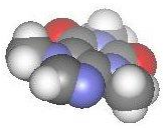


Software und Datenbanken in der Biochemie

PD Dr. Eric Fontain
Technische Universität München
Fakultät für Chemie





■ Molekülstrukturen am Computer

- Discovery Studio Viewer
- Stereobilder
- POV-Ray

■ Molekülstrukturen im Internet

- Organische Moleküle mittlerer Größe (Cambridge Structural Database)
- Proteine (Protein Database)

■ Einfaches Molecular Modeling

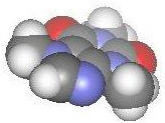
- Molecular Mechanics mit HyperChem
- Potentielle Energie
- Strukturoptimierung
- Molecular Dynamics
- Simulated Annealing
- Periodic Boxes

■ Literatur Online

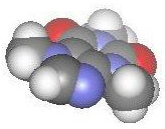
- EZB, SciFinder, Reaxys, MedLine etc.

■ Protein Datenbanken und Services

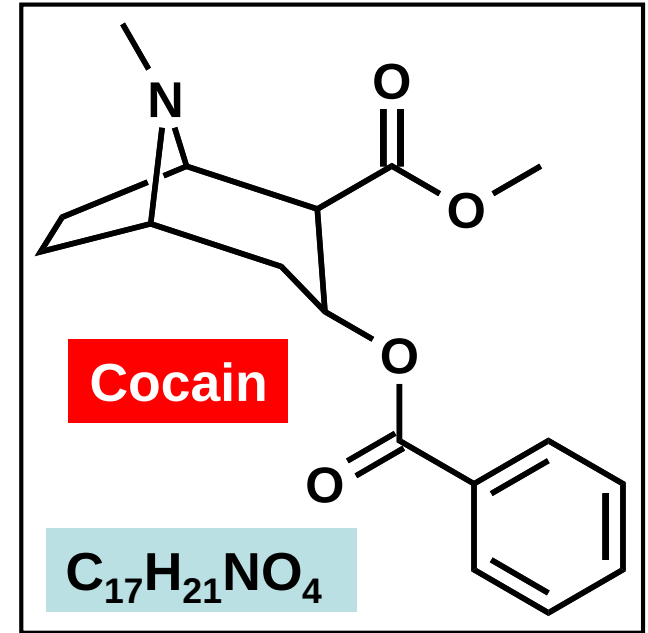
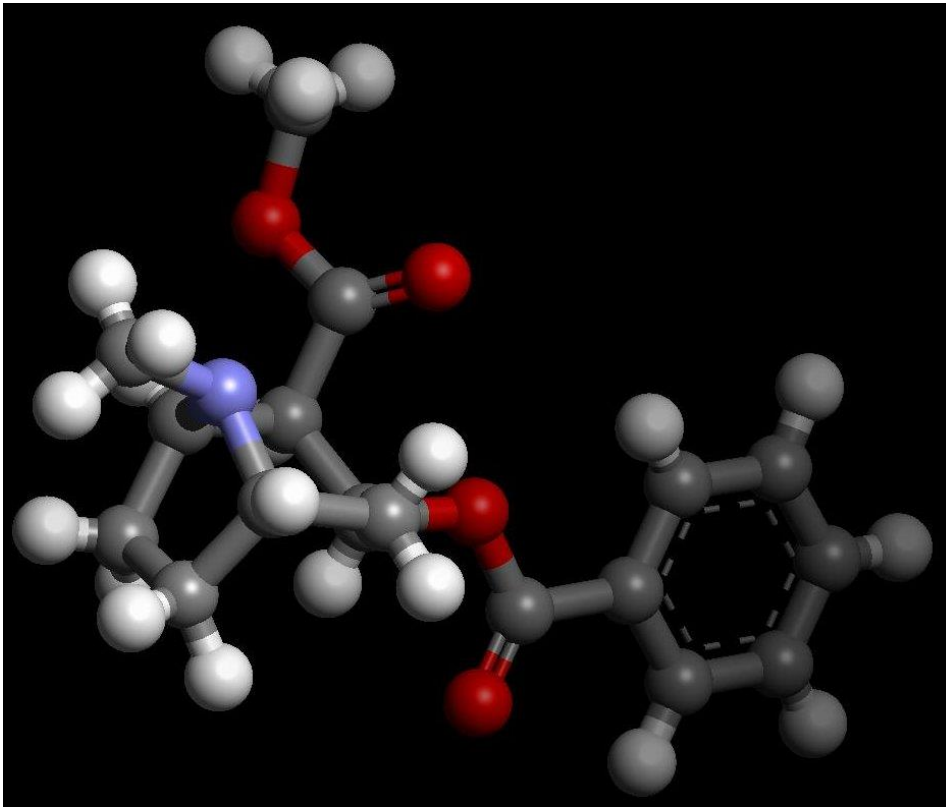
- CATH, PROSITE, PROMOTIF, ExPASy etc.
- Enzyme Commission Numer (E.C.)
- Consensus Pattern



