

Solvatationsmodelle

Wozu ?

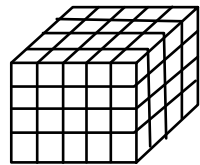
Alle biologisch relevanten Reaktionen finden in wässriger Lösung statt

Vorhersage von molekularen Eigenschaften:

- Solvatationsenergien
- Bindungsaffinitäten in Enzym-Ligand Systemen
- Protonierungszustände von Residuen und Liganden, pK_a

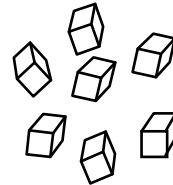
Was ist Solvation ?

Moleküle/Ionen werden durch elektrostatische Kräfte zusammengehalten: Gitterenergie



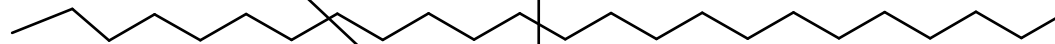
solid

sublimation

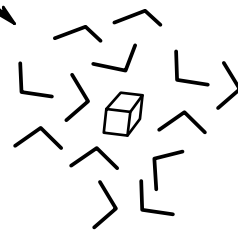


gas / vapour

solvation

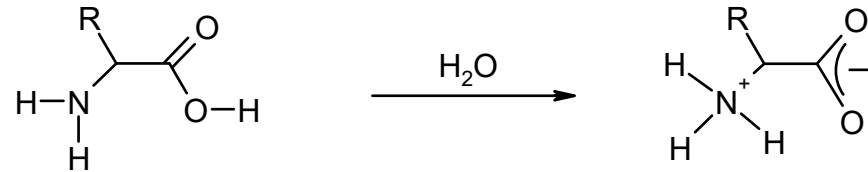


solvent



Elektrostatische Wechselwirkungen mit den Lösungsmittelmolekülen

Aminosäuren in Wasser (1)



Einzelne Aminosäuren dissoziieren in Wasser:

Carbonsäuren als schwache Säuren zu Carboxylaten

Amine als schwache Basen bilden Ammoniumgruppen

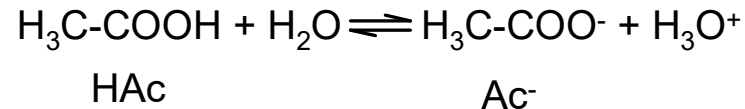
Auch die Seitenketten können ihren Protonierungszustand ändern (titrierbare Gruppen eines Proteins)

Säuren: Asp, Glu (His)

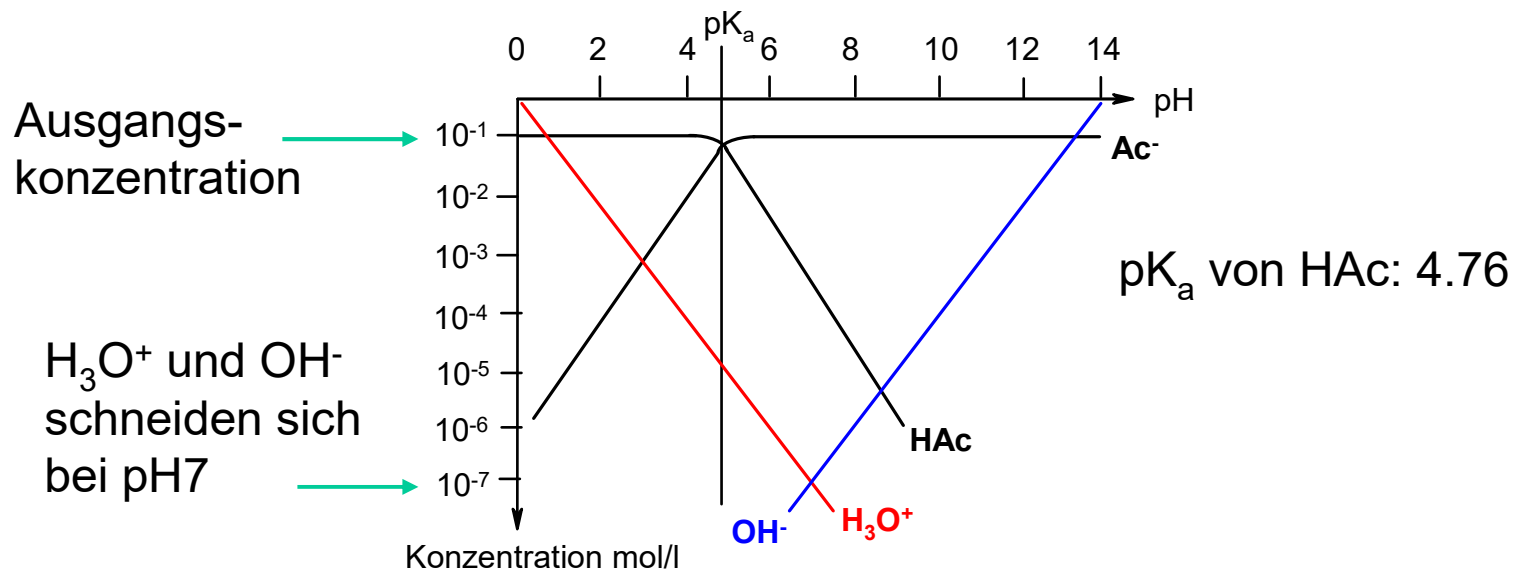
Basen: Lys, Arg

Aminosäuren in Wasser (2)

Wie verhält sich die Carbonsäuregruppe der Seitenkette von Asp oder Glu bei unterschiedlichem pH ?

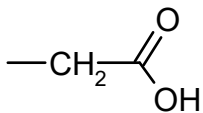
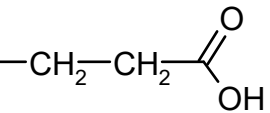
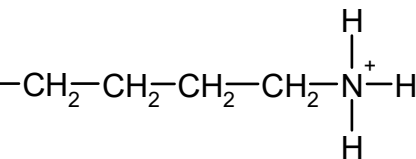
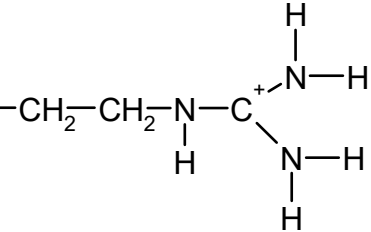
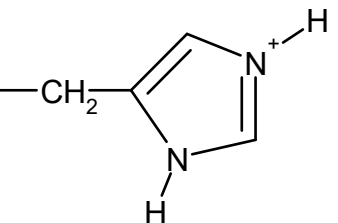


Die Konzentration der jeweiligen Ionen kann aus dem Hägg-Diagramm entnommen werden:



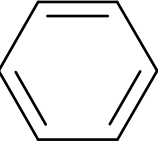
Aminosäuren in Wasser (3)

typische pK_a -Werte von Aminosäureseitenketten
(dargestellt in ihrer protonierten Form als Säure)

