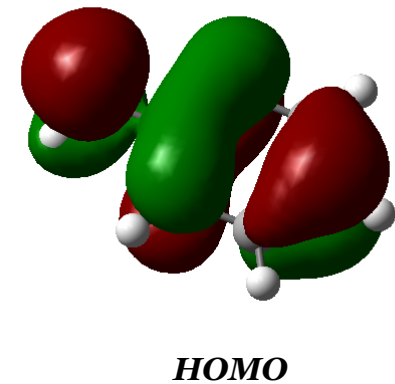
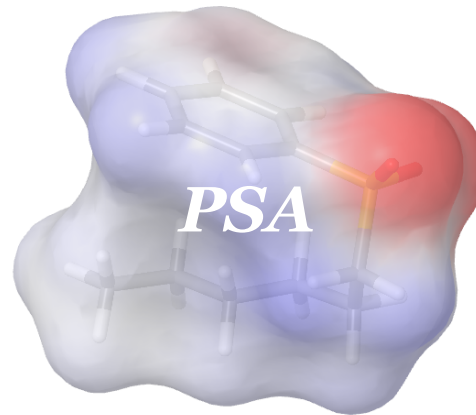
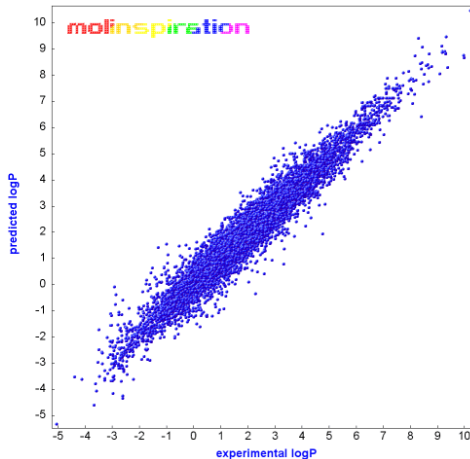
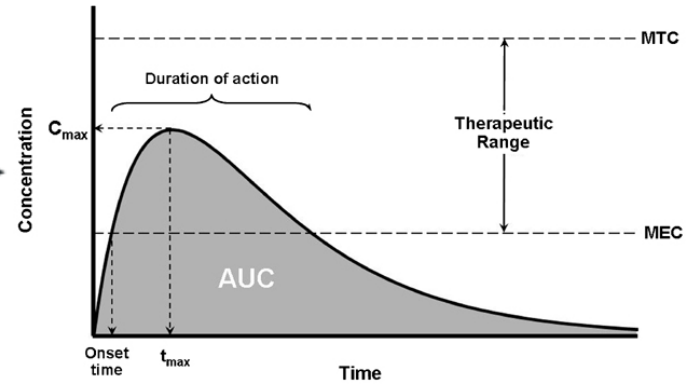
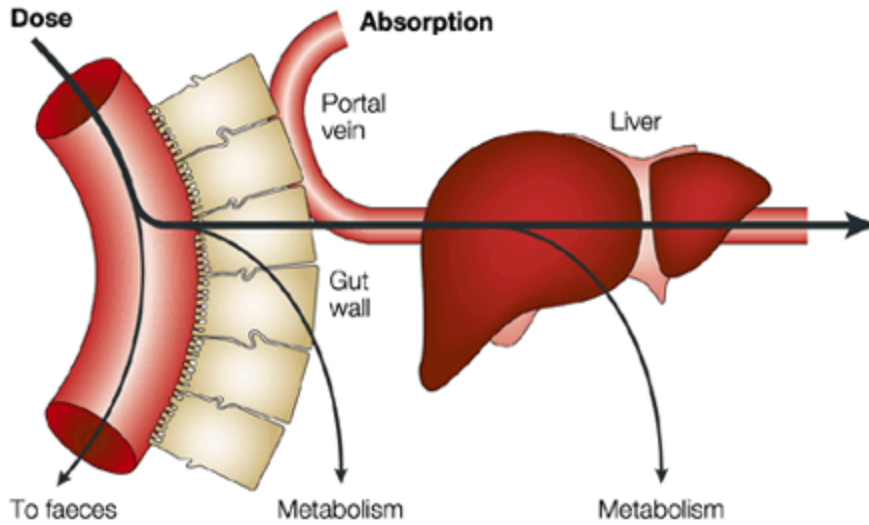


# Vorlesung 12205: Einführung ins Molecular Modeling

M. Smieško & A. Vedani — Departement Pharmazeutische Wissenschaften, Universität Basel, HS-2017



# *in silico* ADMET (Pharmakokinetik)





# ADMET

**A**bsorption + **D**istribution + **M**etabolism + **E**limination + **T**oxicity

Computer Aided ADMET = Predicting (estimating) biopharmaceutical properties of drugs and drug candidates, e.g. solubility, permeability, partitioning coefficient, dissociation constant, using various computational methods.

**database approach** – searching database for similar compounds → similar properties

**computational approach** – calculation of properties from the first principles or using algorithms trained on experimental data (Quantitative Structure-Property Relationships)

“**Rule of 5**” publication by Lipinski et. al. has the highest impact on drug design and development process in the latest years



## Lipinski's Rule of 5

“... ‘the rule of 5’ predicts that poor **absorption** or **permeation** is more likely when there are more than **5 H-bond donors**, **10 H-bond acceptors**, the **molecular weight** (MWT) is greater than **500** and the calculated **LogP** (CLogP) is greater than **5** (or MlogP > 4.15).”

*Christopher A. Lipinski, Franco Lombardo, Beryl W. Dominy, Paul J. Feeney*

Experimental and computational approaches to estimate solubility and permeability  
in drug discovery and development settings

*Advanced Drug Delivery Reviews (1997), 23(1–3), 3-25.*

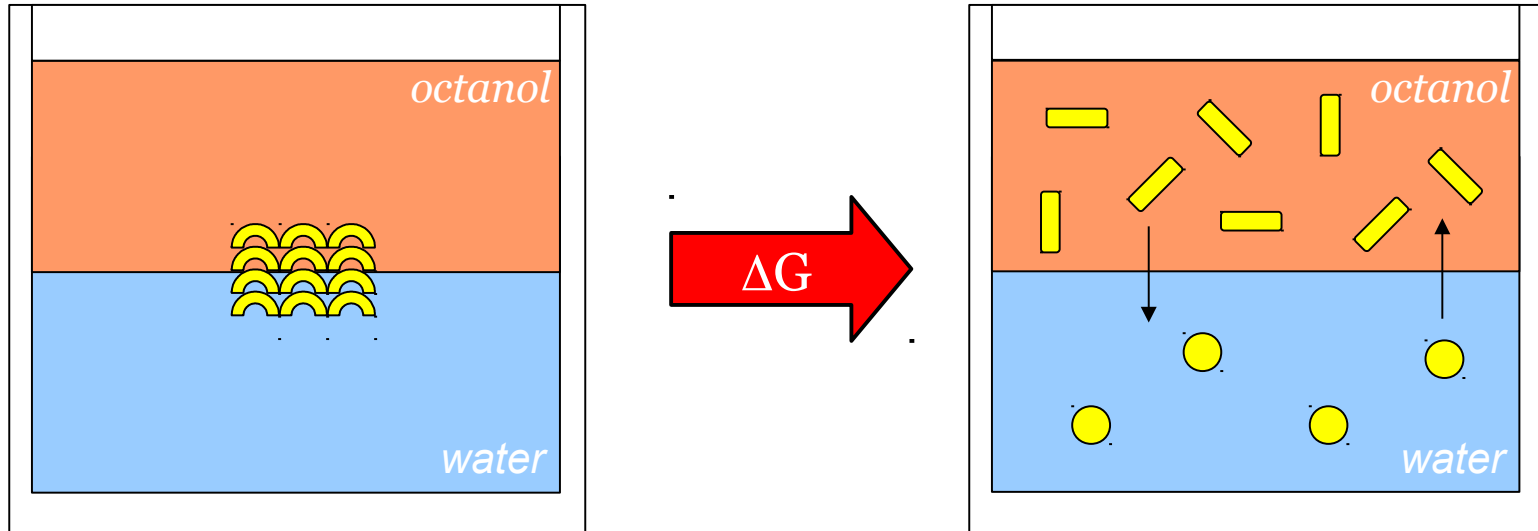
Nowadays fine-tuned, some exceptions identified, extended by additional rules (e.g. rule about rotatable bonds → entropic penalty) and specialized (e.g. blood-brain barrier permeation → less donors, lower flexibility, less polar surface area...)





# Logarithm of the Partition Coefficient – LogP

The most important lipophilicity descriptor



 crystal conformation       collapsed conformation       extended conformation

$$P = c_{\text{octanol}} / c_{\text{water}} = 8 \text{ units} / 4 \text{ units} = 2$$