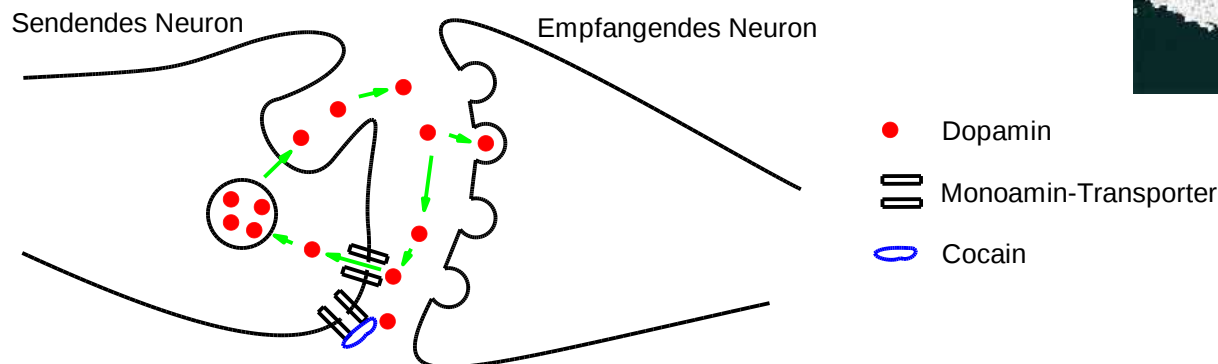
Moleküle am Computer

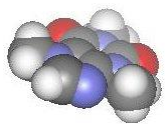


Struktur erzeugt Wirkung

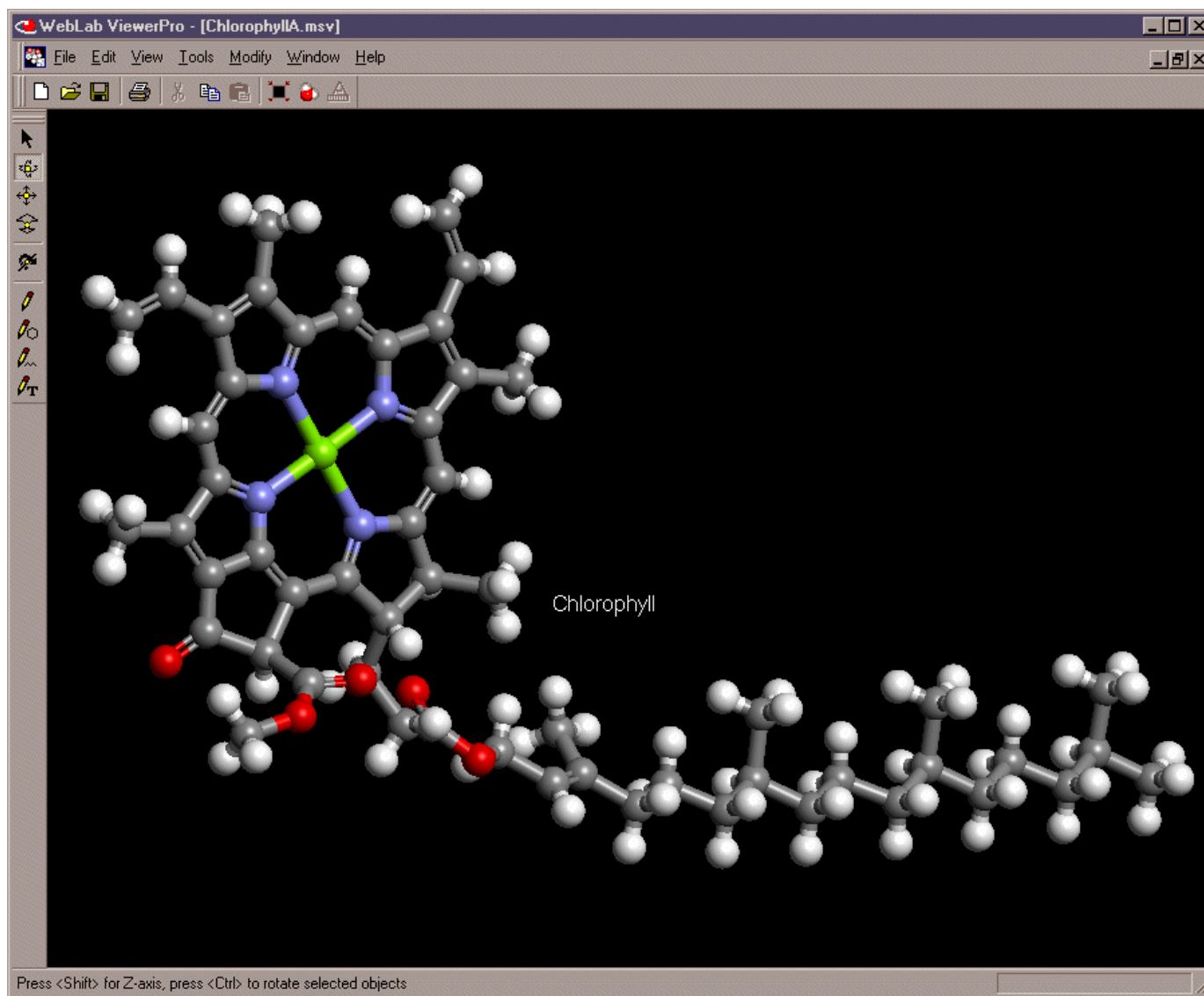


Richard Willstätter
Nobelpreis 1915