Asp		3.9
Glu		4.1
Lys		10.8
Arg		12.5
His		6.04 Protonierung kann sich ändern

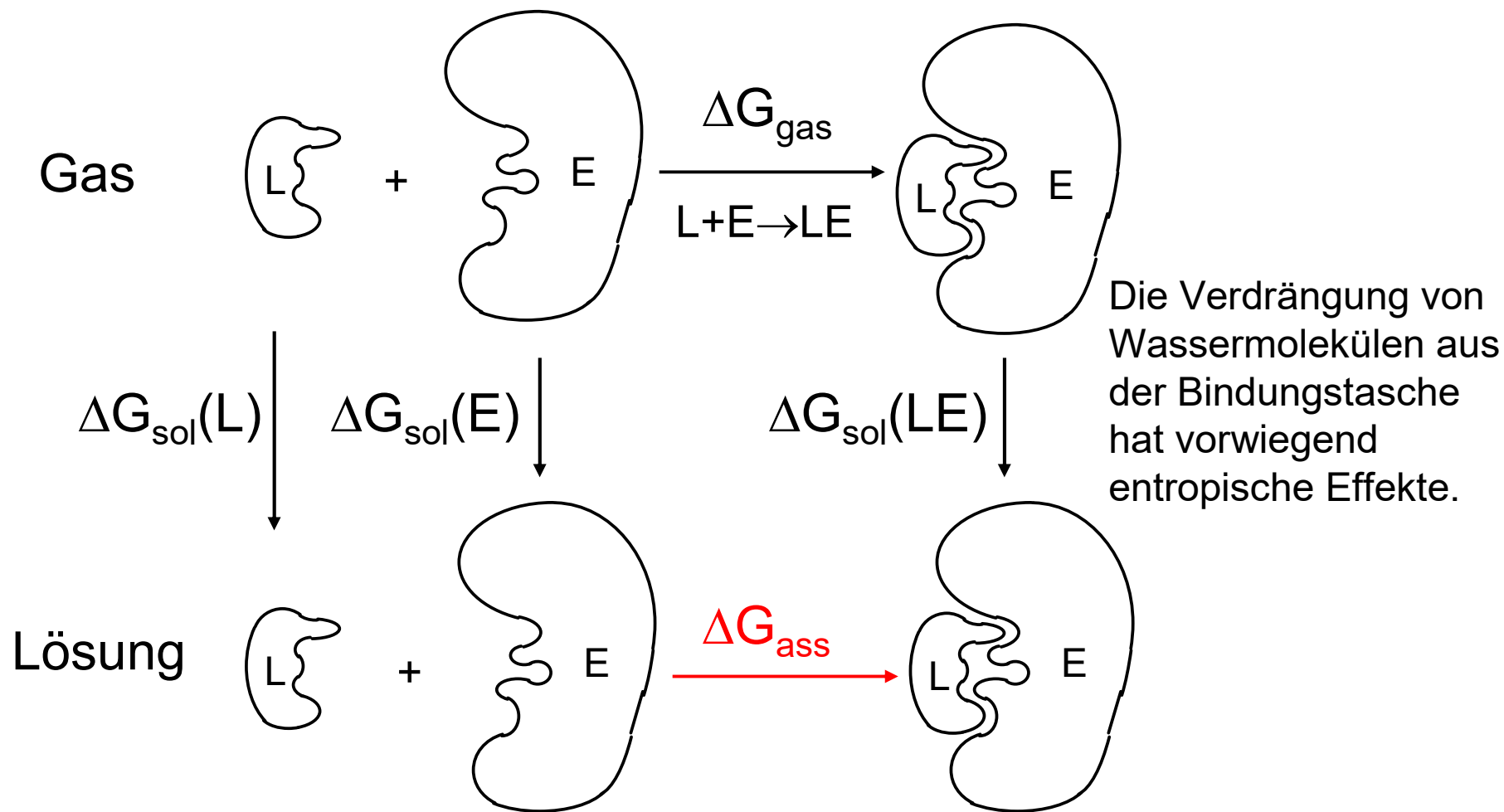
Aminosäuren in Wasser (4)

typische pK_a -Werte von Aminosäureseitenketten
(dargestellt in ihrer protonierten Form als Säure)

Ser	$-\text{CH}_2-\text{OH}$	13	
Thr	$\begin{array}{c} -\text{CH}-\text{OH} \\ \\ \text{CH}_3 \end{array}$	13	
Tyr	$-\text{CH}_2-$  $-\text{OH}$	10.1	
Cys	$-\text{CH}_2-\text{SH}$	8.3	Oxidation zu Disulfidbrücken

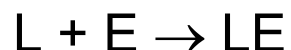
Diese pK_a -Werte wurden für einzelne Aminosäuren in Lösung bestimmt (*bulk properties*). Im Protein kann sich der Wert je nach Umgebung ändern.

Komplexbildung in Lösung (1)



$$\Delta G_{\text{ass}} = \Delta G_{\text{gas}}(\text{L}+\text{E} \rightarrow \text{LE}) + \Delta G_{\text{sol}}(\text{LE}) - \Delta G_{\text{sol}}(\text{L}) - \Delta G_{\text{sol}}(\text{E})$$

Komplexbildung in Lösung (2)



Massenwirkungsgesetz: $\Delta G = -RT \ln K_{on}$

$$= -RT \ln ([LE] / [L] [E])$$

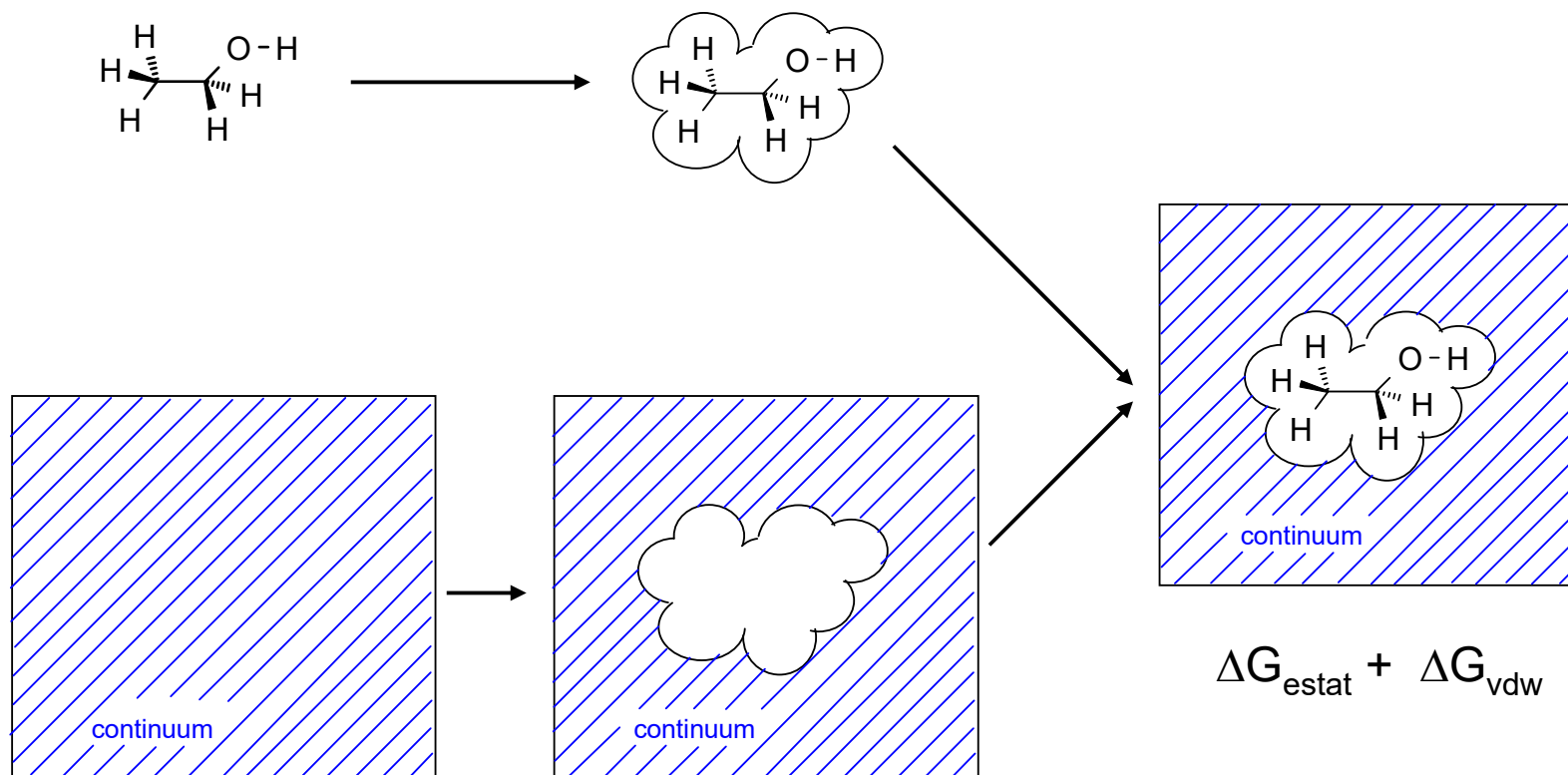
Für pharmazeutische Wirkstoffe/Inhibitoren sollte die Bindungskonstante K_{on} idealerweise im nanomolaren Bereich sein:

+ Höhere Affinität zum entsprechenden Enzym/Rezeptor als das natürliche Substrat

+ nur niedrige Dosierung erforderlich

milli	micro	nano	pico	femto
m	μ	n	p	f
10^{-3}	10^{-6}	10^{-9}	10^{-12}	10^{-15}

Vorgänge bei der Solvatation



$$\Delta G_{\text{cav}} + \Delta G_{\text{estat}} + \Delta G_{\text{vdw}}$$
$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{estat}} + \Delta G_{\text{vdw}}$$

Berechnung der Terme (1)

$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{estat}} + \Delta G_{\text{vdw}}$$

Cavity energy ΔG_{cav} Hohlrumbildungsenergie

- Ausformen einer Kavität gegen den Druck des Solvens
- Reorganisation der Solvensmoleküle

proportional zu der Wahrscheinlichkeit einen entsprechenden Hohlraum zwischen den Lösungsmittelmolekülen zu finden

R. A. Pierotti *Chem.Rev.* **76** (1976) 717.

Berechnung der Terme (2)

$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{estat}} + \Delta G_{\text{vdw}}$$

van der Waals Energie ΔG_{vdw}

- schwache Wechselwirkungen durch fluktuierende elektrische Momente (Dispersion)
- u.a. abhängig von der Polarisierbarkeit

Von Bedeutung falls das Lösungsmittel eine kleine Dielektrizitätskonstante hat ($\epsilon < 10$) und das gelöste Molekül keine Ladung und kein permanentes Dipolmoment hat. Dann ist ΔG_{estat} relativ klein.