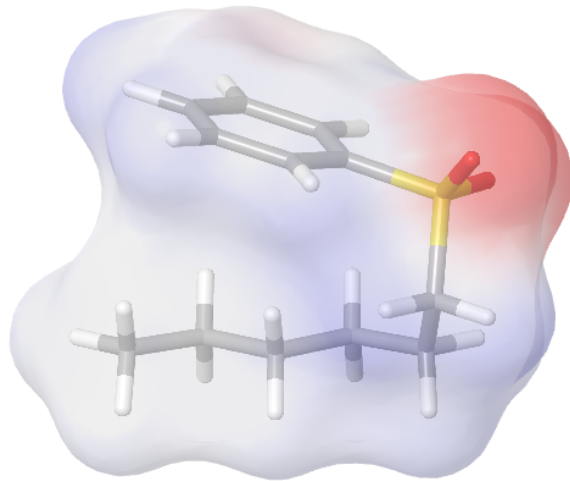
$$\text{LogP} = \log(2) = 0.301$$



# Impact of solvent on conformation and properties

● **collapsed** in water

minimize contact area of lipophilic parts  
with highly polar water molecules



$$3D\text{-PSA} = 37 \text{ \AA}^2$$

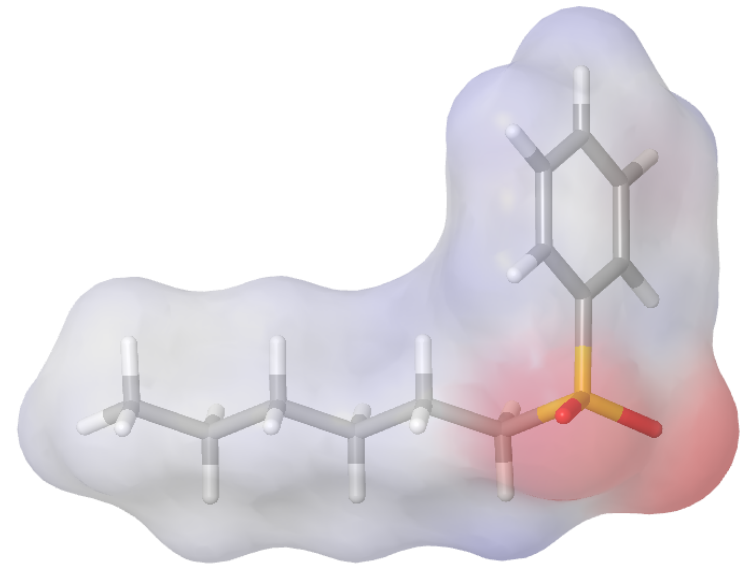
$$3D\text{-SASA} = 437 \text{ \AA}^2$$

$$\Delta G_s \text{ water} = -11.8 \text{ kcal/mol}$$

$$\Delta G_s \text{ octanol} = -11.3 \text{ kcal/mol}$$

■ **extended** in octanol

maximize contact of lipophilic parts  
with lipophilic octanol molecules



$$3D\text{-PSA} = 37 \text{ \AA}^2$$

$$3D\text{-SASA} = 505 \text{ \AA}^2$$

$$\Delta G_s \text{ water} = -11.7 \text{ kcal/mol}$$

$$\Delta G_s \text{ octanol} = -11.8 \text{ kcal/mol}$$

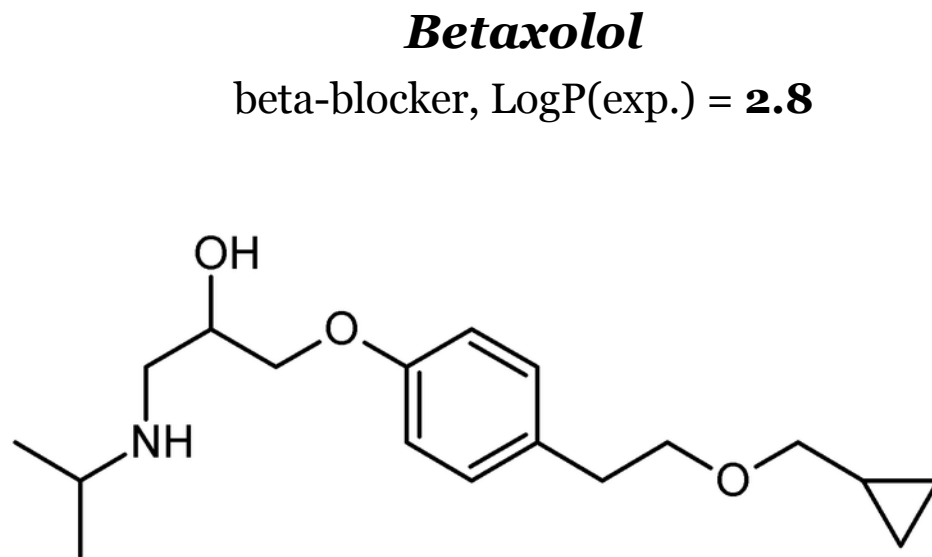
calculated with OPLS2005 force-field



## cLogP (computed log P)

Fragment	$\pi$ -value
Phenyl	+2.0
-Cl	+0.5
C (aliphatic)	+0.5
-ONO <sub>2</sub>	+0.2
-S-	0.0
-NO <sub>2</sub> (aromat.)	-0.28
O=C-O- (carboxyl)	-0.7
O=C-N- (amide)	-0.7
-NO <sub>2</sub> (aliphatic)	-0.85
-O- (hydroxyl, ether)	-1.0
N (amine)	-1.0

$$\text{cLogP} = \sum_{i=1}^n \pi_i$$



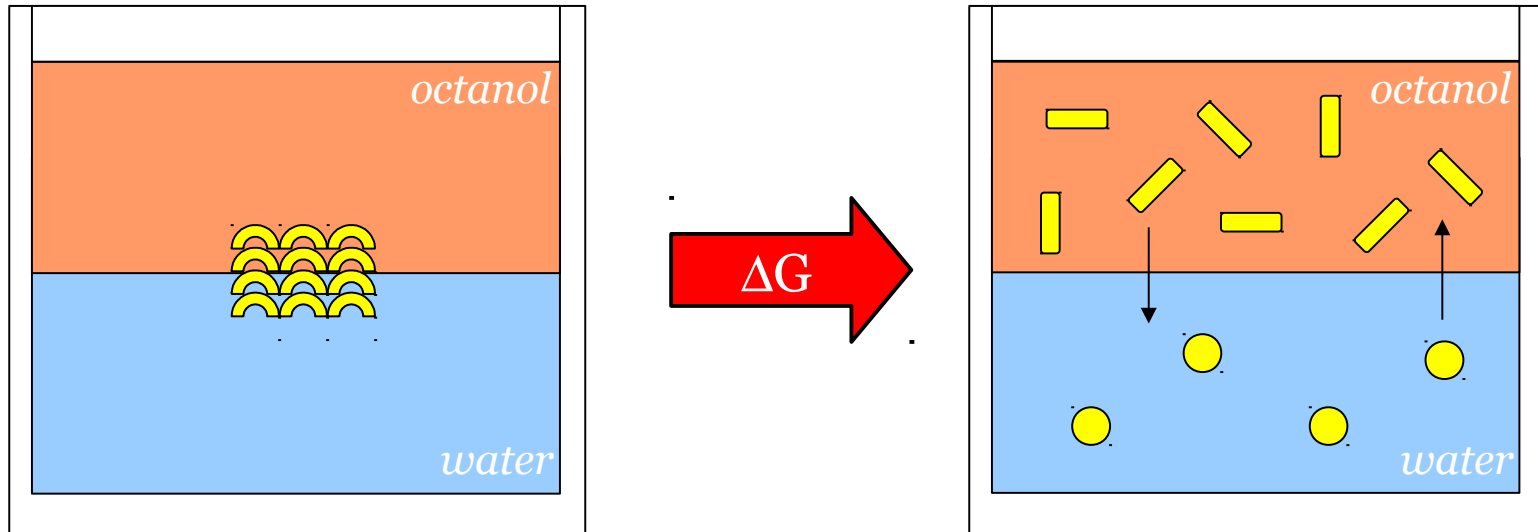
12 x aliphatic C	+6.0
1 x phenyl	+2.0
3 x O	-3.0
1 x N	-1.0

---

**cLogP = +4.0**



## Partition Coefficient - LogP



$$\Delta G = \Delta G_{s\text{water}} - \Delta G_{s\text{octanol}}$$

$$\text{LogP}_{(o/w)} = \Delta G / 2.303 RT$$

Simple fragment based approach → counting together *per* atom (group) contributions



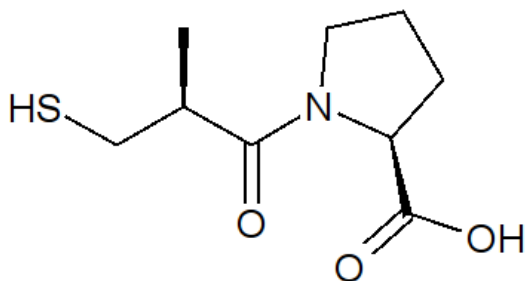
# Distribution Coefficient LogD - including protonation

$$D = c_{\text{octanol}} / (c_{\text{water ionized}} + c_{\text{water neutral}})$$

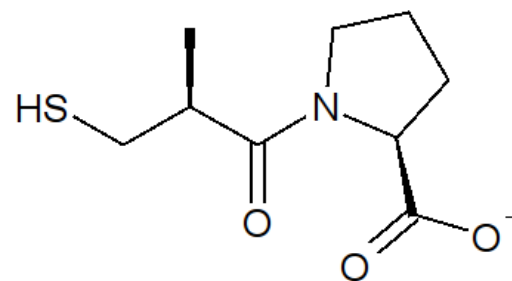
LogD depends on pH as it includes de/protonated states

LogD  $\approx$  LogP (for neutral at pH far from pKa)

pKa's  $\rightarrow$  carboxyl 4.0, sulfhydryl 10.0



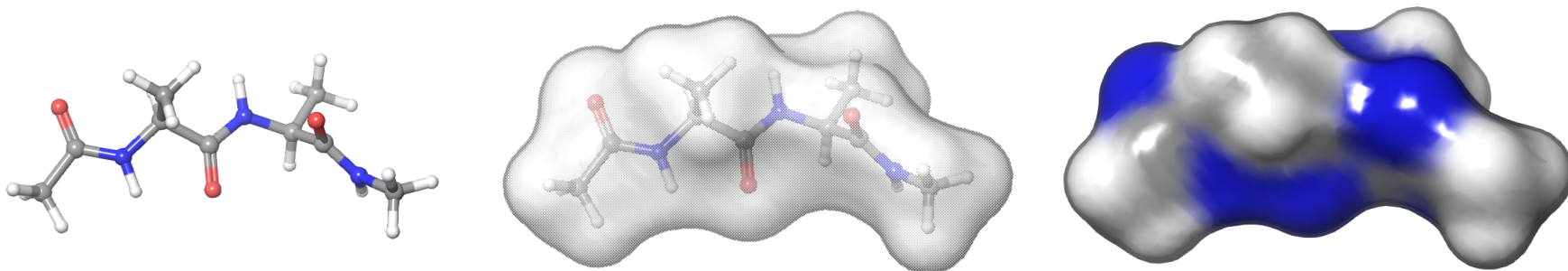
Captopril (logP = 1.02)



Captoprilat (logD<sub>7.4</sub> = -2.00)



## Polar Surface Area - PSA



**Ertl P., Rohde B., Selzer P:** Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties, *J. Med. Chem.* (2000), 43, 3714-3717:

*“Molecular polar surface area (PSA), i.e., surface belonging to polar atoms, is a descriptor that was shown to correlate well with passive molecular transport through membranes and, therefore, allows prediction of transport properties of drugs.”*

**Veber D.F, Johnson S.R., Cheng H.-Y., Smith B.R., Ward K.W., Kopple K.D.:** Molecular Properties That Influence the Oral Bioavailability of Drug Candidates, *J. Med. Chem.* (2002), 45, 2615-2623:

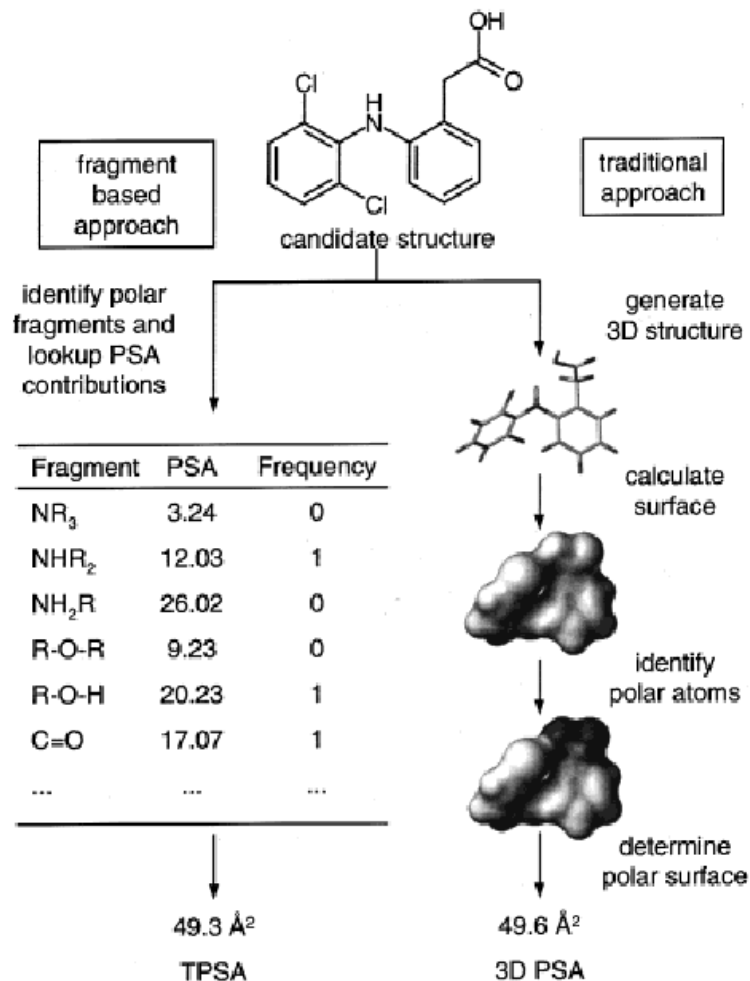
*“...observations suggest that compounds which meet only the two criteria of (1) 10 or fewer rotatable bonds and (2) polar surface area equal to or less than  $140 \text{ \AA}^2$  (or 12 or fewer H-bond donors and acceptors) will have a high probability of good oral bioavailability in the rat.”*



# Polar Surface Area – 3D PSA vs. tPSA

atom type <sup>a</sup>	PSA contrib	atom type <sup>a</sup>	PSA contrib
[N](-*)(-*)-*	3.24	[nH](:*):*	15.79
[N](-*)=*	12.36	[n+](:*):(*):*	4.10
[N]#*	23.79	[n+](-*)(:*):*	3.88
[N](-*)(=*)=* <sup>b</sup>	11.68	[nH+](:*):*	14.14
[N](=*)#* <sup>c</sup>	13.60	[O](-*)-*	9.23
[N]1(-*)-*-* <sup>-1d</sup>	3.01	[O]1-*-* <sup>-1d</sup>	12.53
[NH](-*)-*	12.03	[O]=*	17.07
[NH]1-*-* <sup>-1d</sup>	21.94	[OH]-*	20.23
[NH]=*	23.85	[O-]-*	23.06
[NH2]-*	26.02	[o](:*):*	13.14
[N+](-*)(-*)(-*)-*	0.00	[S](-*)-*	25.30
[N+](-*)(-*)=*	3.01	[S]=*	32.09
[N+](-*)#* <sup>e</sup>	4.36	[S](-*)(-*)=*	19.21
[NH+](-*)(-*)-*	4.44	[S](-*)(-*)(=*)=*	8.38
[NH+](-*)=*	13.97	[SH]-*	38.80
[NH2+](-*)-*	16.61	[s](:*):*	28.24
[NH2+]=*	25.59	[s](=*)(*):*	21.70
[NH3+]-*	27.64	[P](-*)(-*)-*	13.59
[n](:*):*	12.89	[P](-*)=*	34.14
[n](*)(*):*	4.41	[P](-*)(-*)(-*)=*	9.81
[n](-*)(*):*	4.93	[PH](-*)(-*)=*	23.47
[n](=*)(*):* <sup>f</sup>	8.39		

<sup>a</sup> An asterisk (\*) stands for any non-hydrogen atom, - for a single bond, = for a double bond, # for a triple bond, : for an aromatic bond; atomic symbol in lowercase means that the atom is part of an aromatic system. <sup>b</sup> As in nitro group. <sup>c</sup> Middle nitrogen in azide group. <sup>d</sup> Atom in a three-membered ring. <sup>e</sup> Nitrogen in isocyano group. <sup>f</sup> As in pyridine *N*-oxide.

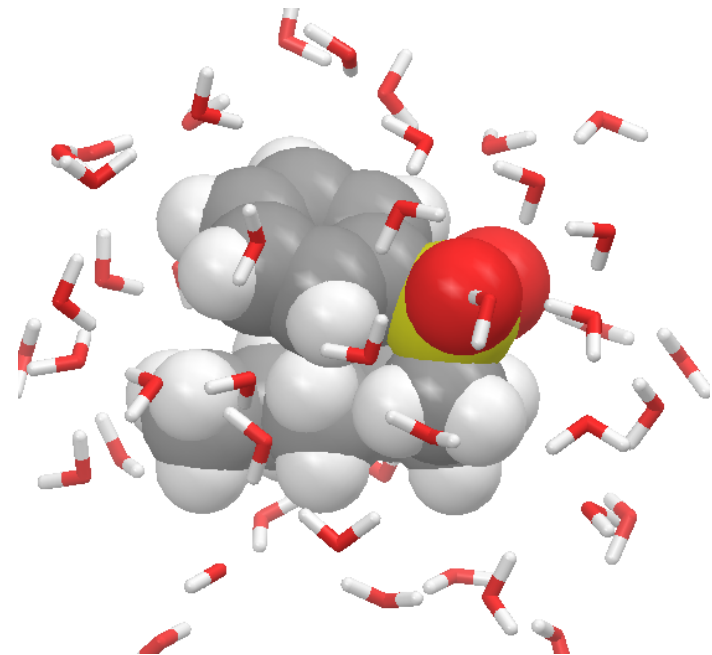
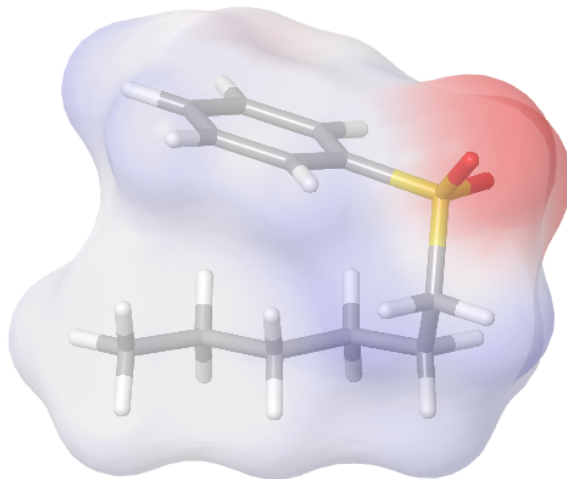






## Solvation energy

- energy released at transfer of a given molecule from gas phase to solvent phase
- very important descriptor, physical constant
- various computational methods available:
  - Continuum methods (surface, cavity, volume, mapped properties; MM and QM)
  - Explicit methods – interaction with explicit water molecules (MDs)



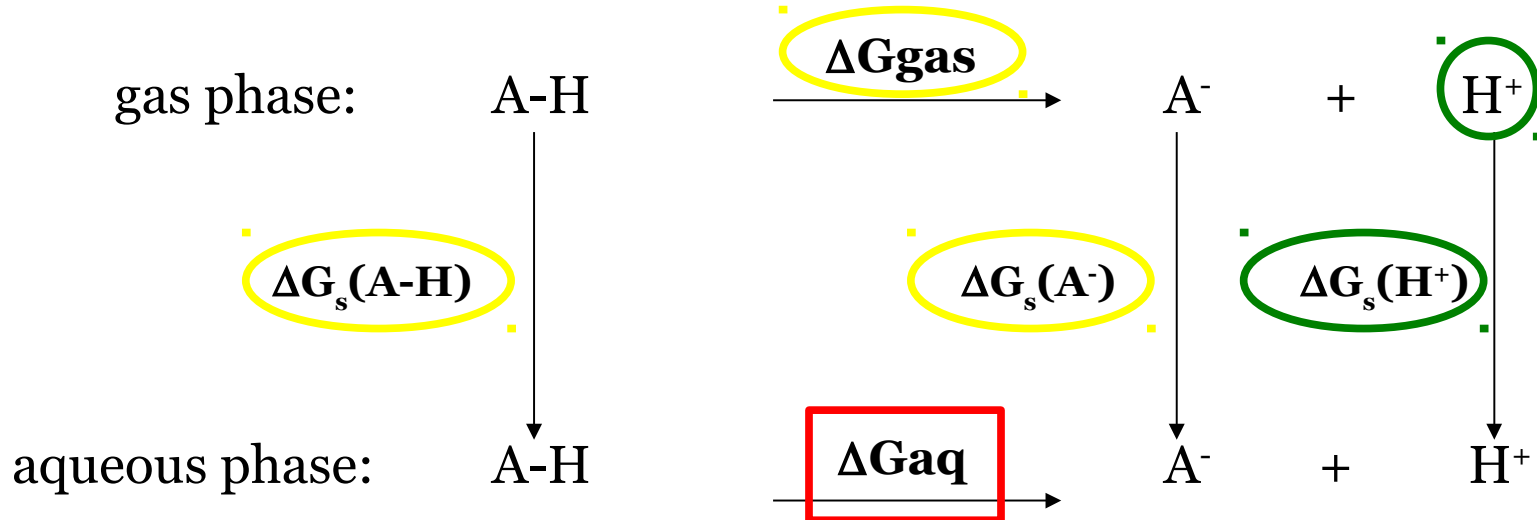


## Proton dissociation constant - $pK_a$

$$pK_a = -\log K_a$$

$$\Delta G = -RT \ln K_a = -2.303 \cdot RT \log K_a$$

$$pK_a = \Delta G_{aq} / -2.303 RT$$

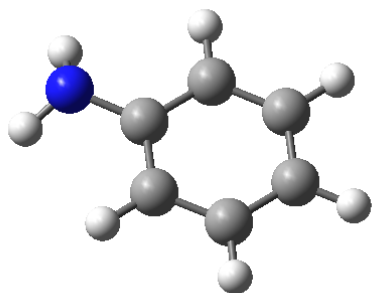




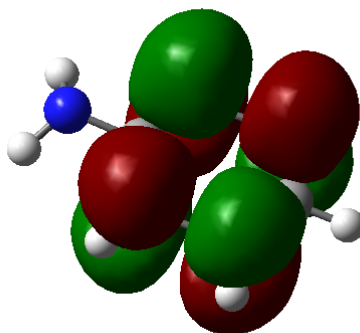
## Metabolism

Identification of labile/exposed regions relevant for metabolic stability (→ metabolites)

