants of Some Environmentally Important Aldehydes. *Environ. Sci. Technol.* **1988**, 22, 1415-1418.
- [L73] N. J. Fendinger, D. E. Glotfelty A Laboratory Method for the Experimental Determination of Air-Water Henry's Law Constants for Several Pesticides. *Environ. Sci. Technol.* **1988**, 22, 1289-1293.
- [L74] N. J. Fendinger, D. E. Glotfelty, H. P. Freeman Comparison of Two Experimental Techniques for Determining Air/Water Henry's Law Constants. *Environ. Sci. Technol.* **1989**, 23, 1528-1531.
- [L75] D. E. Glotfelty, J. N. Seiber, L. A. Liljedahl Pesticides in Fog. *Nature* **1987**, 325, 603-605.
- [L76] I. Mokbel, T. Guetachew, J. Jose Vapor Pressures and Sublimation Pressures of 14 Polycyclic Aromatic Hydrocarbons (C11-C18) At Pressures in the Range from 0.5 Pa to 30 KPa. *ELDATA, Int. Electron. J. Phys. Chem. Data* **1995**, 1, 167.
- [L77] J. C. Sagebiel, J. N. Seiber Studies On Occurence and Distribution of Wood Smoke Marker Compoundss in Foggy Atmospheres. *Environ. Toxicol. Chem.* **1993**, 12, 813-822.
- [L78] W.-H. Xie, W.-Y. Shiu, D. Mackay A Review of the Effect of Salts On the Solubility of Organic Compounds in Seawater. *Mar. Environ. Res.* **1997**, 44, 429-444.
- [L79] S. P. Verevkin Thermochemical Properties of Branched Alkylsubstituted Benzenes. *J. Chem. Thermodynamics* **1998**, 30, 1029-1040.
- [L80] B. Koutek, M. Hoskovec, P. Vrkocová, K. Konecny, L. Felzl Gas Chromatographic Determination of Vapour Pressures of Pheromone-Like Compounds. II. Alcohols. *J. Chromatogr. A.* **1994**, 679, 307-317.
- [L81] B. Koutek, M. Hoskovec, P. Vrkocová, K. Konecny, L. Felzl, J. Vrkoc Gas Chromatographic Determination of Vapour Pressures of Pheromone-Like Compounds. III. Aldehydes. *J. Chromatogr. A.* **1996**, 719, 391-400.
- [L82] J. Dewulf, H. van Langenhove, S. Graré Sediment/Water and Octanol/Water Equilibrium Partitioning of Volatile Organic Compounds: Temperature Dependence in the 2-25°C Range. *Wat. Res.* **1999**, 10, 2424-2436.
- [L83] P.-C. Chiang, C.-H. Hung, J. C. Mar, E. E. Chang Henry's Constants and Mass Transfer Coefficients of Halogenated Organic Pollutants in an Air Stripping Packed Column. *Water. Sci. Technol.* **1998**, 38, 287-294.
- [L84] K. Hauff, R. G. Fischer, K. Ballschmiter Determination of C₁-C₅ Alkyl Nitrates in Rain, Snow, White Frost, Lake, and Tap Water By a Combined Codistillation Head-Space Gas Chromatography Technique. Determination of Henry's Law Constants By Head-Space GC. *Chemosphere* **1998**, 37, 2599-2615.