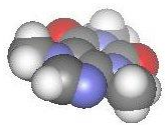
Discovery Studio Viewer



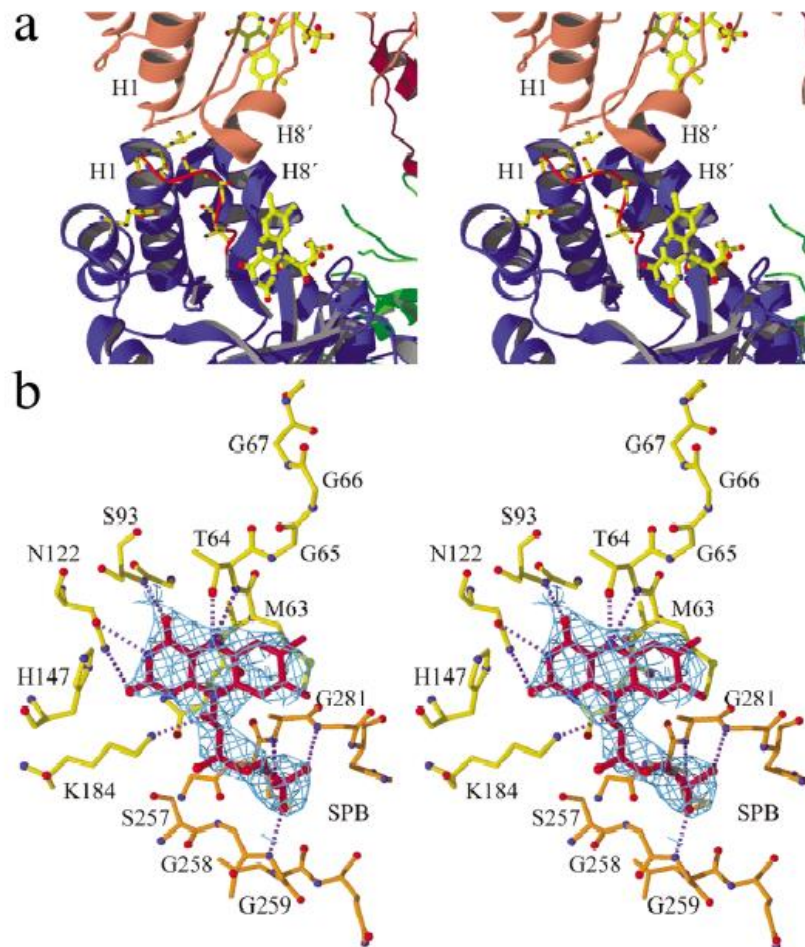
ViewerLite bei:
www.ch.tum.de/oc1/EFontain

ViewerPro bei Accelrys
(auch als Trial-Version)