Berechnung der Terme (3)

$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{estat}} + \Delta G_{\text{vdw}}$$

Elektrostatische Energie ΔG_{estat}

- starke Wechselwirkung, deshalb dominierender Term

ΔG_{solv} –1-10 kcal/mol für neutrale Moleküle

ca. –100 kcal/mol für Ionen (Na^+ , Ca^{2+} , ...)

In der Regel sind ΔG_{estat} (negativ) und ΔG_{cav} (positiv) größenordnungsmäßig gleich

Solvatationsmodelle (1)

Solvens als strukturloses
Kontinuum

Explizite
Lösungsmittelmoleküle
(Supermolekülansatz)
zumeist in MD-Simulationen

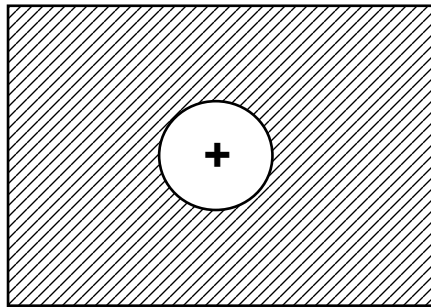
Erste Lösungsmittelhülle explizit,
weiteres Lösungsmittel als Kontinuum

Vorteil von expliziten Wassermolekülen: H-Brücken mit dem Lösungsmittel werden besser erfasst als im reinen Kontinuumansatz, da Wasser nicht isotrop ist.

Alternative: An entsprechenden Stellen platziert man einzelne, explizite Wassermoleküle.

Solvatationsmodelle (2)

Born (1920) Ladung/Ion in einer kugelförmigen cavity



ΔG_{estat} = Arbeit um ein Ion aus dem Vakuum in die Lösung zu bringen. Negativ, da ein einzelnes Ion im Vakuum keine elektrostatischen Wechselwirkungen ausbildet, wohl aber in einem polaren Medium.

$$\Delta G_{\text{estat}} = -\frac{q^2}{2a} \left(1 - \frac{1}{\epsilon} \right)$$

q : Ladung in e

a : Radius der cavity

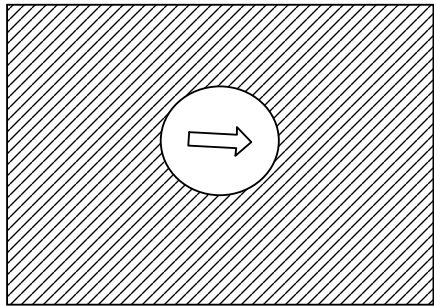
ϵ : Dielektrizitätskonstante (DK)

Wasser = 80 Vakuum = 1

Protein = 2-20

Solvatationsmodelle (3)

Onsager/Kirkwood (1930) Dipol in einer cavity



Das Dipolmoment μ des gelösten Moleküls induziert ein Dipolmoment im Lösungsmittel, das seinerseits ein elektrisches Feld ϕ in der cavity erzeugt, das sog. Reaktionsfeld.

$$\Phi_{RF} = \frac{2(\varepsilon - 1)}{(\varepsilon + 1)a^3} \mu$$

Unter Berücksichtigung der Ladungsverteilung in der cavity erhält man

$$\Delta G_{estat} = -\frac{\Phi_{RF} \mu}{2}$$

Im Falle von geladenen Molekülen kommt noch ein entsprechender Born-Term hinzu.

Solvatationsmodelle (4)

Dieser Reaktionsfeldansatz kann in quantenmechanischen Methoden verwendet werden und zwar als Störung des ursprünglichen Hamiltonians

$$\hat{H}_{tot} = \hat{H}_o + \hat{H}_{RF} \quad \text{wobei} \quad \hat{H}_{RF} = -\hat{\mu}^T \frac{2(\varepsilon - 1)}{(2\varepsilon + 1)a^3} \langle \psi | \hat{\mu} | \psi \rangle$$

Schwachpunkt ist die kugel- oder ellipsoide Form der cavity, die für die meisten Moleküle unzutreffend ist.

- Für willkürliche Formen benutzt man deshalb Punktladungen an der Cavityoberfläche die aus der Wellenfunktion (QM) oder Atomladungen (Kraftfelder) abgeleitet werden. Dies wird als Point Charge Method (PCM) bezeichnet.

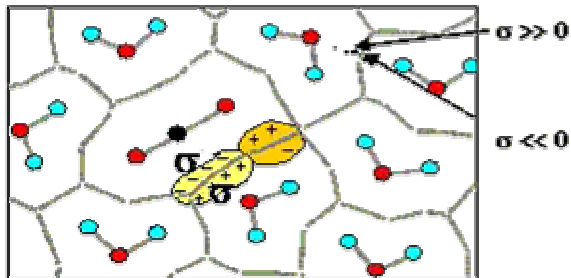
Solvatationsmodelle (5)

Variante der *point charge method*:

In COSMO (A. Klamt) ist die cavity in einem Leiter mit unendlich hohem ϵ untergebracht (metallischer Leiter) und die Ergebnisse werden auf realistische ϵ skaliert.

Verbesserung: COSMO-RS (*real solvent*)

Beschreibt die Wechselwirkung in einer Flüssigkeit als lokale, paarweise Interaktion der Moleküloberflächen



Vorteile: H-Brücken werden besser als in Kontinuumsansätzen beschrieben

Anwendbar auf Mischungen von Flüssigkeiten

http://www.cosmologic.de/theory_background.html

Generalized Born Equation (1)

Kontinuummethode für Kraftfelder, N Atome als Partikel mit Radius a_i und Ladung q_i

Die gesammte freie elektrostatische Energie ist dann die Summe aus Coulomb und freier Born Energie

$$\Delta G_{estat} = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{\epsilon r_{ij}} - \frac{1}{2} \left(1 - \frac{1}{\epsilon} \right) \sum_{i=1}^N \frac{q_i^2}{a_i}$$

In den üblichen Programmen z.B. AMSOL (Cramer & Truhlar) oder OPLS, wird folgende Formel verwendet:

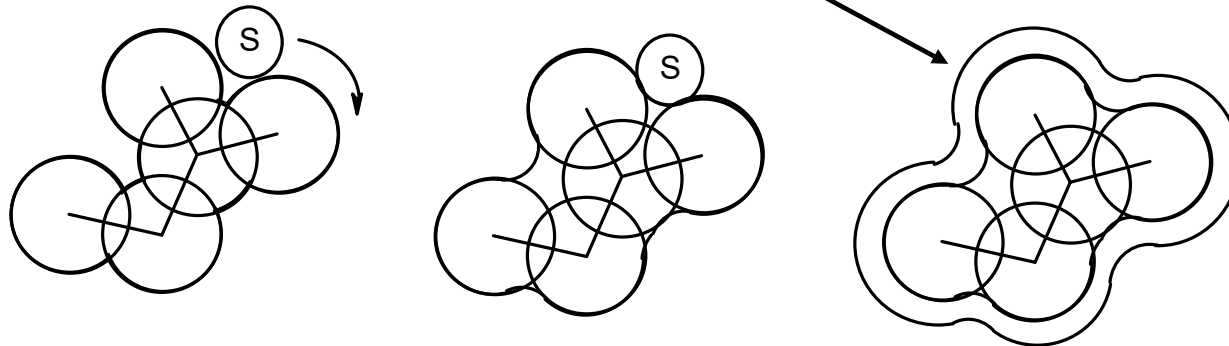
$$\Delta G_{estat} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon} \right) \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{f(r_{ij}, a_{ij})}$$

Generalized Born Equation (2)

Die Funktion $f(r_{ij}, a_{ij})$ berücksichtigt dabei, ob die Ladungen q_i und q_j nahe bei einander sind (\rightarrow Onsager Theorie) oder weit entfernt sind $r \gg a_i$ wobei die Energie dann die Summe aus Coulomb und Born Term ist.

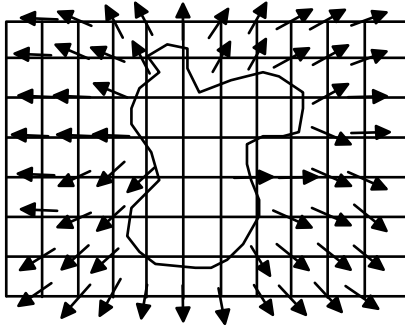
$$f(r_{ij}, a_{ij}) = \sqrt{r_{ij}^2 + a_{ij}^2} e^{-D} \text{ mit } a_{ij} = \sqrt{a_i a_j} \text{ und } D = \frac{r_{ij}^2}{(2a_{ij})^2}$$

Aufwendig ist die Bestimmung des effektiven Born Radius a_i für jede Atomladung im System während der Rechnung. U.a. wird dazu die *solvent accessible area* (SASA) der Atome benötigt.