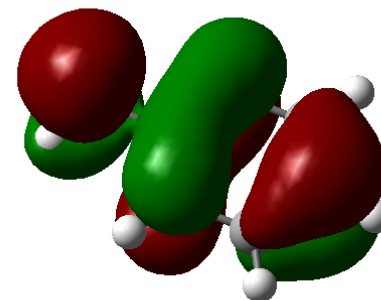
- HOMO (the **H**ighest **O**ccupied **M**olecular **O**rbital)
- LUMO (the **L**owest **U**noccupied **M**olecular **O**rbital)



*aniline*



*LUMO*



*HOMO*

*calculated at B3LYP/6-31G(d) level of theory in Gaussian09*

**3D approaches:** predicting affinity for various liver metabolic enzymes CYP-450s using docking or grid based techniques, accessibility (exposure) of fragments for reactions

**limited or no metabolism → better pharmacological properties!**



## Toxicity

- avoid **reactive functional groups** forming covalent bonds with the target protein
  - avoid **reactive metabolites**
  - avoid chelating groups (2D, 3D), flat polyaromatic systems, extreme properties in general
  - avoid extremely high affinity → optimal range usually low nanomolar ( $IC_{50}$  or  $K_i \sim 1-100$  nM) possible problems: complicated drug formulation, easy overdose → toxicity
  - **minimize off-target binding** – examine binding to known receptors by molecular docking
- anti-targets: hERG  $K^+$  channel = arrhythmia, nuclear receptors = endocrine disruption, liver metabolic enzymes CYP-450s = drug-drug interactions, DNA binding = carcinogenicity

**VirtualToxLab** – in silico estimation of toxic potential of drugs and chemicals



## Conclusion - *in silico* ADMET

- integral part of the lead compound optimization
- useful biopharmaceutical properties/descriptors can be – rapidly and accurately – calculated without need to have compound physically available (synthesized) → pre-screening/compound prioritization
- helps to make drugs safer / more efficient
- saves costs (less experiments needed)
- ethical dimension (reduced animal testing)



# Quantitative Struktur–Wirkungs-Beziehungen

(Quantitative structure-activity relationships, QSAR)

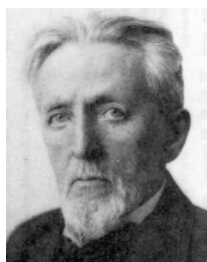


# QSAR: Quantitative Struktur–Wirkungs-Beziehungen

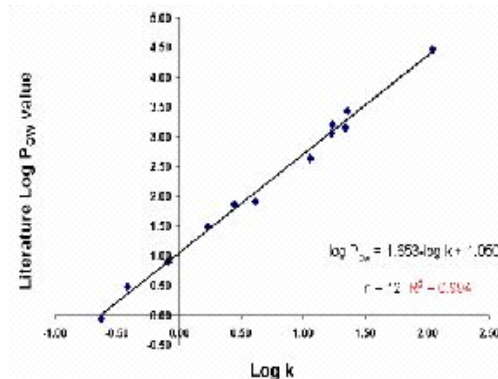
Mathematische Verfahren, die eine Korrelation zwischen der chemischen Struktur einer Substanz und deren biologischer Aktivität (z.B. Bindungsaffinitäten), chemischer Reaktivität (z.B. Reaktionskonstanten) oder toxikologischen Endpunkten (z.B. Karzinogenizität) erstellen, fasst man unter dem Begriff **QSAR** (Quantitative Struktur–Wirkungs-Beziehungen) zusammen:

**Aktivität =  $f$ (physikochemischen und/oder strukturellen Eigenschaften)**

Die erste Struktur-Wirkungs-Beziehung (SAR) wurde 1863 vom Arzt Antoine F. Cros aufgestellt: Er beobachtete, dass die Giftigkeit von Alkoholen (bei Säugern) zunahm, wenn die Wasserlöslichkeit der Substanz abnahm. 1899 fanden Hans-Horst Meyer und Charles E. Overton eine Beziehung zwischen der Aktivität von 51 Narkotika und deren Akkumulation in der lipophilen Phase:  $\log(1/C) = 0.94 \log P + 0.87$ .



Hans-Horst Meyer



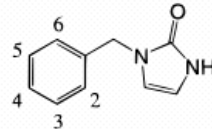




# QSAR: Quantitative Struktur–Wirkungs-Beziehungen

1935–1938 korrelierte Lois Hammett elektronische Eigenschaften von organischen Säuren und Basen mit deren  $pK_a$ -Wert:  $\log(K) = \log(K_0) + \rho \cdot \sigma$ . Free und Wilson fanden 1964 eine Aufschlüsselungsmethode, wie stark die verschiedenen funktionellen Gruppen zur Aktivität einer Substanz beitragen, und so eine Möglichkeit, Wirkstoffe gezielt zu analysieren. Im selben Jahr veröffentlichten Hansch und Fujita eine Struktur–Wirkungs-Beziehung, welche die Toxizität von Phenolen auf grampositive und gramnegative Bakterien erklärt. Seither haben Forscher versucht, QSAR zur Voraussage der Aktivität neuer Wirksubstanzen einzusetzen. So auch Martin Karplus (Nobelpreis 2013) im Jahre 1997 zur Berechnung von Enzym-Bindur

Protein = Dopamin  $\beta$ -Hydroxylase



*J. Med. Chem.* **1997**, 40, 4360-4371

no.	R	pIC <sub>50</sub>	no.	R	pIC <sub>50</sub>	no.	R	pIC <sub>50</sub>
1	2,6-(CH <sub>3</sub> ) <sub>2</sub>	3.00	17	3-NO <sub>2</sub> , 4-OCH <sub>3</sub>	3.45	33	H	4.48
2	2,6-Cl <sub>2</sub>	3.15	18	4-OCH <sub>3</sub>	3.69	34	3-NO <sub>2</sub> , 4-OH	4.51
3	2,6-(OCH <sub>3</sub> ) <sub>2</sub>	3.30	19	3-OCH <sub>3</sub>	3.80	35	3,4-Cl <sub>2</sub>	4.55
4	2-Cl	3.45	20	3-OH	3.83	36	2,4-Cl <sub>2</sub>	4.77
5	2-CH <sub>3</sub>	3.47	21 <sup>c</sup>	3-CF <sub>3</sub> , 4-OH	3.92	37	3-Br, 4-OH	4.92
6	3,4-(OCH <sub>3</sub> ) <sub>2</sub>	3.47	22	2,4,6-Cl <sub>3</sub>	3.99	38	3-Cl	4.92
7	4-CF <sub>3</sub>	3.70	23	2,5-Cl <sub>2</sub>	4.01	39	3-F	5.25
8	3-CF <sub>3</sub> , 4-OCH <sub>3</sub>	3.76	24	4-Cl	4.02	40	4-OH	5.59
9	2,6-Cl <sub>2</sub> , 4-OCH <sub>3</sub>	3.81	25	2,6-Cl <sub>2</sub> , 4-OH	4.12	41	3,5-Cl <sub>2</sub>	5.62
10	4-CH <sub>3</sub>	3.83	26	2,3,5,6-F <sub>4</sub> , 4-OH	4.21	42	3,4-(OH) <sub>2</sub>	5.66
11	4-Br	3.94	27	4-NO <sub>2</sub>	4.28	43	3-Cl, 4-OH	5.70
12	3-Br, 4-OCH <sub>3</sub>	4.08	28	2,3-Cl <sub>2</sub>	4.28	44	3-F, 4-OH	5.82
13	3-F, 4-OCH <sub>3</sub>	4.13	29	3-CH <sub>3</sub> , 4-OH	4.31	45	3,5-F <sub>2</sub>	5.92
14	2-OCH <sub>3</sub>	4.13	30	4-F	4.33	46	3,5-Cl <sub>2</sub> , 4-OH	6.17
15	3-CH <sub>3</sub> , 4-OCH <sub>3</sub>	4.16	31	3,5-Cl <sub>2</sub> , 4-OCH <sub>3</sub>	4.33	47	3,5-F <sub>2</sub> , 4-OH	7.13
16	2-OH	3.24	32	3,5-F <sub>2</sub> , 4-OCH <sub>3</sub>	4.44			



## QSAR: Quantitative Struktur–Wirkungs-Beziehungen

Typischerweise werden den Wirksubstanzen Deskriptoren zugeordnet, die physikochemische (wie log P, pK<sub>a</sub>, Molekulargewicht, Dipolmoment, Polarisierbarkeit) oder topologische Eigenschaften (Volumen, Anzahl Wasserstoffbrückendonoren und -akzeptoren, Anzahl frei drehbare Bindungen, Lösungsmittelzugängliche Oberfläche) repräsentieren. Durch unterschiedliche Gewichtung dieser Terme kann eine (oft komplexe) mathematische Beziehung gefunden werden, welche daraus Affinitäten (Reaktivitäten, toxische Endpunkte) berechnet.

The screenshot shows the Tstar software interface. The main window displays a table with columns for Row ID, Whole M, Subst. 1, Subst. 2, Subst. 3, Name (Whole M, Subst. 1, Subst. 2, Subst. 3), log(ED30), PI (Aromatic), and MR (Aromatic). The table contains 11 rows of data. An ADME Screen dialog box is open, showing screening properties: Molecular weight (< 500), H-bond donors (<= 5), H-bond acceptors (<= 10), logP value (< 5), and Rotatable bond count (<= 8). The Tstar Visualizer window shows a 3D ball-and-stick model of a molecule.

Row ID	Whole M	Subst. 1	Subst. 2	Subst. 3	Name (Whole M)	Name (Subst. 1)	Name (Subst. 2)	Name (Subst. 3)	log(ED30)	PI (Aromatic)	MR (Aromatic)	
1	A				A	CHLORO	HYDROG	CHLORO	0.43	0.71	6.03	0
2	B									0.86	8.88	0
3	C									0.86	8.88	0
4	D									0.71	6.03	0
5	E									0.56	5.65	0
6	F									0.71	6.03	0
7										0.71	6.03	0
8										0.56	5.65	0
9										0.71	6.03	0
10						METHYL	METHYL	HYDROG	2.91	0.56	5.65	0
11						HYDROG	HYDROG	HYDROG	4.4	0	1.03	0

The screenshot shows the Accelrys Discovery Studio software interface. The main window displays a table with columns for Structure, Name, and Index. The table contains 6 rows of data. A 3D PCA plot is shown on the right, with axes labeled PCA1\_PCA2, PCA1\_PCA3, and PCA1\_PCA4. The plot shows a cluster of points colored by activity, ranging from 0 (blue) to 1 (red). A legend for Activity is provided below the plot.