Stereodarstellung in Publikationen

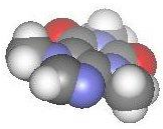


"Crystal Structure of the Type II Isopentenyl Diphosphate: Dimethylallyl Diphosphate Isomerase from *Bacillus subtilis*"

Stefan Steinbacher, Johannes Kaiser, Stefan Gerhardt, Wolfgang Eisenreich, Robert Huber, Adelbert Bacher and Felix Rohdich

J. Mol. Biol. (2003) 329, 973–982

Figure 3. FMN binding site. (a) Close-up of the octameric contacts and the glycine-rich sequence motif T64GGG. (b) $F_o - F_c$ simulated annealing omit electron density of FMN at 2.8 Å resolution and contoured at 2.5 sigma. The phosphate ion is hydrogen bonded to the amide of Gly259 of the glycine-rich motif Ser257GlyGly and to the amides of Ala280 and Gly281 that form the standard phosphate binding (SPB) motif. The isoalloxazine ring is fixed by hydrogen bonds to Thr64, S93 and N122. The *si*-side of FMN is exposed for reactions. The conserved His147 is placed near FMN.



Betrachten von Stereobildern

■ Normales Sehen

- Blick auf die Papierebene
- Focus nah
- Parallaxe nah
- Wir sehen:
 - 2 einzelne Bilder

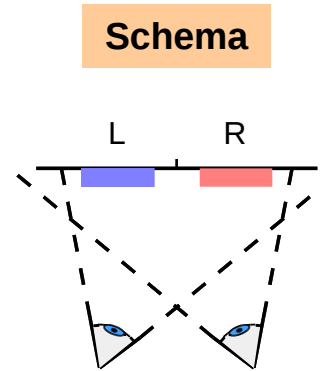
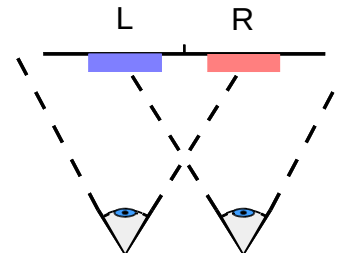


Bild im Gehirn



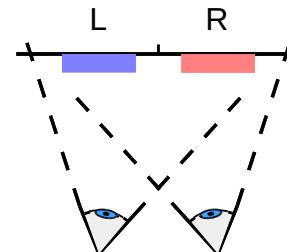
■ Mehrfachbild (relaxed eyes)

- Blick weit hinter die Papierebene
- Focus weit
- Parallaxe weit
- Wir sehen:
 - 4 unscharfe Bilder

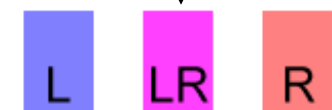


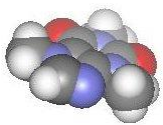
■ Stereobild (relaxed eyes)

- Blick auf die Papierebene
- Focus nah
- Parallaxe zwischen nah und weit
- Wir sehen:
 - 3 scharfe Bilder mit Stereobild in der Mitte