Langevin Dipol Modell

Warshel & Levitt (1976)

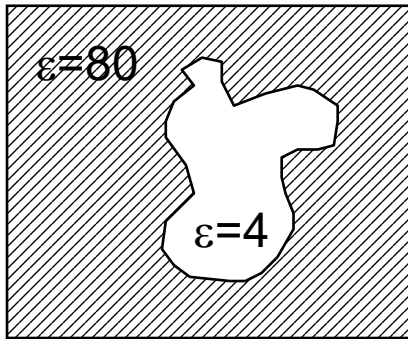


Auf einem dreidimensionalen Gitter werden die Lösungsmittelmoleküle als Dipole μ_i dargestellt, wobei E_i^o das Feld aufgrund des Moleküls allein ist. Damit erhält man die Solvationsenergie als

$$\Delta G_{sol} = -\frac{1}{2} \sum_i \mu_i E_i^o$$

Poisson-Boltzmann Gleichung (1)

Geeignet für große Moleküle (Proteine, DNA-Stränge)



Die Poisson Gl. stellt einen Zusammenhang zwischen dem Potential Φ und der Ladungsdichte ρ in einem Medium mit homogenem ϵ dar

$$\nabla^2 \phi(r) = -\frac{4\pi\rho(r)}{\epsilon}$$

Im Falle von Punktladungen in einem konstanten Dielektrikum geht die Poisson Gl. in das Coulombsche Gesetz über. Variiert man nun ϵ mit der Position r dann ist

$$\nabla \epsilon(r) \nabla \phi(r) = -4\pi\rho(r)$$

Poisson-Boltzmann Gleichung (2)

Kommen nun noch bewegliche Ionen in der Lösung vor, so werden diese mittels einer Boltzmann-Verteilung beschrieben

$$n(r) = N \exp\left(\frac{-V(r)}{k_b T}\right)$$

Zusammen erhält man die Poisson-Boltzmann Gleichung

$$\nabla \varepsilon(r) \nabla \phi(r) - \kappa' \sinh(\phi(r)) = -4\pi\rho(r)$$

κ' erhält man über die inverse Debye-Hückel-Länge κ mit I als der ionischen Stärke der Lösung

$$\kappa'^2 = \frac{\kappa'^2}{\varepsilon} = \frac{8\pi N_A e^2 I}{1000 \varepsilon k_b T}$$

$$I = \frac{1}{2} \sum_i c_i z_i^2$$

c_i Konzentration

z_i Ladungszahl

Poisson-Boltzmann Gleichung (3)

Der *sinh* läßt sich alternativ als Taylor-Reihe darstellen

$$\nabla \varepsilon(r) \nabla \phi(r) - \kappa' \phi(r) \left[1 + \frac{\phi(r)^2}{6} + \frac{\phi(r)^4}{120} + \dots \right] = -4\pi\rho(r)$$

In der linearisierten PB-Gl. wird nur der erste Term verwendet

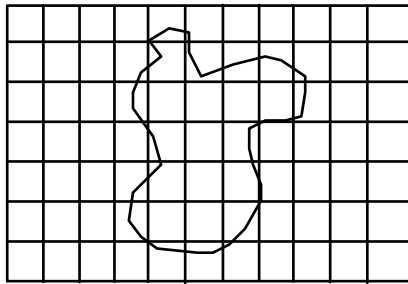
$$\nabla \varepsilon(r) \nabla \phi(r) - \kappa' \phi(r) = -4\pi\rho(r)$$

Analytisch lassen sich nur sehr einfache geometrische Formen der Moleküle lösen:

- Proteine als Kugeln oder Ellipsen (Tanford-Kirkwood)
- DNA als Zylinder
- Membranen als Ebenen (Gouy-Chapman)

Poisson-Boltzmann Gleichung (4)

Als numerische Lösung kommt v.a. die *finite difference* Methode zum Einsatz (Warwick & Watson 1982) z.B. verwendet im Programm DelPhi (Barry Honig)



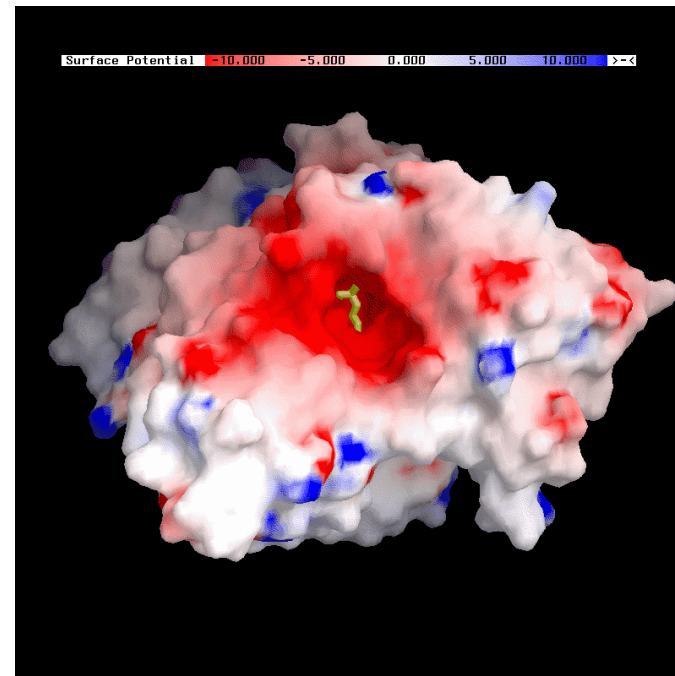
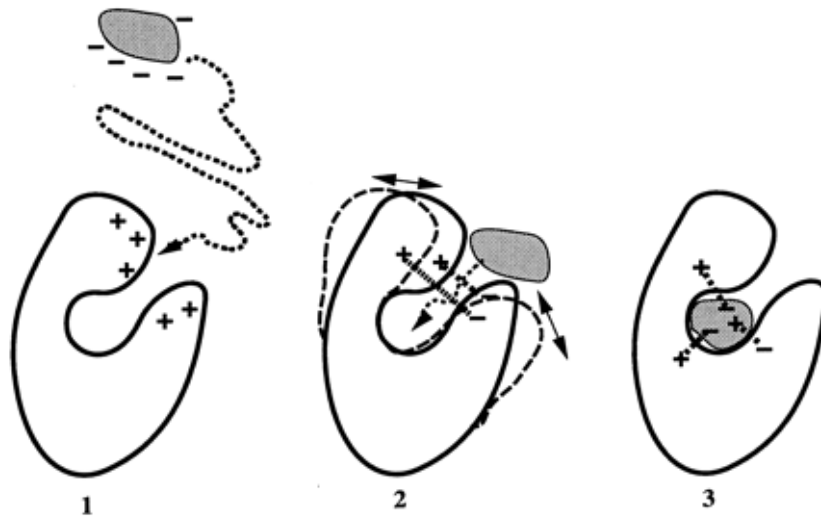
Dazu werden elektrostatisches Potential Φ , Ladungsdichte ρ , Dielektrizitätskonstante ϵ und ionische Stärke I jedem Gitterpunkt zugewiesen. Da sich die Werte gegenseitig beeinflussen, erfolgt die Lösung iterativ.

Entsprechende Atomladungen kann man aus Kraftfeldparametern entnehmen. Umgekehrt lassen sich damit Ladungen für bestimmte Fragmente bestimmen, z.B. Ladung des Eisens in Häm-Gruppen, die je nach Oxidationszustand variiert.

Poisson-Boltzmann Gleichung (5)

Anwendungen:

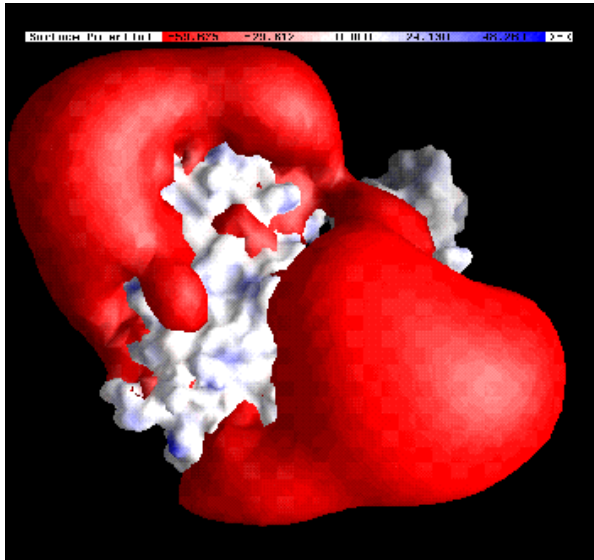
electrostatic steering



Beispiel: Acetylcholin an Acetylcholinesterase

Sehr schnelle Spaltung des Neurotransmitters Acetylcholins

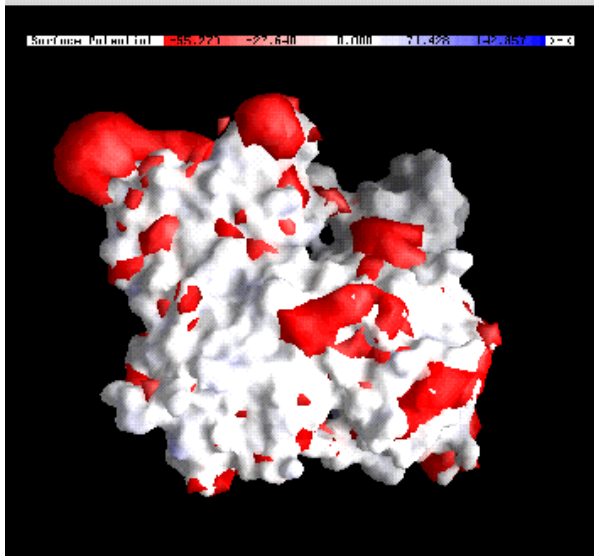
Poisson-Boltzmann Gleichung (6)