Referenzen der experimentellen Stoffeigenschaften

- [L85] J. Dewulf, H. Van Langenhove, P. Everaert Determination of Henry's Law Coefficients By Combination of the Equilibrium Partitioning in Closed Systems and Solid-Phase Microextraction Techniques. *J. Chromatogr. A* **1999**, *830*, 353-363.
- [L86] R. Sander (1999): Compilation of Henry's Law Constants for inorganic and organic species of potential importance in environmental chemistry (Version 3). <http://www.mpch-mainz.mpg.de/~sander/res/henry.html>.
- [L87] E. Wilhelm, R. Battino, R. J. Wilcock Low-Pressure Solubility of Gases in Liquid Water. *Chem. Rev.* **1977**, *77*, 219-262.
- [L88] R. Bone, P. Cullis, R. Wolfenden Solvent Effects On Equilibria of Addition of Nucleophiles to Acetaldehyde and the Hydrophilic Character of Diols. *J. Am. Chem. Soc.* **1983**, *105*, 1339-1343.
- [L89] C. D. S. Tomlin, Ed.; *The Pesticide Manual*; British Crop Protection Council: Farnham, Surrey, UK, 1997.
- [L90] J. Altschuh, R. Brüggemann, H. Santl, G. Eichinger, O. G. Piringer Henry's Law Constant for a Diverse Set of Organic Chemicals: Experimental Determination and Comparison of Estimation Methods. *Chemosphere* **1999**, *39*, 1871-1887.
- [L91] A. Klamt, V. Jonas, T. Bürger, J. C. W. Lohrenz Refinement an Parametrization of COSMO-RS *J. Phys. Chem. A* **1998**, *102*, 5074-5085
- [L92] Acros Organics Catalog of Fine Chemicals: <http://www.fisher1.com/fb/itv?24..f97.1....1>
- [L93] Industrieverband Agrar (ed.): Wirkstoffe in Pflanzenschutz- und Schädlingsbekämpfungsmitteln physikalisch-chemische und toxikologische Daten. BLV Verl., München 1990.
- [L94] Umweltbundesamt (UBA), B.- und L. Institut für Wasser- (eds.): Dokumentation wassergefährdender Stoffe. Datenblattsammlung zur Verwaltungsvorschrift wassergefährdender Stoffe (VwVwS) nach Par. 19g WHG. S. Hirzel Wissenschaftliche Verlagsgesellschaft, Stuttgart 1996 ff. Loseblattsammlung
- [L95] R.A. Baker (ed.): Organic Substances and Sediments in Water 2. Processes and analytical. CRC, Boca Raton 1991.
- [L96] US EPA Fact Sheets on New Active Ingredients: www.epa.gov/opprd001/factsheets
- [L97] D. A. Leo (1999): MedChem99 database. Daylight Chemical Information Systems. Irvine (CA, USA).
- [L98] D. A. Leo (2000): MedChem00 database. Daylight Chemical Information Systems. Irvine (CA, USA).
- [L99] C. D. S. Tomlin, Ed.; *The Pesticide Manual*; British Crop Protection Council: Farnham, Surrey, UK, 1997.
- [L100] R. J. L. Andon, J. D. Cox, E. F. G. Herington Phase Relationships in the Pyridine Series. Part V. The Thermodynamic Properties of Dilute Solutions of Pyridine Bases in Water At 25° and 40°. *J. Chem. Soc.* **1954**, 3188-3196.

A.3 Fischöl/Wasser-Verteilungskoeffizient-Messungen

Material

Fischöl: Medizinal-Dorschlebertran, der dankenswerterweise von der Fa. Henry Lamotte GmbH, Bremen zu Forschungszwecken zur Verfügung gestellt wurde.

Wasser: bidestilliert

Testchemikalien: Merck, Darmstadt

Durchführung der Fischöl/Wasser-Verteilungsgleichgewichtsmessungen

Es wurden je Testsubstanz 3 Glasrührflaschen ($V = 600$ mL) eingesetzt, die während des Versuchs mit einem Thermostat F 20 (Julabo) auf $25 \pm 0,5$ °C temperiert wurden. Nach vorsichtiger Überschichtung der gegenseitig vorgesättigten Phasen wurden diese mit einem glasummantelten Magnetrührstab und einem Rührwerk MR 300D (Heidoph) mit 150 U/min gerührt.