Structure	Name	Index	Act
392	3716087	392	Ina
393	849477	393	Ina
394	861822	394	Ina
395	4243389	395	Ina
396	845361	396	Ina

accelrys.com



# QSAR: Quantitative Struktur–Wirkungs-Beziehungen

In der Arzneistoffentwicklung vermögen solche einfachen QSAR oft nicht zu genügen, denn die Aktivität eines Wirkstoffes ist eine komplexe Grösse und hängt insbesondere von seinen Wechselwirkungen mit dem Makromolekül ab:

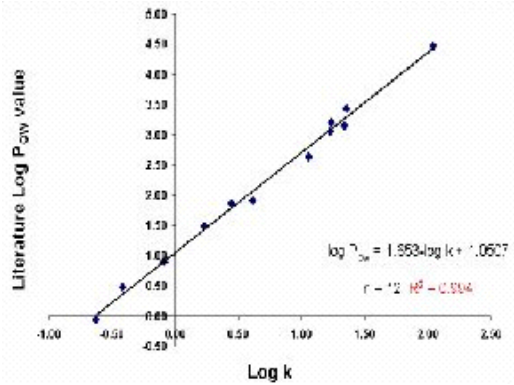
$$E_{\text{Bindung}} = E_{\text{Ligand-Protein}} - E_{\text{Innere Spannung}} - E_{\text{Ligand-Desolvatation}} - T\Delta S - E_{\text{Induced Fit}}$$

Die bioaktive Konformation (und Orientierung) lässt sich aber nur in an einer experimentellen Struktur (bzw. einem Homologiemodell) identifizieren. Im so genannten “mixed-model QSAR” werden beide Methoden kombiniert: Der Bindungsmodus wird durch flexibles Docking (Simulation des *induced fit*) an die Proteinstruktur identifiziert und anschliessend mit QSAR bewertet. Quantitative Struktur-Wirkungs-Beziehungen können aufgrund ihrer “Dimensionalität” klassiert werden, wobei unterschieden werden muss, ob es sich um mathematische, virtuelle oder strukturelle Modelle handelt. Im Molecular Modeling werden strukturelle und virtuelle Modelle verwendet – quasi-atomistische Ansätze stellen ein Hybrid (Kombination) dar.

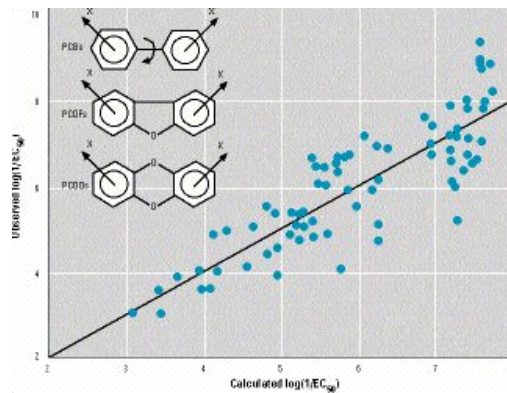
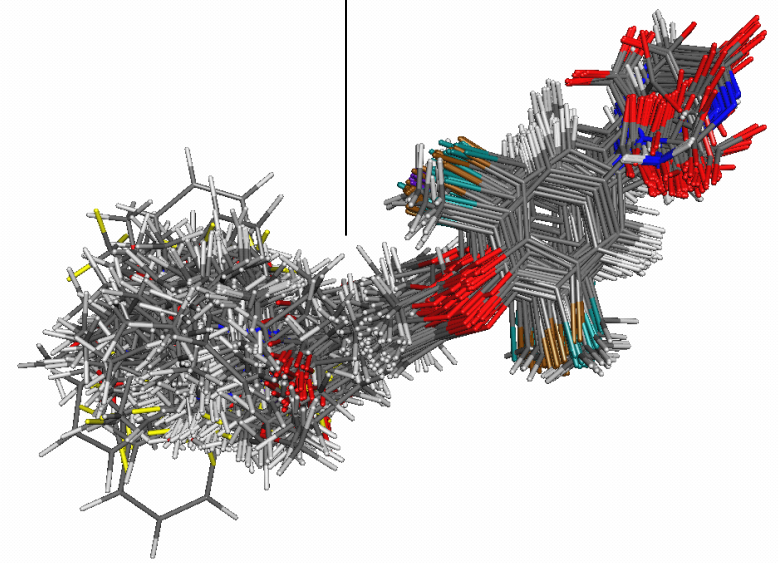
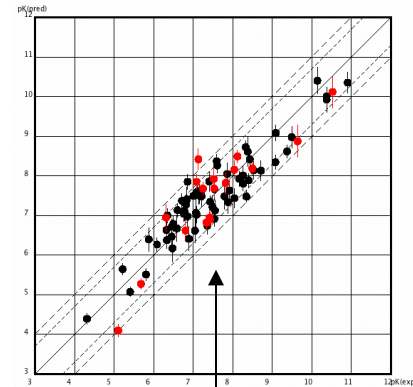
Einfachere Fragestellungen – wie beispielsweise Haut- oder Augenreizungen durch Chemikalien – lassen sich aber durchaus mit rein physikochemikalischen Parametern beschreiben, denn hier geht es lediglich um die Eigenschaften einer Substanz (z.B. Korrosivität) und nicht um deren Wechselwirkungen mit einem Makromolekül.



# Klassische QSAR-Ansätze



**1D-QSAR:** Aktivität korreliert mit einer physikochemischen Eigenschaft



**2D-QSAR:** Aktivität korreliert mit strukturellen Motiven (Verknüpfungsmuster)

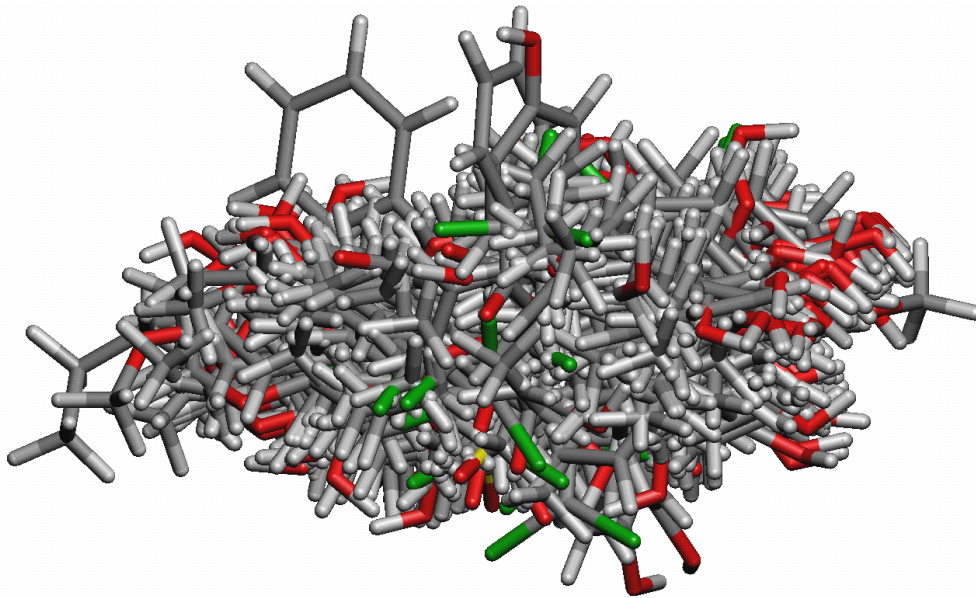
**3D-QSAR:** Aktivität korreliert mit räumlicher Struktur



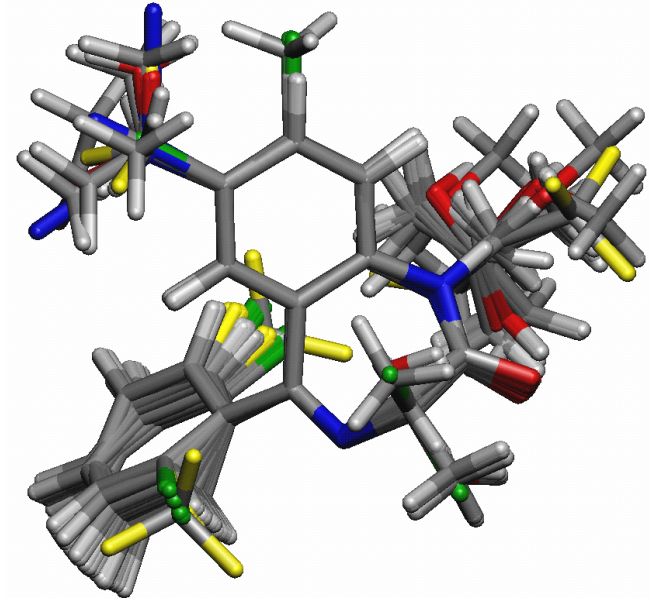


## Erstellen einer QSAR

Voraussetzung zur Erstellung eines QSAR ist ein Satz von (typischerweise 20–50) **Strukturen** (3D: Koordinaten, 2D: Verknüpfungsmuster, 1D: physikochemische Eigenschaften) sowie die experimentell bestimmte **Aktivität** (Reaktivität, toxischer Endpunkt). Im 3D-QSAR werden die drei-dimensionalen Strukturen der Wirkstoffe bevorzugterweise durch Andocken an eine 3D-Proteinstruktur (ggf. Homologiemodell) erzeugt. Im ligand-basierten Design können sie mittels einer Pharmakophorhypothese (Überlagerung aller Strukturen in deren bioaktiver Konformation) gefunden werden.