Anwendungen:

Elektrostatistisches Potential an der Oberfläche in Abhängigkeit des umgebenden Salzgehaltes

Oben: Isocontourfläche von freiem Actin



Unten: Actin mit 5 gebundenen Ca²⁺-Gegenionen

Poisson-Boltzmann Gleichung (7)

Berechnung des elektrostatischen Beitrags an der Solvatationsenergie

Es sind zwei Simulationen nötig:

1. System in Vakuum ($\epsilon = 1$)
2. System in Lösung ($\epsilon = 80$)

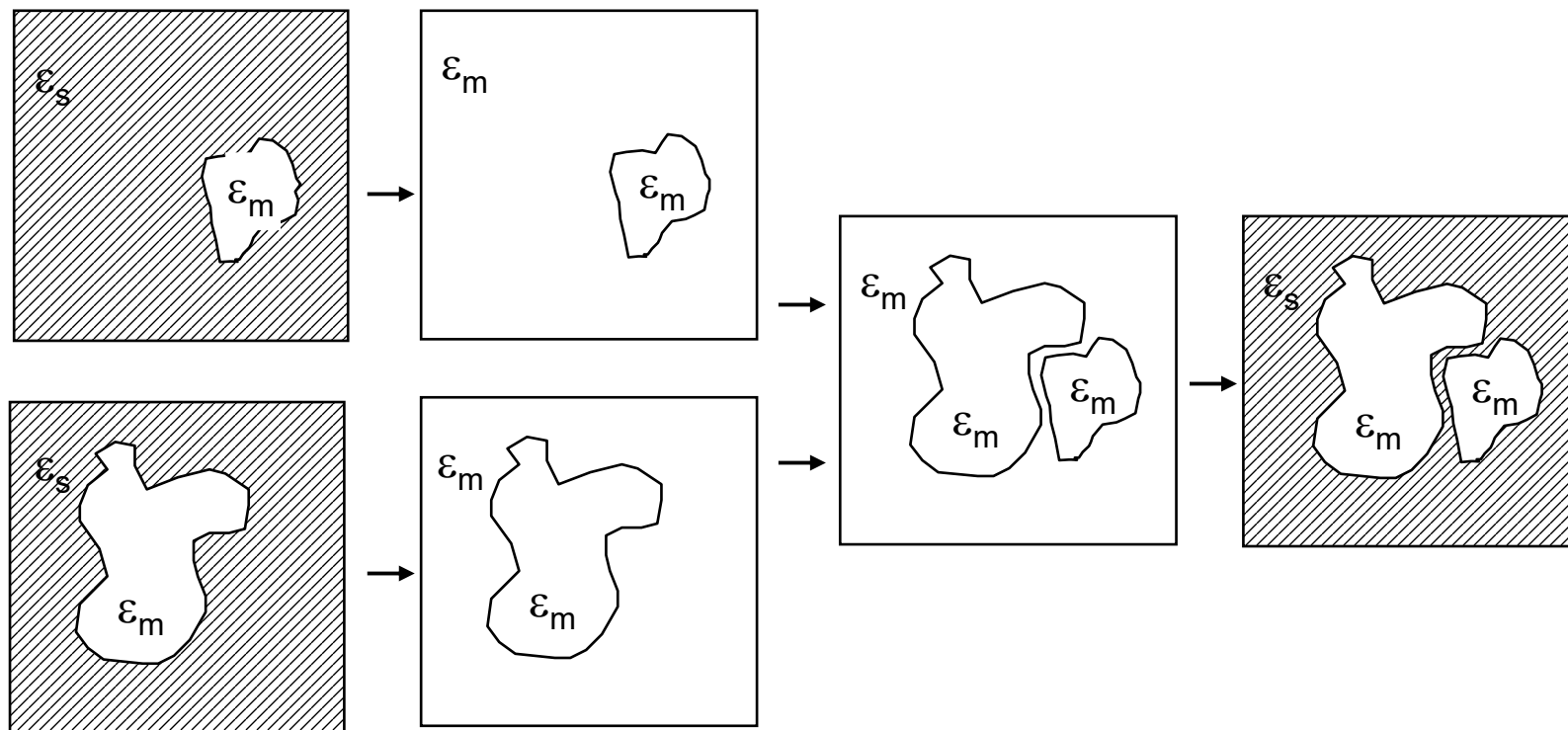
Somit ist $\Delta G_{estat} = \frac{1}{2} \sum_i q_i (\phi_i^{\epsilon=80} - \phi_i^{\epsilon=1})$ für alle Ladungen q_i im System

Typischerweise werden dazu Atomladungen aus Kraftfeldparametern verwendet. Da auch alle H-Atome berücksichtigt werden, müssen die Protonierungszustände der Aminosäureseitenketten bestimmt werden.

z.B. Propka (pdb2pqr) oder ProToss Webserver

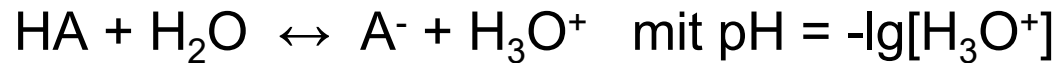
Poisson-Boltzmann Gleichung (8)

Ebenso lässt sich die freie Komplexierungsenergie eines Enzym-Ligand Komplexes berechnen, indem man die einzelnen Molekülsysteme aus der Lösung in ein Medium niedriger DK ($\epsilon_m > 1$) bringt, vereinigt und anschließend zusammen solvatisiert



Poisson-Boltzmann Gleichung (9)

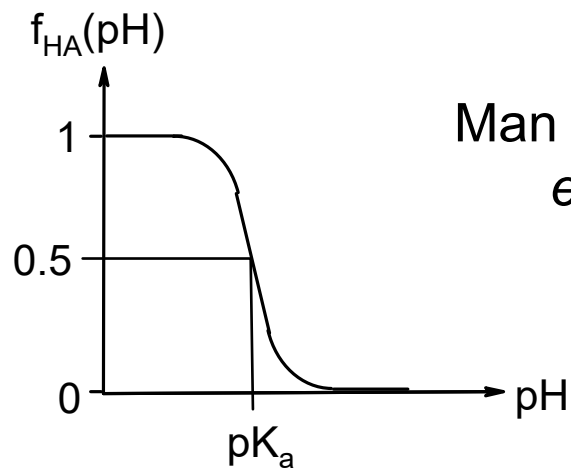
pK_a Berechnung (z.B. von Aminosäuren in Proteinen)



Henderson-Hasselbach: $pH = pK_a + \lg([A^-]/[H_3O^+]) = pK_a + \lg K_G$

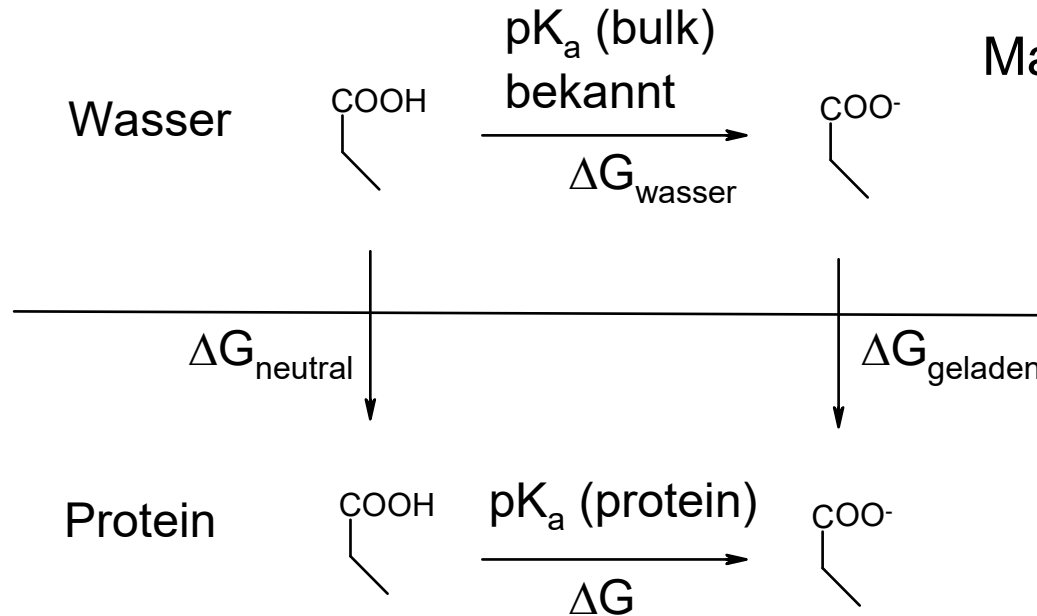
Durch Umformen erhält man:

$$f_{HA}(pH) = \frac{[HA]}{[HA] + [A^-]} = \frac{1}{K_G + 1} = \frac{1}{10^{pH - pK_a} + 1}$$