Die Verteilungsgleichgewichte wurden für jede Substanz bei drei verschiedenen Konzentrationsniveaus bestimmt. Die Ansätze waren wie folgt:

- Flasche 1: 500 mL Wasser und 20 mL Fischölphase (mit 0,01 mol/L Testsubstanz)
- Flasche 2: 500 mL Wasser und 20 mL Fischölphase (mit 0,005 mol/L Testsubstanz)
- Flasche 3: 500 mL Wasser und 20 mL Fischölphase (mit 0,001 mol/L Testsubstanz).

Die Equilibrierzeit betrug 6 Tage. Nach ca. halbstündiger Absetzzeit wurden beide Phasen beprobt. Die pH-Wert-Messung wurde an 100 mL Wasserphase in einem 100 mL–Erlenmeyer-Kolben mit einem pH-Meter CG 841 (Schott) vorgenommen. Die Proben aus der Fischöl-Phase wurden für die Analyse mit Hexan verdünnt.

Analysebedingungen

Für die Bestimmung der Gleichgewichtskonzentrationen von *Tetrachlormethan in Fischöl- und Wasserphase* wurde ein Gaschromatograph HP 5890 II mit Elektroneneinfangdetektor (=ECD) unter den nachfolgenden GC-Bedingungen eingesetzt:

- Träger- & Make-Up-Gas: N_2
- Injektortemperatur: 250 °C;
- Säule: OPTIMA delta-6 (60 m \times 0,25 mm ; $d_f = 0,25$ μ m)
- Temperaturprogramm: 40 °C (4 min) \rightarrow 10 °/min \rightarrow 100 °C \rightarrow 20 °/min \rightarrow 280 °C (5 min)
- Detektortemperatur: 300 °C

Jeweils 1 μ L Fischölphase (mit Hexan verdünnt) wurden per Autosampler bei einem Split von 39:1 injiziert. Von den Wasserphasen-Proben wurden jeweils 5 mL in ein Vial überführt und bei Raumtemperatur der Festphasen-Mikroextraktion (=SPME) unterzogen. Eingesetzt wurde dafür eine mit 75 μ m Carboxen/PDMS beschichtete SPME-Faser (Supelco). Die Faser wurde unter Rühren bei 1000 U/min für 15 min in der Probe exponiert. Durch thermische Desorption der Faser (2 min) im Injektor wurde die extrahierte Analytenmenge quantitativ in den GC überführt, wobei mit einem Split 20:1 gearbeitet wurde.

Für die Quantifizierung der *aromatischen Testsubstanzen in der Fischölphase* wurde mit einem

Gaschromatograph HP 5890 II mit massenselektivem Detektor (=MSD) 5971 im Single-Ion-Modus gearbeitet, wobei die folgenden Ionen-Massenzahlen detektiert wurden:

<i>Phenol</i>	94 , 66
<i>2-Chlorphenol</i>	128 , 100, 92
<i>2-Nitrophenol</i>	139 , 109, 81

<i>2,4-Dichlorphenol</i>	163, 63
<i>Anilin</i>	93, 66
<i>3-Nitroanisol</i>	153, 77

Die GC-Bedingungen waren dabei wie folgt:

- Trägergas: He
- Injektortemperatur: 250 °C
- Splitless-Zeit: 1 min
- Split: 25:1
- Säule: CP-SIL 8CB-MS (50 m × 0,32 mm ; $d_f = 0,12 \mu\text{m}$)
- Temperaturprogramm: 50 °C (1 min) → 8 °/min → 150 °C → 10 °/min → 280 °C (10 min)
- Detektortemperatur: 280 °C
- MP-Spannung: + 200 V
- Ausblendzeit: 4,5 min

Per Autosampler wurden jeweils 1 μL Fischölphase (mit Hexan verdünnt) injiziert.