Östrogenrezeptor → : Überlagerung aus Docking ans Protein

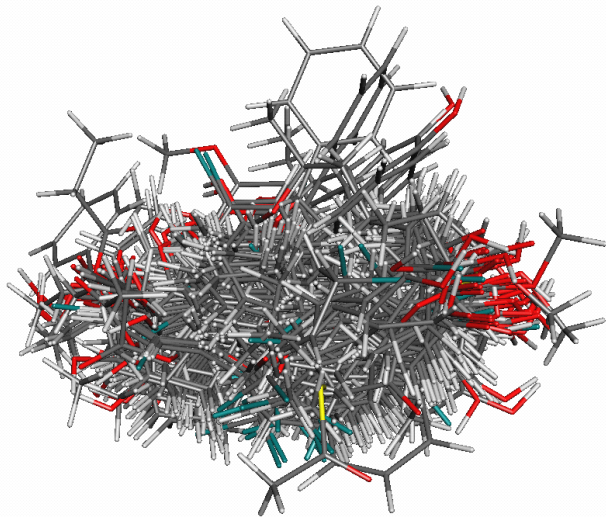


GABA<sub>A</sub>-Rezeptor: Überlagerung aus Pharmakophorhypothese

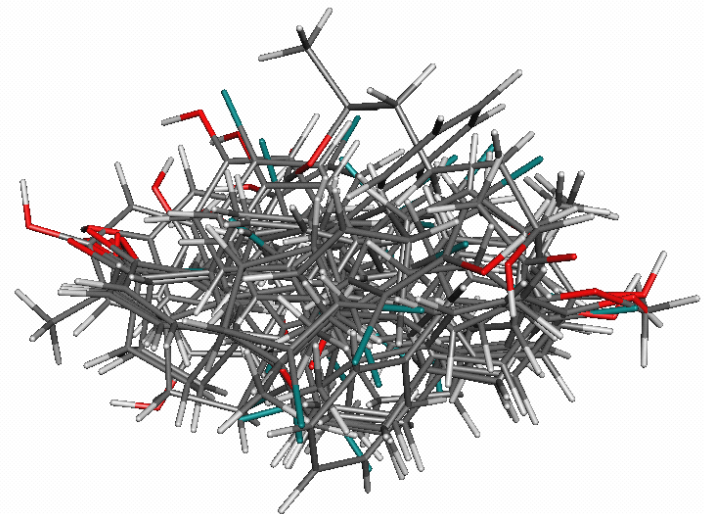


## QSAR: Trainings- und Testsatz

Weil – genügend Parameter/Variablen vorausgesetzt – im Prinzip (fast) alles erklärt werden kann, müssen Protokolle angewendet werden, die sicherstellen, dass ein erstelltes QSAR auch eine Voraussagekraft hat. Der wichtigste Ansatz hierzu ist es, die Wirksubstanzen in einen so genannten “Trainingsatz” und einen “Testsatz” aufzuteilen. Am Trainingsatz wird das Modell entwickelt und optimiert; am Testsatz wird die Voraussagekraft ermittelt. Typischerweise werden Trainings- und Testsatz im Verhältnis 3:1 oder 4:1 erstellt. Zentral ist, dass der Trainingsatz “repräsentativ” ist, d.h. alle (im ganzen Datensatz vorkommenden) Eigenschaften (im 3D-QSAR = funktionelle Gruppen) müssen vertreten sein, denn sonst kann diese Eigenschaft im Testsatz – und später bei den wirklichen Voraussagen – nicht verwertet werden.



Androgenrezeptor: Trainingsatz (85 Substanzen)



Androgenrezeptor: Testsatz (20 Substanzen)



# QSAR: Berechnen von Eigenschaften

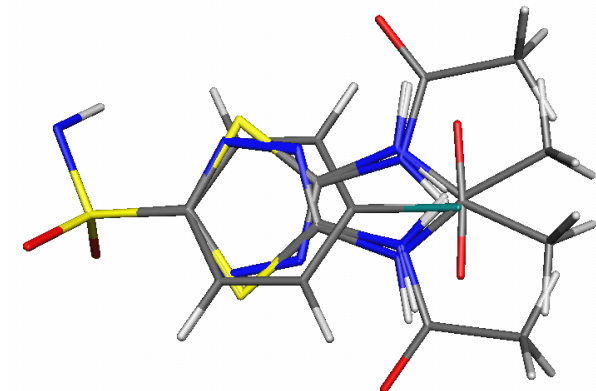
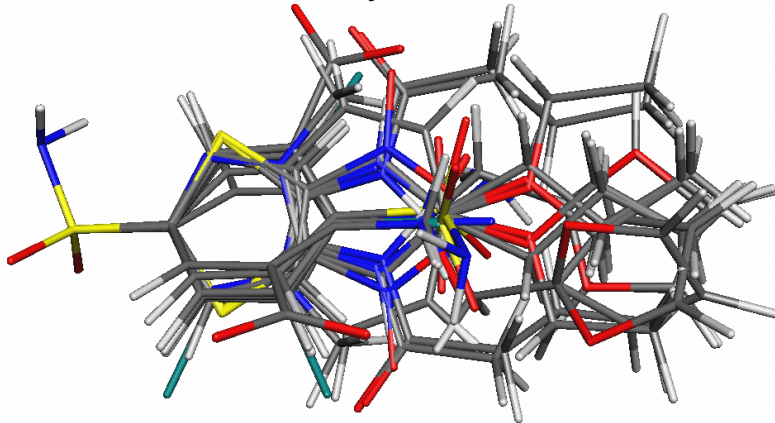
$$E_{\text{Bindung}} = E_{\text{Ligand-Protein}} = E_{\text{elektrostatisch}} + E_{\text{van der Waals}} + E_{\text{Wasserstoffbrücken}} + E_{\text{Polarisation}}$$

Trainingsatz ( $n = 12$ , [kcal/mol])

**EZA:**  $E_{bdg} = -52.4$   
**NTS:**  $E_{bdg} = -49.6$   
**MTZ:**  $E_{bdg} = -49.1$   
**BAA:**  $E_{bdg} = -46.7$   
**NBS:**  $E_{bdg} = -46.0$   
**DBS:**  $E_{bdg} = -42.5$   
**SBS:**  $E_{bdg} = -40.8$   
**CBS:**  $E_{bdg} = -38.9$   
**YBS:**  $E_{bdg} = -38.6$   
**MBS:**  $E_{bdg} = -37.1$   
**FUR:**  $E_{bdg} = -32.3$   
**SAM:**  $E_{bdg} = -32.6$

Testsatz ( $n = 4$ , [kcal/mol])

**AAA:**  $E_{bdg} = -43.8$   
**AAH:**  $E_{bdg} = -40.7$   
**LBS:**  $E_{bdg} = -39.3$   
**BSA:**  $E_{bdg} = -35.4$







# QSAR: Erstellen einer Regression

## Trainingsatz

<b>EZA:</b> $\Delta G(\text{exp}) = -11.8 \leftarrow K(\text{exp}) = 1.599\text{e-}09$	$E_{\text{bdg}}(\text{ber}) = -52.3$	$\Delta G(\text{ber}) = -11.6 \rightarrow K(\text{ber}) = 2.154\text{e-}09$
<b>NTS:</b> $\Delta G(\text{exp}) = -11.2 \leftarrow K(\text{exp}) = 4.199\text{e-}09$	$E_{\text{bdg}}(\text{ber}) = -49.6$	$\Delta G(\text{ber}) = -11.0 \rightarrow K(\text{ber}) = 6.155\text{e-}09$
<b>MTZ:</b> $\Delta G(\text{exp}) = -10.6 \leftarrow K(\text{exp}) = 1.171\text{e-}08$	$E_{\text{bdg}}(\text{ber}) = -49.1$	$\Delta G(\text{ber}) = -10.9 \rightarrow K(\text{ber}) = 7.468\text{e-}09$
<b>BAA:</b> $\Delta G(\text{exp}) = -10.5 \leftarrow K(\text{exp}) = 1.562\text{e-}08$	$E_{\text{bdg}}(\text{ber}) = -46.7$	$\Delta G(\text{ber}) = -10.4 \rightarrow K(\text{ber}) = 1.876\text{e-}08$
<b>NBS:</b> $\Delta G(\text{exp}) = -9.7 \leftarrow K(\text{exp}) = 6.249\text{e-}08$	$E_{\text{bdg}}(\text{ber}) = -46.0$	$\Delta G(\text{ber}) = -10.2 \rightarrow K(\text{ber}) = 2.438\text{e-}08$
<b>DBS:</b> $\Delta G(\text{exp}) = -9.6 \leftarrow K(\text{exp}) = 6.412\text{e-}08$	$E_{\text{bdg}}(\text{ber}) = -42.5$	$\Delta G(\text{ber}) = -9.4 \rightarrow K(\text{ber}) = 9.018\text{e-}08$
<b>SBS:</b> $\Delta G(\text{exp}) = -9.1 \leftarrow K(\text{exp}) = 1.543\text{e-}07$	$E_{\text{bdg}}(\text{ber}) = -40.8$	$\Delta G(\text{ber}) = -9.1 \rightarrow K(\text{ber}) = 1.734\text{e-}07$
<b>CBS:</b> $\Delta G(\text{exp}) = -8.8 \leftarrow K(\text{exp}) = 2.700\text{e-}07$	$E_{\text{bdg}}(\text{ber}) = -38.9$	$\Delta G(\text{ber}) = -8.6 \rightarrow K(\text{ber}) = 3.629\text{e-}07$
<b>YBS:</b> $\Delta G(\text{exp}) = -8.6 \leftarrow K(\text{exp}) = 4.120\text{e-}07$	$E_{\text{bdg}}(\text{ber}) = -38.6$	$\Delta G(\text{ber}) = -8.6 \rightarrow K(\text{ber}) = 4.100\text{e-}07$
<b>MBS:</b> $\Delta G(\text{exp}) = -8.4 \leftarrow K(\text{exp}) = 5.125\text{e-}07$	$E_{\text{bdg}}(\text{ber}) = -37.1$	$\Delta G(\text{ber}) = -8.2 \rightarrow K(\text{ber}) = 7.310\text{e-}07$
<b>FUR:</b> $\Delta G(\text{exp}) = -7.1 \leftarrow K(\text{exp}) = 5.147\text{e-}06$	$E_{\text{bdg}}(\text{ber}) = -32.3$	$\Delta G(\text{ber}) = -7.2 \rightarrow K(\text{ber}) = 4.433\text{e-}06$
<b>SAM:</b> $\Delta G(\text{exp}) = -7.0 \leftarrow K(\text{exp}) = 6.175\text{e-}06$	$E_{\text{bdg}}(\text{ber}) = -32.6$	$\Delta G(\text{ber}) = -7.2 \rightarrow K(\text{ber}) = 4.029\text{e-}06$

Regression —  $\Delta G(\text{exp})$  gegen  $E_{\text{bdg}}(\text{ber})$  — ergibt eine Steigung  $a = 0.222$  ( $r^2 = 0.96$ )  
 $\Delta G(\text{ber})$  wird durch Multiplikation mit 0.222 aus  $E_{\text{bdg}}(\text{ber})$  erhalten