Man erhält so eine virtuelle Titrationskurve
einer Gruppe/Seitenkette

Poisson-Boltzmann Gleichung (10)



Man erstellt eine Titrationskurve in Abhängigkeit eines äußeren pH und ermittelt so den pK_a der Residue

$$\lg K = pH - pK_a$$

$$\Delta G = -kT \ln K$$

$$\Delta G_{(\text{geladen}-\text{neutral})} = -kT \ln(pH - pK_a)$$

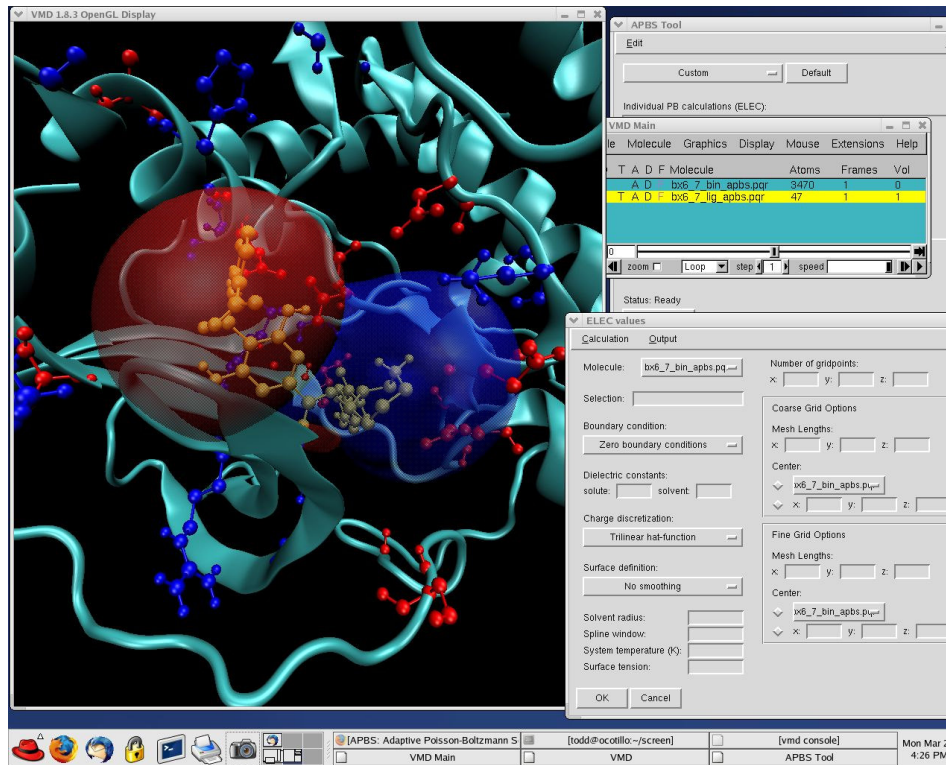
$$pK_a = -\lg \frac{[\text{H}_3\text{O}^+][\text{A}^-]}{[\text{HA}]} = \frac{\Delta G_{\text{diss}}^0}{RT \ln 10}$$

$$pK_a = \frac{\Delta G^0(\text{A}^-) + \Delta G^0(\text{H}_3\text{O}^+) - \Delta G^0(\text{HA})}{RT \ln 10}$$

Schwierigkeit: Wechselwirkungen mit anderen titrierbaren Residuen in unmittelbarer Nachbarschaft.

Poisson-Boltzmann Gleichung (11)

APBS: Adaptive Poisson Boltzmann Solver
www.poissonboltzmann.org/apbs



Benutzt das Finite Element Toolkit um die Poisson-Boltzmann Gleichung numerisch zu lösen.

Nathan Baker, J. Andrew McCammon, Michael Holst *et al.*

Zusammenfassung Solvation

Ionen sind im Wasser von einer Solvationshülle umgeben, und damit energetisch günstiger als im Vakuum (Solvationsenergie). Deshalb sind entsprechende Aminosäureseitenketten auch allermeistens im geladenen Zustand anzutreffen (an der Proteinoberfläche, bzw. gepaart als Salzbrücken im Protein-Inneren).

Der zahlenmäßig größte Beitrag zur Solvationsenergie ist der Born Term (Monopol). Dispersive Beiträge (Dipol-Dipol wie bei van der Waals WW) spielen vor allem in unpolaren Medien (kleine Dielektrizitätskonstante) eine Rolle.

Da auch entropische Beiträge (teilweise) in die Solvationsenergie einfließen, ist es schwierig gute Übereinstimmungen mit experimentellen Werten zu bekommen (z.B. Bindungsaffinitäten).

Docking und Scoring Funktionen (1)

Ziel: Vorhersage der biologischen Affinität einer Verbindung

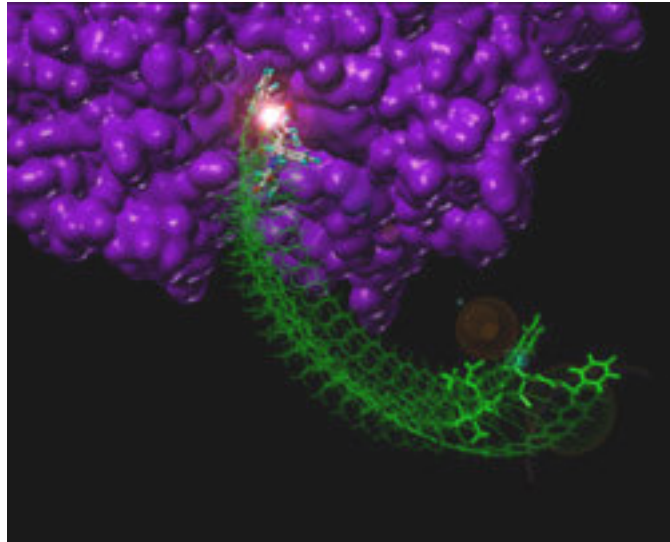
Warum: Bei der Suche nach neuen pharmakologischen Wirkstoffen wird zunächst eine virtuelle Bibliothek von mehr als 100.000 denkbaren Verbindungen konstruiert. Von diesen bleiben auch nach Anwendung von ADME Filtern noch tausende übrig die in die Bindungstasche des *target* Proteins passen könnten. Nur ein Bruchteil davon kann auch tatsächlich im Labor synthetisiert werden. → Docking Algorithmen notwendig

Das virtuelle Docking von potentiellen Wirkstoffen an ein Enzymmodell ist der Flaschenhals im Computer Aided Drug Design.

Für ca. 50% der *target* Proteine von denen man die Struktur in molekularem Detail kennt, gibt es noch keine geeigneten Wirkstoffe. → Nicht alle Proteine sind brauchbare *targets*.

Docking und Scoring Funktionen (2)

Docking ist rechenzeitaufwendig da es ein globales Minimumproblem darstellt:



starrer Ligand an starres Protein: 6 Variablen bzw. Freiheitsgrade

flexibler Ligand an starres Protein: $6 + n \cdot \sigma$ Bindungen

flexibler Ligand an flexibles Protein: $6 + n \cdot \sigma + \dots$

Docking und Scoring Funktionen (3)

→ Docking benötigt eine schnell zu berechnende Bewertungsfunktion
Es bieten sich **Energiefunktionen** an, wie sie auch in Kraftfeldern verwendet werden:

Intra-Ligand Konformationsterme (Bindungen, Winkel, Torsionen)

Nicht-bindende Wechselwirkungen zwischen Ligand und Enzym:

- Van der Waals Wechselwirkung (sterisch)
- Elektrostatische Wechselwirkung (Salzbrücken)
- Wasserstoffbrücken (vorwiegend elektrostatisch)
- Solvationsterme (Desolvatation, SASA)
- Entropische Terme (Verlust der translatorischen und rotatorischen Freiheitsgarde des gesamten Liganden, sowie der Rotationen um Einfachbindungen bei der Bindung ans Enzym)

z.B. in FlexX und AutoDock verwendet

M. Rarey et al. *J. Mol. Biol.* **261** (1996) 470

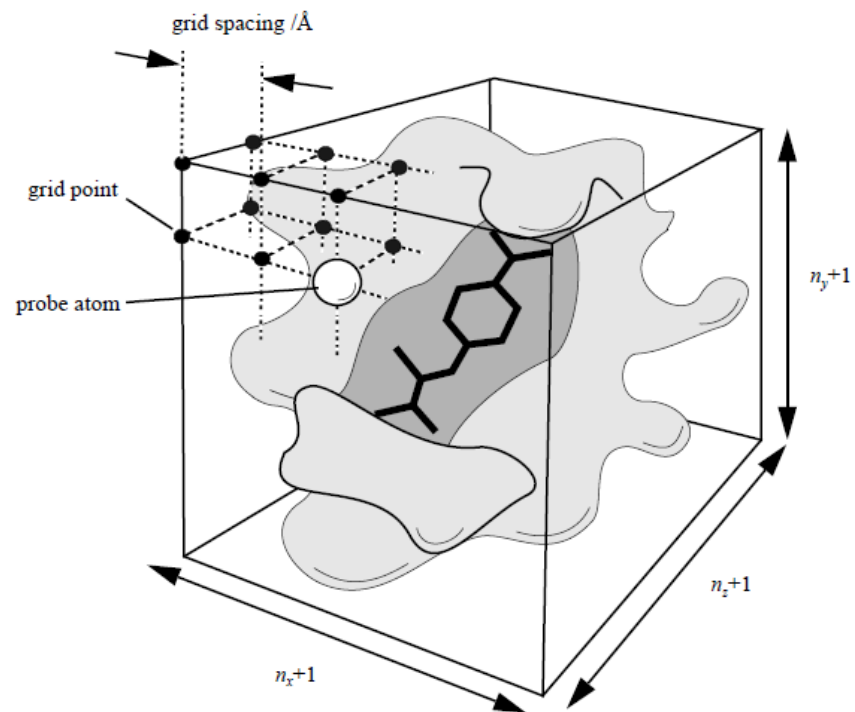
Docking und Scoring Funktionen (4)

Die Berechnung der nicht-bindenden Wechselwirkungen (WW) zwischen Ligand- und Proteinatomen kosten am meisten Rechenzeit.

Um diese zu minimieren kann man die Proteinatome auf regelmäßige Gitterpunkte „mappen“. Dadurch müssen nur noch die Interaktionen der Ligandenatome mit den Gitterpunkten berechnet werden.

AutoDock verwendet dazu für jeden Atomtyp (probe atom) ein eigenes Gitter (map) für die van der Waals WW. Die elektrostatischen WW bekommen ein weiteres Gitter.

Zudem berücksichtigt man nur die polaren H-Atome. Deren H-Brückenenergien werden separat berechnet.



Quelle: AutoDock User's Guide

Docking und Scoring Funktionen (5)

Hinzu kommen bei AutoDock entropische Terme die Solvatationsenergien und den Verlust translatorischer Freiheitsgrade (Drehung um Einfachbindungen) berücksichtigen.

Beispielausgabe von AutoDock (Version 4.0):

```
Estimated Free Energy of Binding      =   -9.35 kcal/mol  [= (1) + (2) + (3) - (4) ]
Estimated Inhibition Constant, Ki     =  140.33 nM (nanomolar)   $\Delta G = RT \ln Ki$ 
  [Temperature = 298.15 K]
(1) Final Intermolecular Energy      =   -7.73 kcal/mol
    vdW + Hbond + desolv Energy      =   -6.52 kcal/mol
    Electrostatic Energy              =   -0.11 kcal/mol
    Moving Ligand-Fixed Receptor      =   -6.63 kcal/mol
    Moving Ligand-Moving Receptor     =   -1.10 kcal/mol
(2) Final Total Internal Energy       =   -1.66 kcal/mol
    Internal Energy Ligand             =    +0.00 kcal/mol
    Internal Energy Receptor           =   -1.66 kcal/mol
(3) Torsional Free Energy              =    +0.00 kcal/mol
(4) Unbound System's Energy          =   -0.04 kcal/mol
```

Vorteil solcher Energiebasierter Scoringfunktionen: Die energetischen Einzelbeiträge sind einfach zu berechnen.

Docking und Scoring Funktionen (6)

Prinzipielles Problem von Energiebasierten Scoring Funktionen:

Leichte Ungenauigkeiten führen zu großen Fehlern bei der Vorhersage der Bindungskonstanten

Massenwirkungsgesetz: $\Delta G = -RT \ln K$

5.7 kJ/mol (= 1.4 kcal/mol) bewirken Faktor 10 für K

Weiteres Probleme:

- Die Genauigkeit von berechneten Solvatationsenergien liegt bei etwa 1-2 kcal/mol
- Entropische Effekte (Verlust von Freiheitsgraden) lassen sich nur schwer abschätzen (ca. 1 kcal/mol pro rotierbare Bindung)

Idee: Die Scoring Funktion wird an einer Reihe von experimentell bekannten Enzym-Ligand Komplexen kalibriert/gefittet.

H.J. Böhm, *J. Comp. Aid. Mol. Des.* **8** (1994) 243.

Nachteil: Nicht für „neue“ Enzymsysteme verwendbar

Docking und Scoring Funktionen (7)

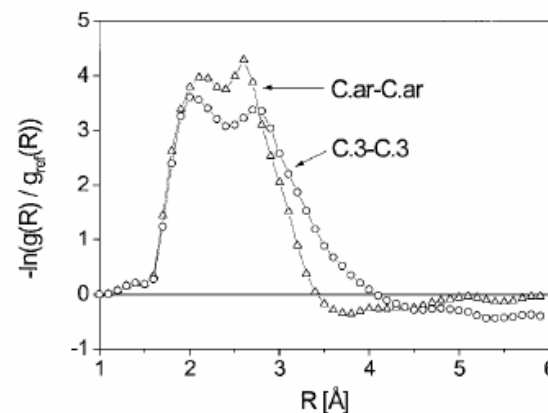
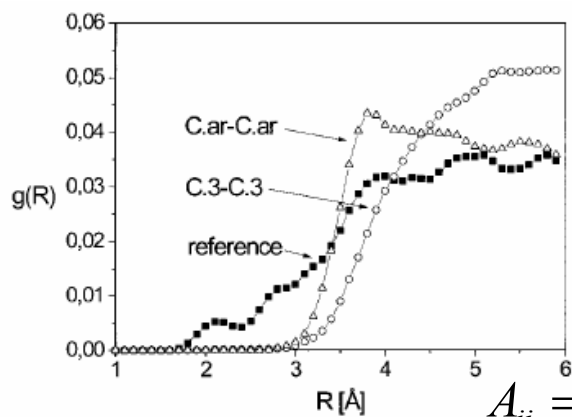
Wissensbasierte Scoring Funktionen

Idee: Ähnliche Typen von Ligandenatomen werden häufiger zu bestimmten Atomtypen im Protein benachbart sein.

Lit. I. Muegge & Y.C. Martin *J. Med. Chem.* **42** (1999) 791

Atomtypen für C: aliphatisch, aromatisch, polar, unpolar, etc. Wesentlich mehr Atomtypen für Liganden als für Proteinatome

Aus der radialen Verteilungsfunktion $g(R)$ zwischen Atomtypen-paaren ij wird ein *potential of mean force* $A(ij)$ erstellt.



$$A_{ij} = -k_B T \ln \left(\frac{\rho_{seg}^{ij}(r)}{\rho_{bulk}^{ij}} \right)$$

$$PMFscore = \sum_{kl} A_{ij}(r_{ij})$$

Ähnlich: Drugscore Gohlke & Klebe *J.Mol.Biol.* **295** (2000) 337,
J.Med.Chem. **48** (2005) 6296.

Docking und Scoring Funktionen (8)