Zur Anreicherung der *aromatische Testsubstanzen aus der Wasserphase* kam wiederum die SPME, jetzt eine mit 85 μm Polyacrylat beschichtete Faser (Supelco), zum Einsatz (Probevolumen und Extraktionsbedingungen wie oben). Die Analyse erfolgte an einem Chrompack GC 9001 mit Flammenionisationsdetektor (=FID) unter folgenden Bedingungen:

- Trägergas: N_2
- Injektortemperatur: 250 °C
- Splitless-Zeit: 2 min
- Split: 25:1
- Säule: CP-SIL 8CB (50 m × 0,32 mm ; $d_f = 0,25 \mu\text{m}$)
- Temperaturprogramm: 50 °C (2 min) → 10 °/min → 280 °C (10 min)
- Detektortemperatur: 280 °C

Kalibrierung

Für GC-Analyse der Fischölphase wurde je Substanz eine 10-Punkt-Kalibrierung mit Standardlösungen (Matrix: Hexan) durchgeführt.

Die SPME/GC-Prozedur der Wasserphase wurde je Substanz mit 5 verschiedenen Standardlösungen (Matrix: fischölgesättigtes Wasser) kalibriert.

A.4 Tabellen zu den temperaturabhängigen Berechnungen in Kapitel 4.2

Tabelle A-3: Henrykonstanten H , Dampfdrücke P_V [Pa], und Wasserlöslichkeiten S_W [mol L⁻¹] von **Propan** inkl. Standardfehler (SE) und mittlerer absolute Abweichung (MAF). Alle Größen sind in logarithmischen Einheiten angegeben. AM1: COSMO-RS/AM1, DFT: COSMO-RS/B88-PW91, Lee: Dampfdruckberechnung nach Lee und Kesler, UNIFAC: KOW-UNIFAC nach Wienke und Gmehling.

T/K	H'	H'^{AM1}	H'^{DFT}	$H'^{PV/SW}$	P_V	P_V^{AM1}	P_V^{DFT}	P_V^{Lee}	S_W	S_W^{AM1}	S_W^{DFT}	S_W^{UNIFAC}
273.15	1.04	0.82	1.57	0.31	5.65	5.76	5.99	5.09		-2.21	-2.92	-1.58
278.15	1.13	0.86	1.60	0.44		5.82	6.05	5.17	-2.46	-2.24	-2.96	-1.63
283.15	1.22	0.89	1.63	0.53		5.88	6.12	5.24	-2.61	-2.27	-3.00	-1.66
288.15	1.30	0.93	1.65	0.65		5.94	6.18	5.32		-2.31	-3.04	-1.71
293.15	1.14	0.96	1.68	0.75	5.94	6.00	6.24	5.39	-2.76	-2.34	-3.07	-1.75
298.15	1.44	0.99	1.71	0.86		6.06	6.29	5.46	-2.82	-2.37	-3.10	-1.79
303.15	1.49	1.02	1.73	0.95	6.03	6.12	6.35	5.52	-2.89	-2.41	-3.13	-1.83
308.15	1.54	1.05	1.75	1.05		6.17	6.40	5.59		-2.44	-3.16	-1.87
313.15	1.58	1.08	1.77	1.13	6.11	6.22	6.45	5.65		-2.46	-3.19	-1.90
318.15	1.62	1.10	1.80	1.23		6.27	6.50	5.71		-2.49	-3.22	-1.94
323.15	1.65	1.13	1.81	1.29	6.24	6.31	6.55	5.76	-3.08	-2.51	-3.25	-1.96
328.15	1.67	1.15	1.83	1.38		6.36	6.59	5.82		-2.54	-3.27	-2.00
333.15	1.69	1.18	1.85	1.46	6.31	6.40	6.64	5.87	-3.21	-2.56	-3.30	-2.03
338.15	1.71	1.20	1.87	1.52		6.45	6.68	5.92		-2.59	-3.32	-2.05
343.15	1.72	1.22	1.88	1.59		6.49	6.72	5.97		-2.6	-3.34	-2.08
348.15	1.73	1.24	1.90	1.67		6.53	6.76	6.02		-2.62	-3.36	-2.11
353.15		1.26	1.91	1.74	6.48	6.57	6.80	6.07		-2.64	-3.38	-2.14
358.15		1.28	1.93	1.80		6.61	6.84	6.11		-2.66	-3.40	-2.16
363.15		1.29	1.94	1.87		6.65	6.88	6.16		-2.68	-3.42	-2.19
368.15		1.31	1.95	1.92	6.63	6.68	6.91	6.20		-2.69	-3.44	-2.21
SE		0.44	0.30	0.47		0.09	0.32	0.48		0.47	0.31	1.03
MAF		-0.43	0.27	0.43		0.08	0.32	-0.48		0.45	-0.28	1.03