## Testatz bzw. neue Substanzen

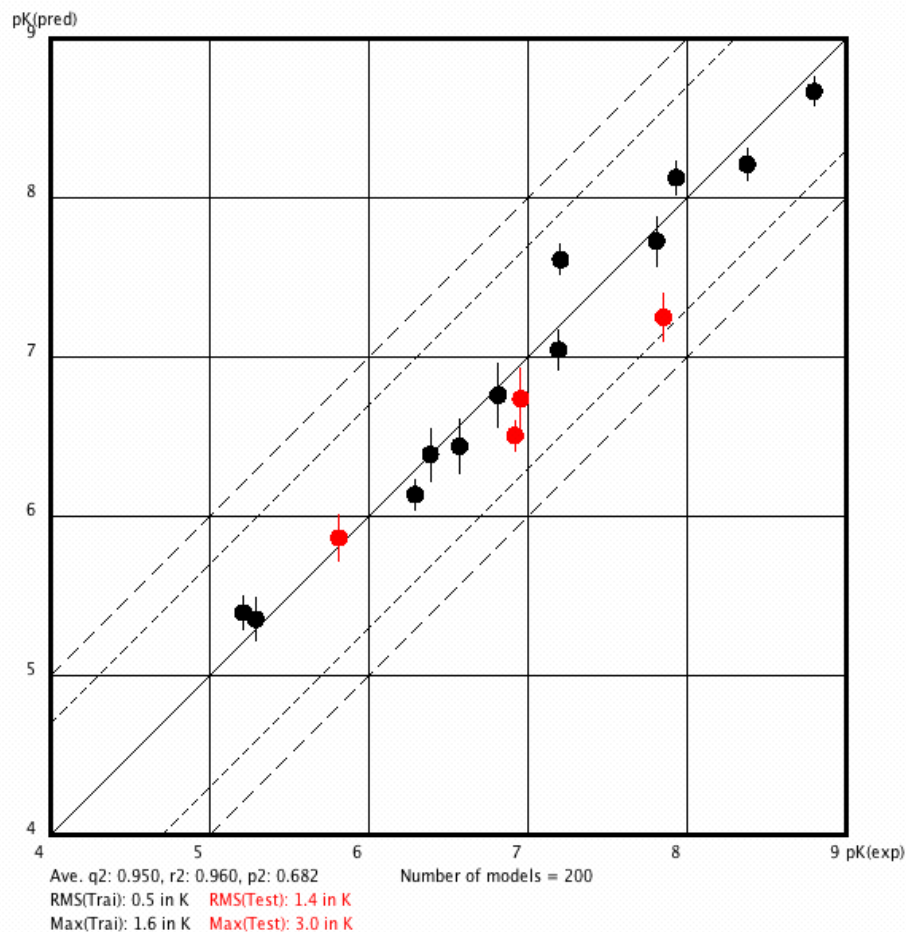
<b>AAA:</b> $E_{\text{bdg}}(\text{ber}) = -43.8$	$\Delta G(\text{ber}) = -9.7 \rightarrow K(\text{ber}) = 5.625\text{e-}08$
<b>AAH:</b> $E_{\text{bdg}}(\text{ber}) = -40.7$	$\Delta G(\text{ber}) = -9.0 \rightarrow K(\text{ber}) = 1.835\text{e-}07$
<b>LBS:</b> $E_{\text{bdg}}(\text{ber}) = -39.3$	$\Delta G(\text{ber}) = -8.7 \rightarrow K(\text{ber}) = 3.103\text{e-}07$
<b>BSA:</b> $E_{\text{bdg}}(\text{ber}) = -35.4$	$\Delta G(\text{ber}) = -7.9 \rightarrow K(\text{ber}) = 1.365\text{e-}06$

Neue Substanzen: hier wird  $E_{\text{bdg}}(\text{ber})$  mit 0.222 multipliziert und so  $\Delta G(\text{ber})$  erhalten

Umrechnungen:  $\Delta G = -RT \ln K$  bzw.  $K = e^{-\Delta G/RT}$  ( $RT = 0.58216$ )



# QSAR: Anwendung der Regression



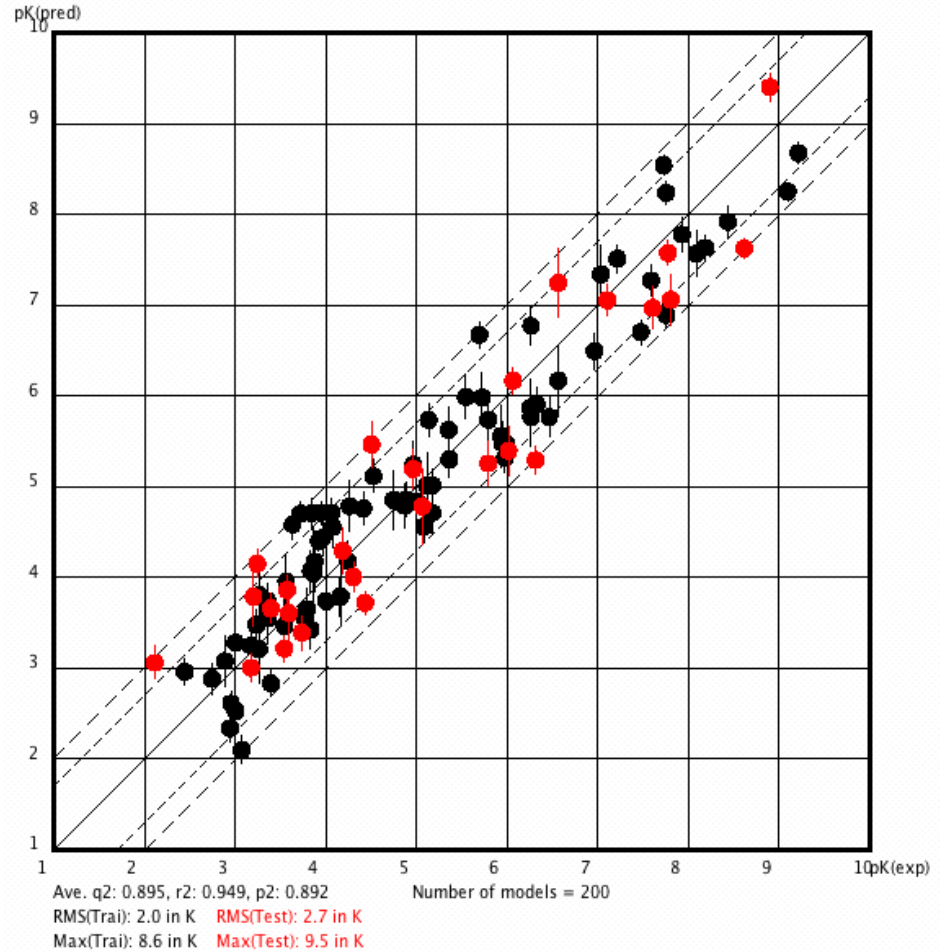
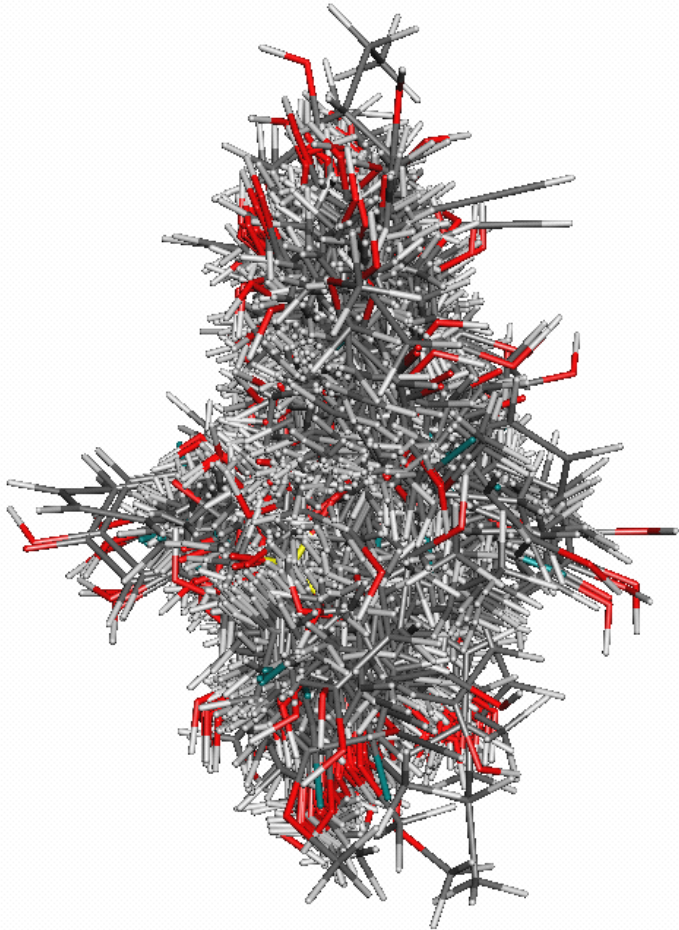
Trainingssubstanzen (12)

Testsubstanzen (4)

**Nach Anwendung der Regression (Steigung a, Achsenabschnitt b)  
können exp. und ber. Affinitäten direkt verglichen werden**



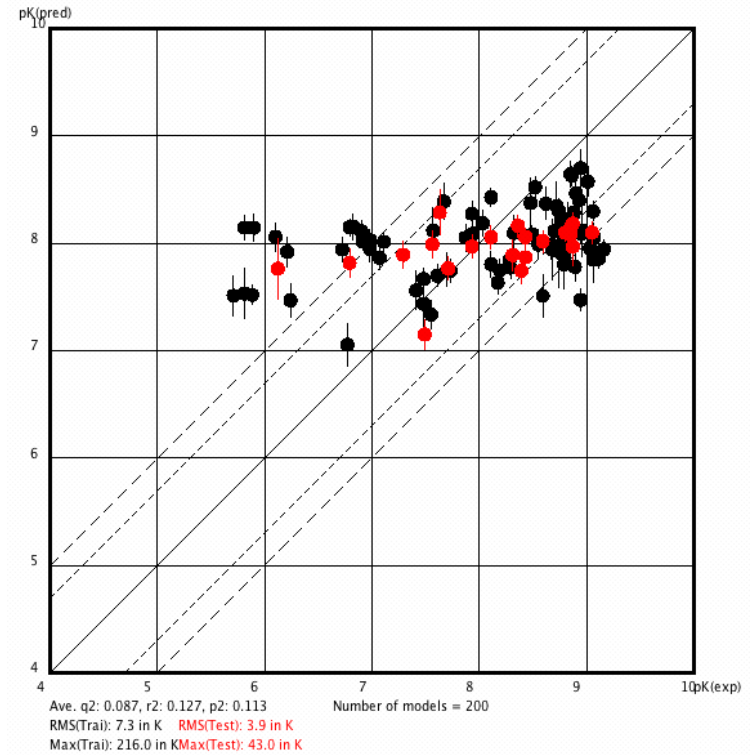
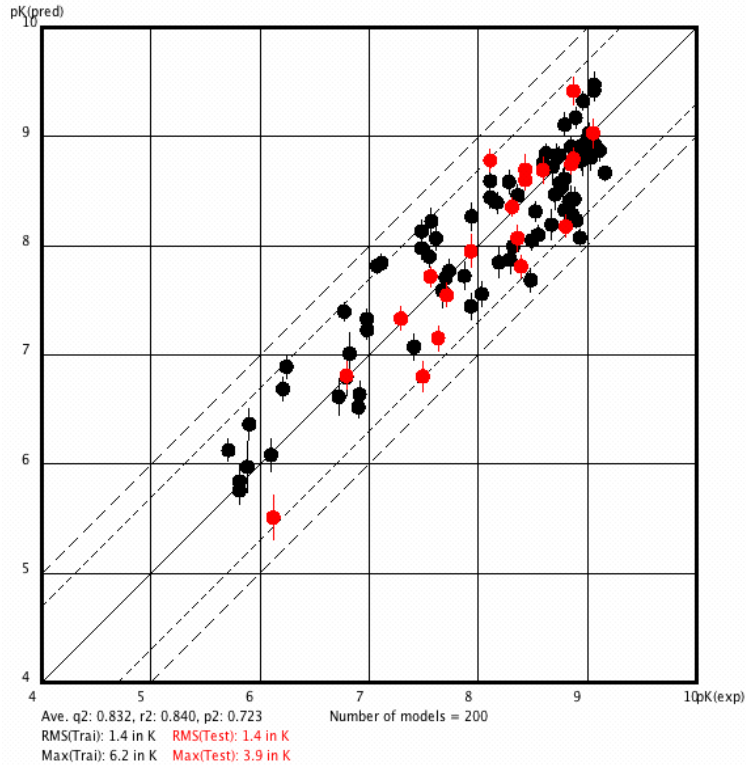
# Validierung I: Trainings- und Testset