Vergleich der Performance von Scoring Funktionen

G. Keserü *J. Comp.-Aided Mol. Design* **15** (2001) 649

Vergleich anhand bekannter Cytochrom P450cam Substrate

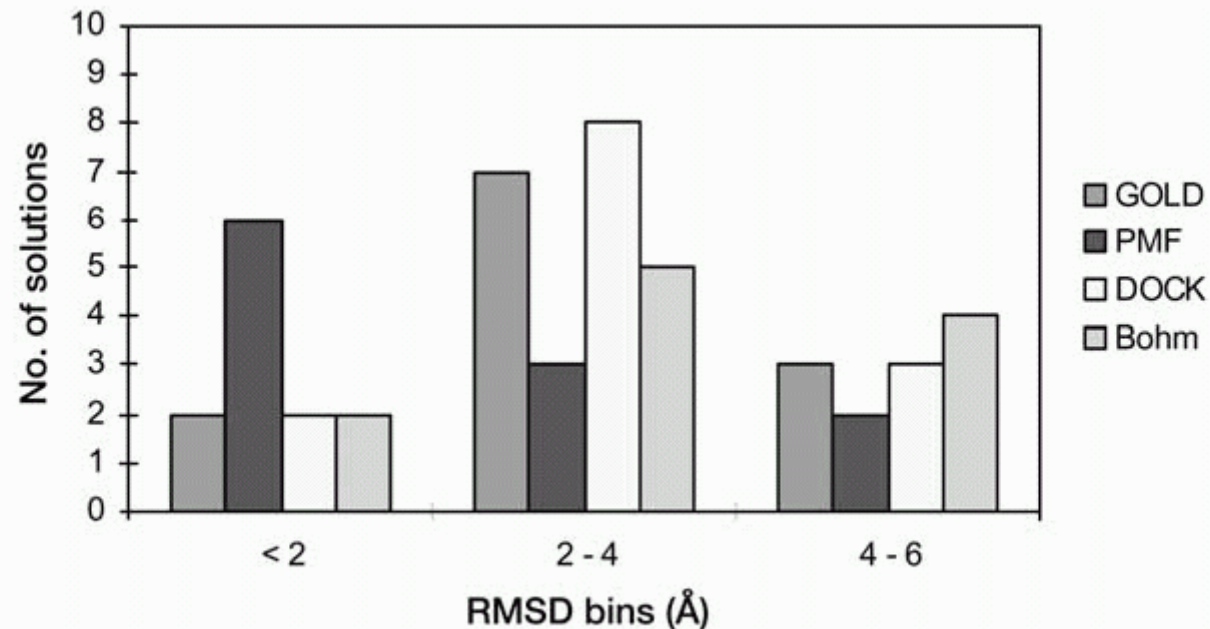


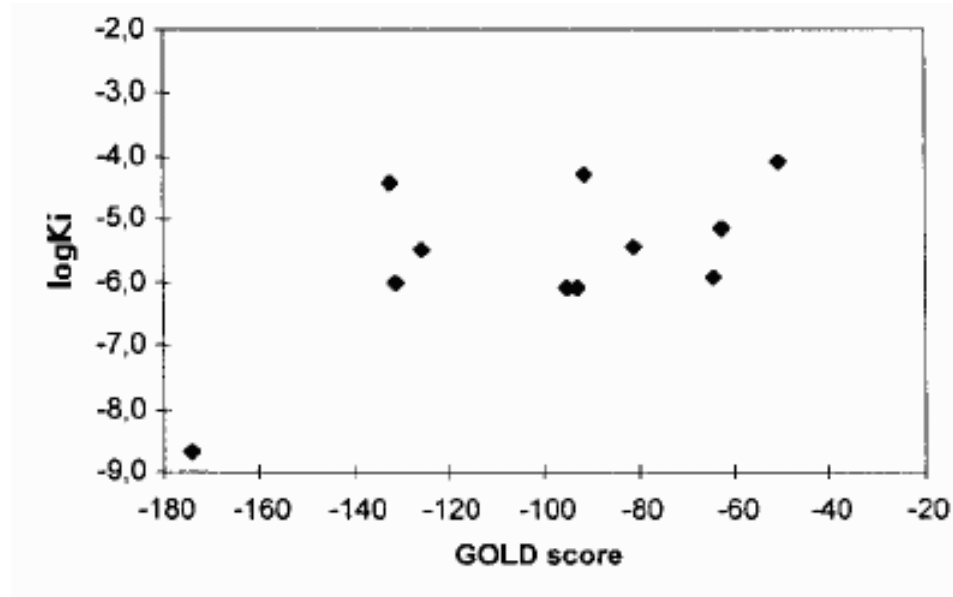
Figure 2. Frequency distributions of the rmsd values with respect to the X-ray conformation of 11 flexibly docked CP450cam ligands.

Docking und Scoring Funktionen (9)

Vergleich der Performance von Scoring Funktionen

G. Keserü *J. Comp.-Aided Mol. Design* **15** (2001) 649

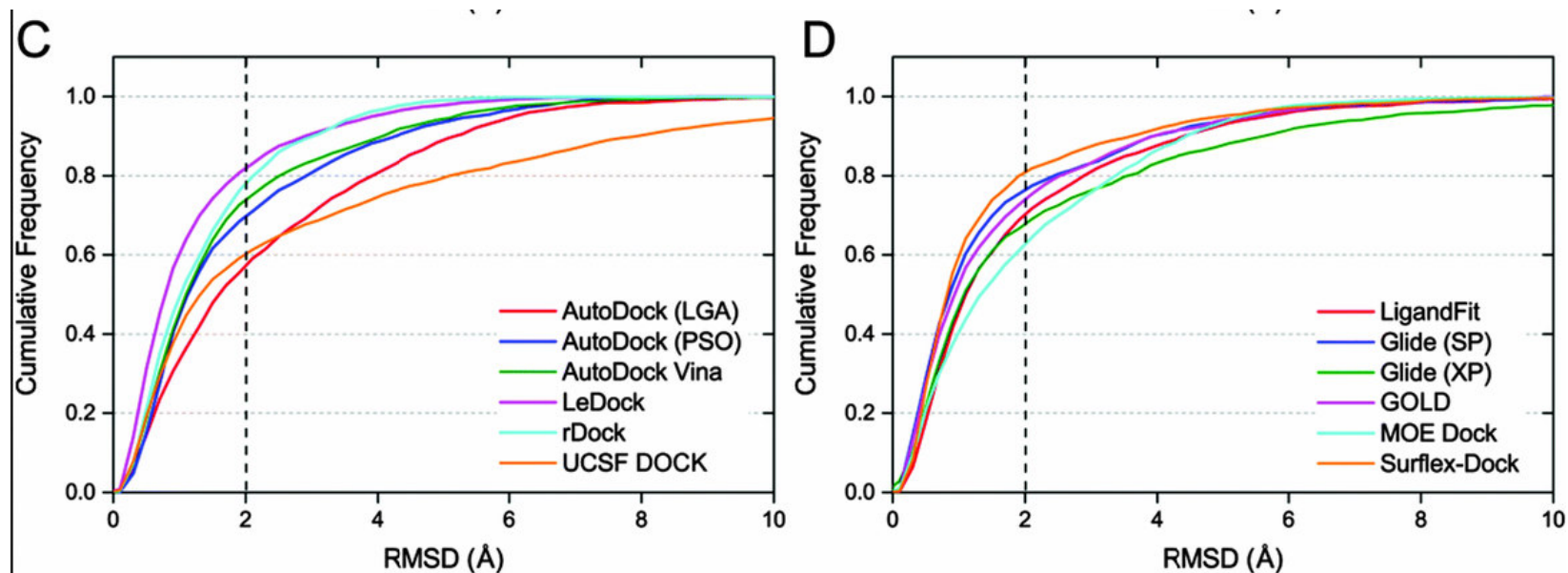
Vergleich anhand bekannter Cytochrom P450cam Substrate



Vergleich berechneter und tatsächlicher Bindungsaffinität. (Pearson) Korrelation ist relativ niedrig. Sinnvoller ist eine *rank correlation*: Trennung von hochaffinen und schwachbindenden Liganden.

Docking und Scoring Funktionen (10)

Vergleich von 10 populären Dockingprogrammen,
Redocking von Liganden



Häufigkeit erhaltener Konformationen in Abhängigkeit der
RMSD-Abweichung zur Kristallstruktur

Lit: Wang et al. Phys.Chem.Chem.Phys. **18** (2016) 12964.

Docking und Scoring Funktionen (11)

Fazit: Kommerzielle Dockingprogramme sind nicht „signifikant“ besser als solche aus der *public domain*. Die Unterschiede kommen vorwiegend aufgrund der zugrundeliegenden Kalibrierung an unterschiedlichen *benchmark sets* zustande.

Deshalb: Consensus scoring

Man bildet die Summe aus verschiedenen Scoring Funktionen.
z.B. die „beste“ gefundene Konformation aus Programm A wird auch mit den Scoringfunktionen von Programm B, C und D bewertet.

Für jede der verwendeten Scoring Funktionen wird eine Schranke definiert, die affine (1) und nicht-affine Liganden (0) separiert.
Das Dockingergebnis für jeden Liganden wird anhand dieser Schranke bewertet (0 oder 1).

Diese Ganzzahlen werden für jede Scoringfunktion aufaddiert.

Beispiel: Erhält man mit 4 unterschiedlichen Scoringfunktionen mindestens 3 als Ergebnis, so wird der Ligand als bindend betrachtet.

Docking und Scoring Funktionen (12)

Performance und Benchmark-Sets für Docking-Programme:

Directory of Useful Decoys DUD-E <http://dude.docking.org/>

Mysinger et al. *J. Med. Chem.* **55** (2012) 6582.

102 targets, viele hochaffine Liganden (aus der ChEMBL Datenbank), sowie ähnliche Moleküle die aber nicht binden, sog. decoys

CCDC/Astex Dataset

Nissink et al. *Proteins* **49** (2002) 457. 205 Protein-Ligand Komplexe

Probleme:

- Liganden sind vorwiegend hochaffine Inhibitoren und nicht Substrate
- Auch X-ray Daten guter Auflösung können fehlerhafte Liganden aufweisen: falsche Konformation von Ringsystemen, falsche Stereochemie, fehlende Atome
- Rotationen werden durch den RMSD Fehler überbewertet