A-4 Henrykonstante, Dampfdruck [Pa] und Wasserlöslichkeit [mol L⁻¹] von **Benzaldehyd** mit Standardfehler (SE) und mittlerer absoluter Abweichung (MAF). Alle Größen sind in logarithmischen Einheiten angegeben. *Werte in Klammern sind aus experimentellen Dampfdrücken und Wasserlöslichkeiten berechnet. AM1: COSMO-RS/AM1, DFT: COSMO-RS/B88-PW91, Lee: Dampfdruckberechnung nach Lee und Kesler, UNIFAC: KOW-UNIFAC nach Wienke und Gmehling

T/K	H'	H'^{AM1}	H'^{DFT}	$H'^{PV/SW}$	P_V	P_V^{AM1}	P_V^{DFT}	P_V^{Lee}	S_W	S_W^{AM1}	S_W^{DFT}	S_W^{UNIFAC}
273.15		-3.58	-4.03	-2.56		2.39	1.32	1.57	-1.11	-0.41	-1.00	-2.22
278.15	-3.62	-3.42	-3.86	-2.42		2.54	1.50	1.75		-0.41	-1.01	-2.19
283.15	-3.46	-3.28	-3.69	-2.29		2.69	1.67	1.92		-0.40	-1.01	-2.16
288.15	-3.30	-3.14	-3.54	-2.16		2.84	1.84	2.09		-0.39	-1.01	-2.13
293.15	-3.13	-3.00	-3.38	-2.04	1.91	2.98	1.99	2.24	-1.17	-0.39	-1.01	-2.10
298.15	-2.96	-2.88	-3.24	-1.92		3.11	2.15	2.40		-0.39	-1.01	-2.08
303.15	(-3.01)	-2.75	-3.10	-1.81	2.23	3.24	2.29	2.54	-1.16	-0.38	-1.01	-2.05
308.15	-2.74	-2.63	-2.96	-1.70		3.37	2.44	2.68		-0.38	-1.01	-2.03
313.15		-2.52	-2.83	-1.60		3.49	2.57	2.82	-1.13	-0.37	-1.01	-2.00
318.15	-2.52	-2.41	-2.71	-1.50		3.61	2.71	2.95		-0.36	-1.01	-1.98
323.15	(-2.49)	-2.30	-2.59	-1.40	2.83	3.72	2.84	3.07	-1.12	-0.36	-1.01	-1.96
328.15		-2.19	-2.47	-1.31	3.00	3.82	2.96	3.19		-0.37	-1.01	-1.93
333.15		-2.09	-2.36	-1.22		3.92	3.08	3.31	-1.06	-0.36	-1.01	-1.91
338.15		-2.00	-2.25	-1.14	3.31	4.03	3.20	3.42		-0.35	-1.00	-1.89
343.15		-1.90	-2.14	-1.06		4.13	3.31	3.53	-1.02	-0.35	-1.00	-1.87
348.15		-1.81			4.22	3.42	3.63			-0.34	-1.00	-1.85
353.15		-1.73			4.32	3.53	3.73			-0.33	-1.00	-1.83
358.15		-1.65			4.41	3.63	3.83			-0.33	-1.00	-1.81
363.15		-1.56			4.50	3.73	3.93			-0.32	-0.99	-1.79
368.15		-1.48			4.58	3.83	4.02			-0.31	-0.99	-1.77
SE		0.14	0.24	1.10		0.95	0.07	0.25		0.74	0.11	0.91
MAF		0.14	-0.24	1.10		0.08	0.00	0.24		0.74	0.06	-0.90