*J. Med. Chem.* **2005**, *48*, 3700–3703



## Validierung II: Scramble Test

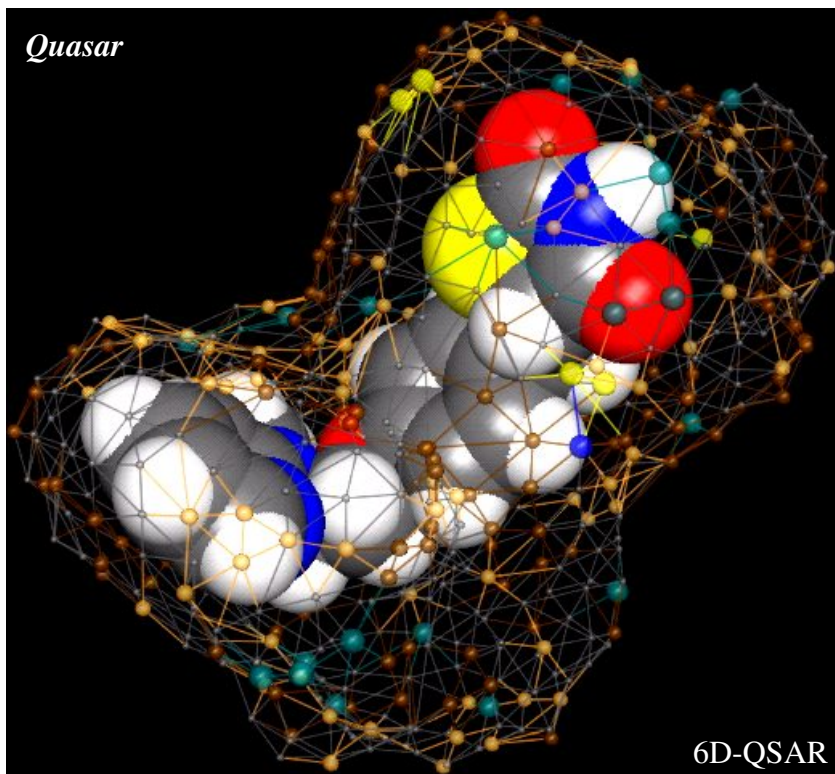


*Tox. Lett.* **2007**, *173*, 17–23.





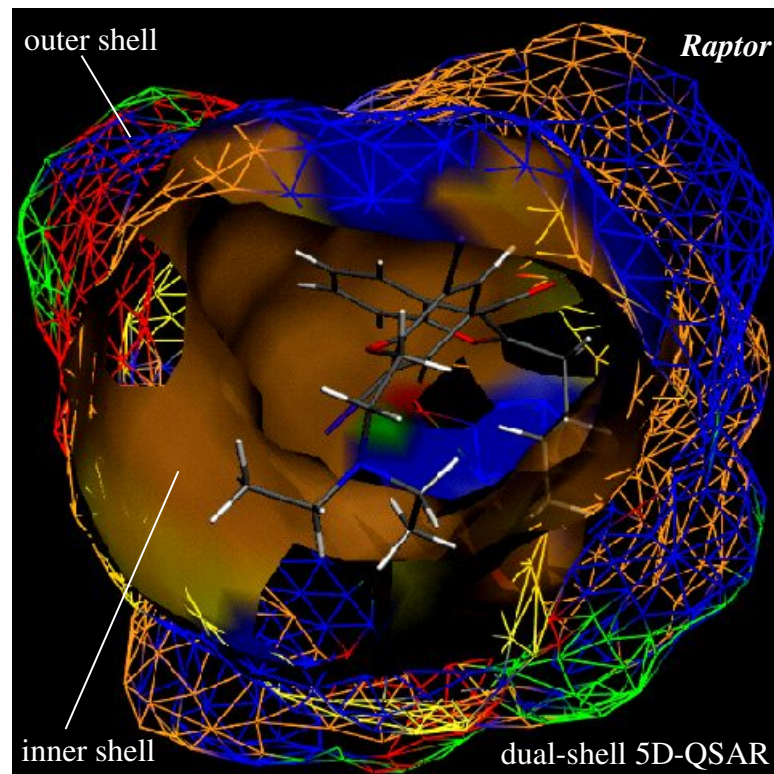
## Validierung III: Consensus Scoring



$$\Delta G_{\text{binding}} \propto E_{\text{prot-lig}} - E_{\text{solv,lig}} - E_{\text{int,lig}} - T\Delta S - E_{\text{IndFit}}$$
$$E_{\text{prot-lig}} = E_{\text{elec}} + E_{\text{vdW}} + E_{\text{HBond}} + E_{\text{polarization}}$$

*J. Med. Chem.* **2002**, *45*, 2139–2149

*J. Med. Chem.* **2005**, *48*, 3700–3703

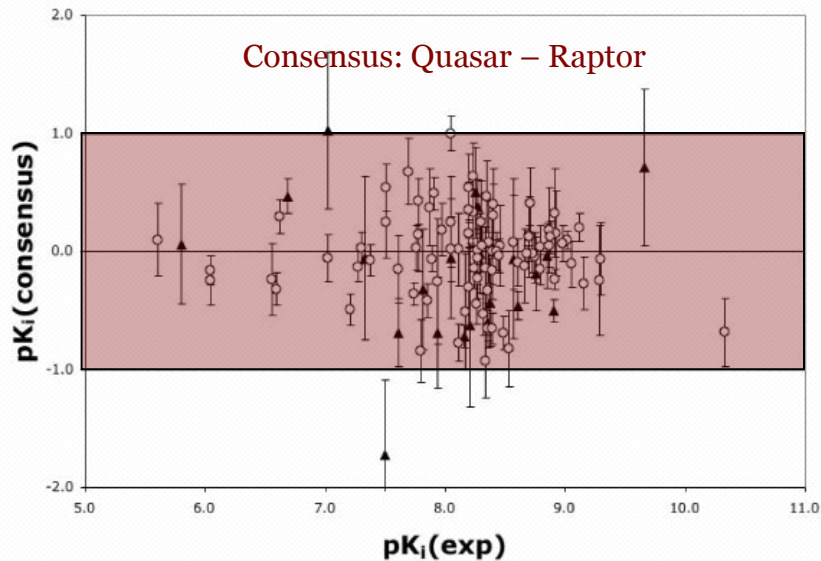
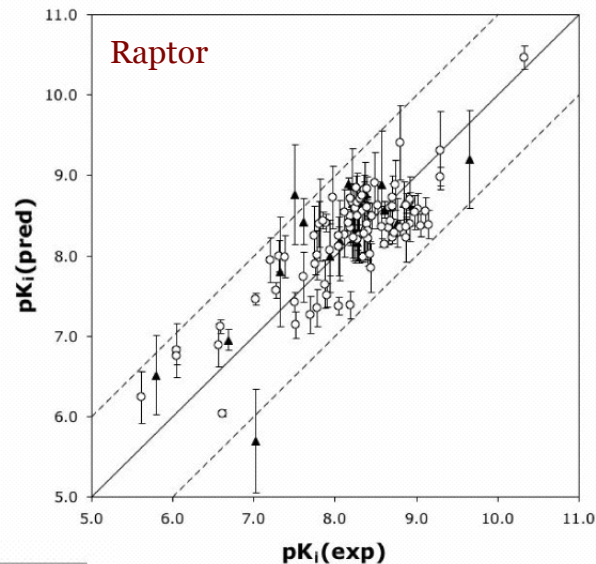
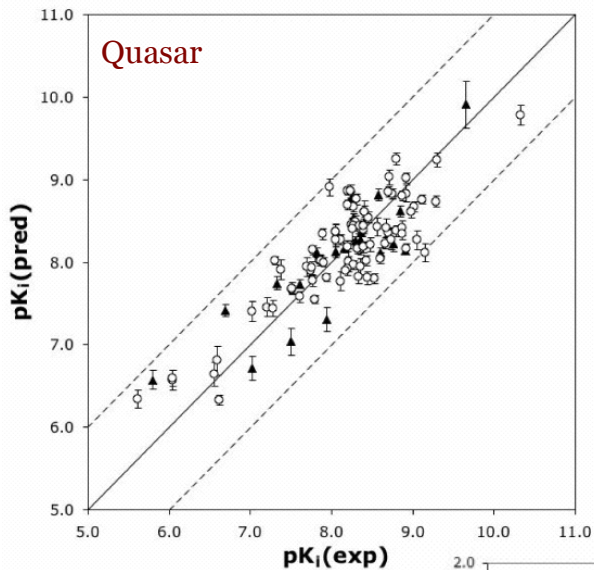


$$\Delta G_{\text{binding}} \propto E_{\text{prot-lig}} - T\Delta S - E_{\text{IndFit}}$$
$$E_{\text{prot-lig}} = E_{\text{HBond}} + E_{\text{hydrophobic}} (\text{shell}_1 + \text{shell}_2)$$

*J. Med. Chem.* **2004**, *47*, 6174–6186



# Consensus Scoring am Glucocorticoid-Rezeptor





## Multi-dimensionales QSAR: mQSAR

<i>Dimension</i>	<i>Methode</i>	<i>Protein</i>
1D-QSAR	Affinität korreliert mit $pK_a$ , logP, molekularem Volumen, etc.	nein
2D-QSAR	Affinität korreliert mit Strukturmotiven oder Verknüpfungsmuster	nein
<b>3D-QSAR</b>	Affinität korreliert mit einer 3D-Struktur der Liganden	möglich
4D-QSAR	Liganden werden als Ensemble von Konformeren, Orientierungen, Protonierungszuständen, Tautomeren und Stereoisomeren repräsentiert	möglich
5D-QSAR	wie 4D + Berücksichtigung verschiedener <i>induced-fit</i> Modelle	ja
6D-QSAR	wie 5D + Berücksichtigung verschiedener Solvationszenarien	ja





## 4D-QSAR: Berücksichtigung verschiedener Bindungsmodi