Tabellen der temperaturabhängigen Berechnungen

Tabelle A-5 Henrykonstante, Dampfdruck [Pa] und Wasserlöslichkeit [mol L⁻¹] von **Acetophenon** mit Standardfehler (SE) und mittlerem absolutem Fehler (MAF). Alle Größen sind in logarithmischen Einheiten angegeben. *Werte in Klammern sind aus experimentellen Dampfdrücken und Wasserlöslichkeiten berechnet. AM1: COSMO-RS/AM1, DFT: COSMO-RS/B88-PW91, Lee: Dampfdruckberechnung nach Lee und Kesler, UNIFAC: KOW-UNIFAC nach Wienke und Gmehling.

T[K]	H'	H'^{AM1}	H'^{DFT}	$H'^{PV/SW}$	P_V	P_V^{AM1}	P_V^{DFT}	P_V^{Lee}	S_w	S_w^{AM1}	S_w^{DFT}	S_w^{UNIFAC}
273.15		-4.03	-4.43	-4.23		1.75	0.63	0.94		-0.59	-1.30	-1.19
278.15		-3.86	-4.24	-4.03		1.92	0.82	1.13		-0.58	-1.30	-1.21
283.15		-3.7	-4.06	-3.84		2.09	1.01	1.32		-0.58	-1.30	-1.22
288.15	-3.61	-3.55	-3.88	-3.66		2.25	1.19	1.49		-0.58	-1.31	-1.23
293.15	(-3.51)*	-3.4	-3.71	-3.48	1.61	2.40	1.37	1.66	-1.27	-0.58	-1.31	-1.24
298.15	-3.43	-3.25	-3.55	-3.32		2.54	1.53	1.83		-0.58	-1.31	-1.24
303.15	(-3.27)	-3.11	-3.40	-3.16	1.90	2.68	1.70	1.99	-1.23	-0.58	-1.31	-1.25
308.15	-3.09	-2.98	-3.25	-3.01		2.82	1.85	2.14		-0.58	-1.31	-1.26
313.15		-2.86	-3.10	-2.87		2.95	2.00	2.28	-1.16	-0.57	-1.31	-1.26
318.15	-2.82	-2.73	-2.96	-2.73		3.08	2.15	2.42		-0.56	-1.31	-1.26
323.15	(-2.79)	-2.61	-2.83	-2.60	2.46	3.21	2.29	2.56	-1.17	-0.56	-1.31	-1.27
328.15		-2.49	-2.70	-2.48		3.33	2.43	2.69		-0.56	-1.31	-1.27
333.15		-2.38	-2.57	-2.36		3.44	2.56	2.82	-1.14	-0.55	-1.31	-1.27
338.15		-2.27	-2.45	-2.24		3.56	2.69	2.94		-0.55	-1.30	-1.27
343.15		-2.17	-2.34	-2.13		3.67	2.82	3.06	-1.09	-0.54	-1.30	-1.27
348.15		-2.07	-2.22	-2.03		3.77	2.95	3.17		-0.54	-1.29	-1.26
353.15		-1.98	-2.11	-1.92		3.88	3.06	3.28	-1.00	-0.53	-1.29	-1.26
358.15		-1.88	-2.01	-1.83		3.98	3.18	3.39		-0.52	-1.29	-1.26
363.15		-1.79	-1.91	-1.73		4.07	3.29	3.49		-0.51	-1.29	-1.26
368.15		-1.7	-1.81	-1.64		4.17	3.40	3.59		-0.51	-1.28	-1.25
SE		0.13	0.18	0.08		0.77	0.21	0.08		0.60	0.17	0.14
MAF		0.11	-0.17	0.06		0.77	-0.20	0.08		0.59	-0.15	-0.11

Tabellen der temperaturabhängigen Berechnungen