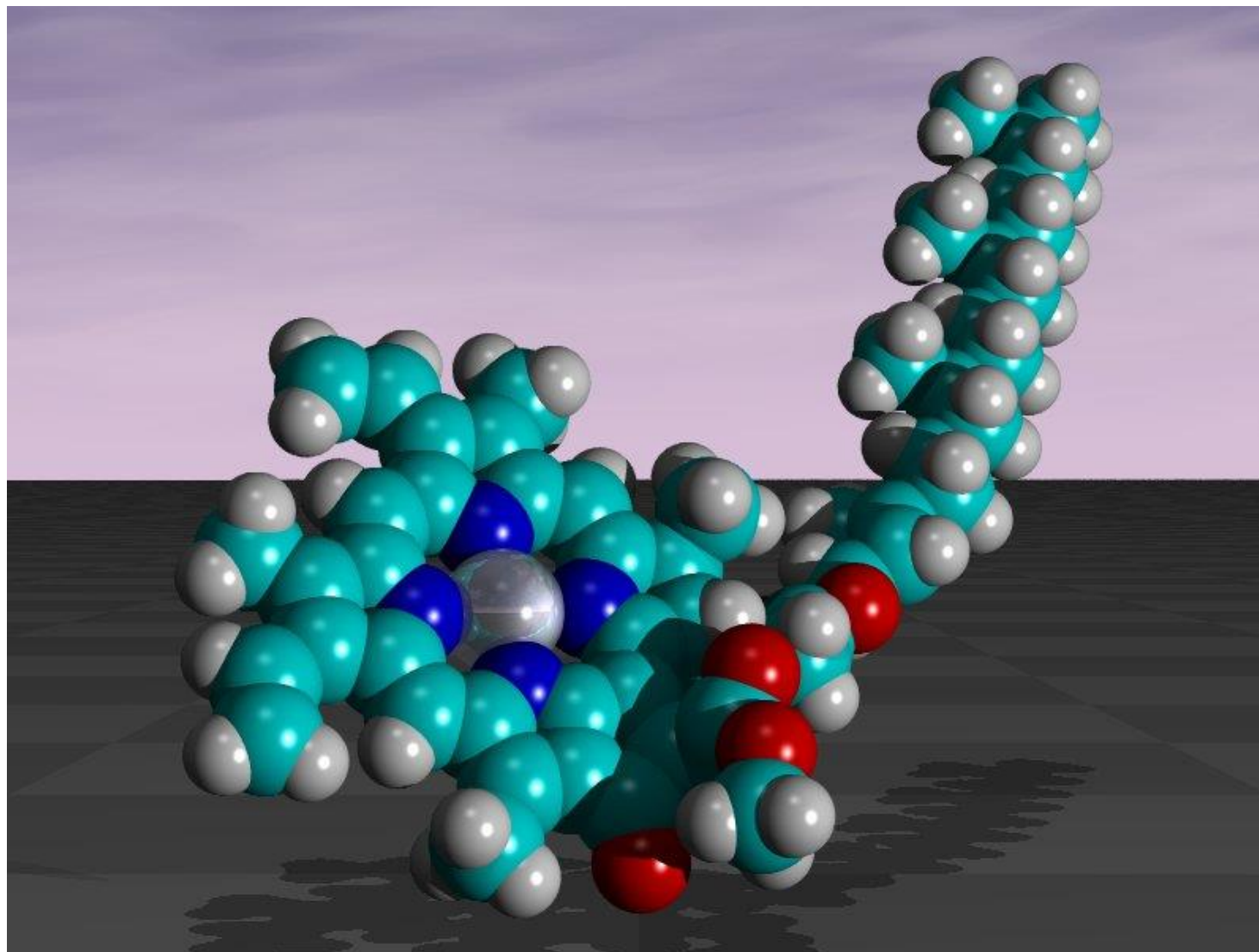


Stereobild



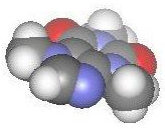


Persistence of Vision Raytracing POV-Ray

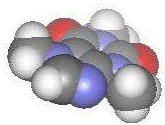


 POVChem

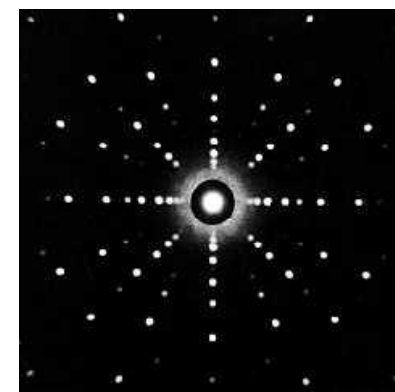
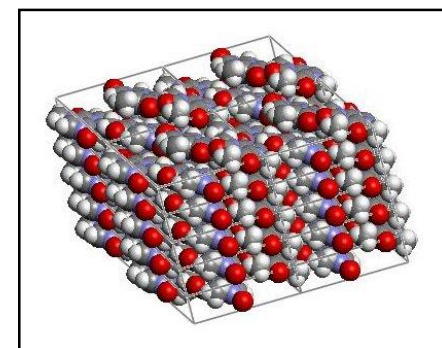
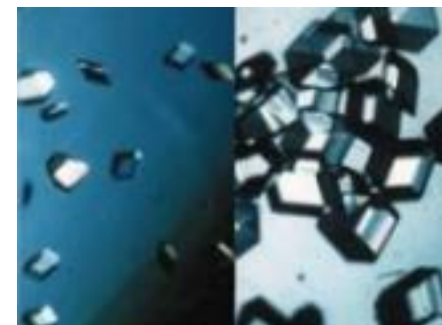
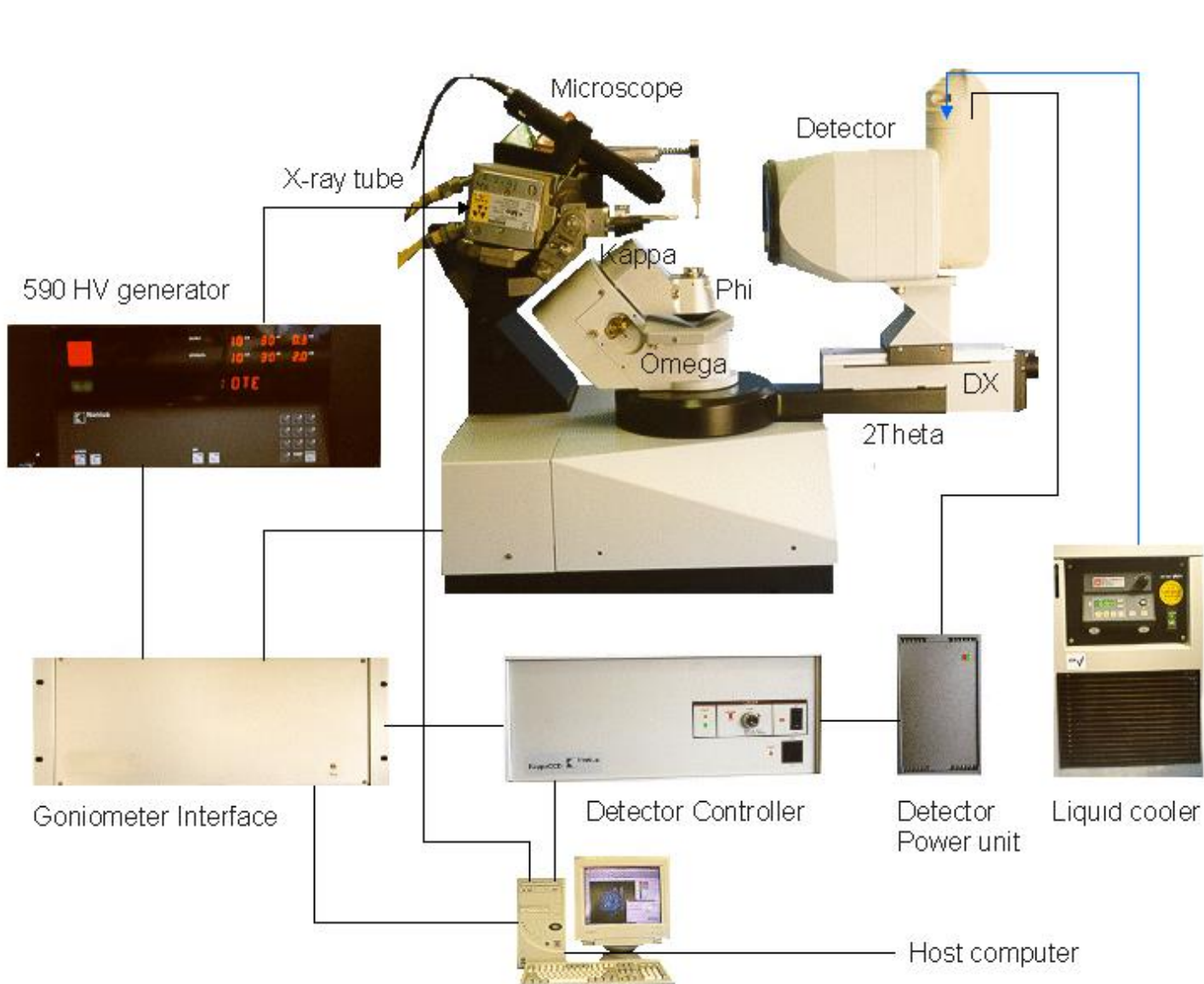
 POV-Ray

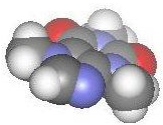


Molekülstrukturen im Internet



Röntgenstrukturanalyse (X-Ray)





- Für kleine und mittelgroße Organische Moleküle (keine Biopolymere)
- Am Cambridge Crystallographic Data Centre (CCDC) gegründet 1965
- Strukturen durch X-Ray und n-Beugung
- 2018 ca. 930.000 Strukturen (+ ca. 60.000 / a)
- Datenformat z.B. MOL2 von Tripos
- Adresse:
 - www.ccdc.cam.ac.uk



CSD



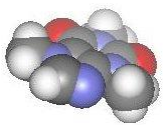
Coffein



MOL2



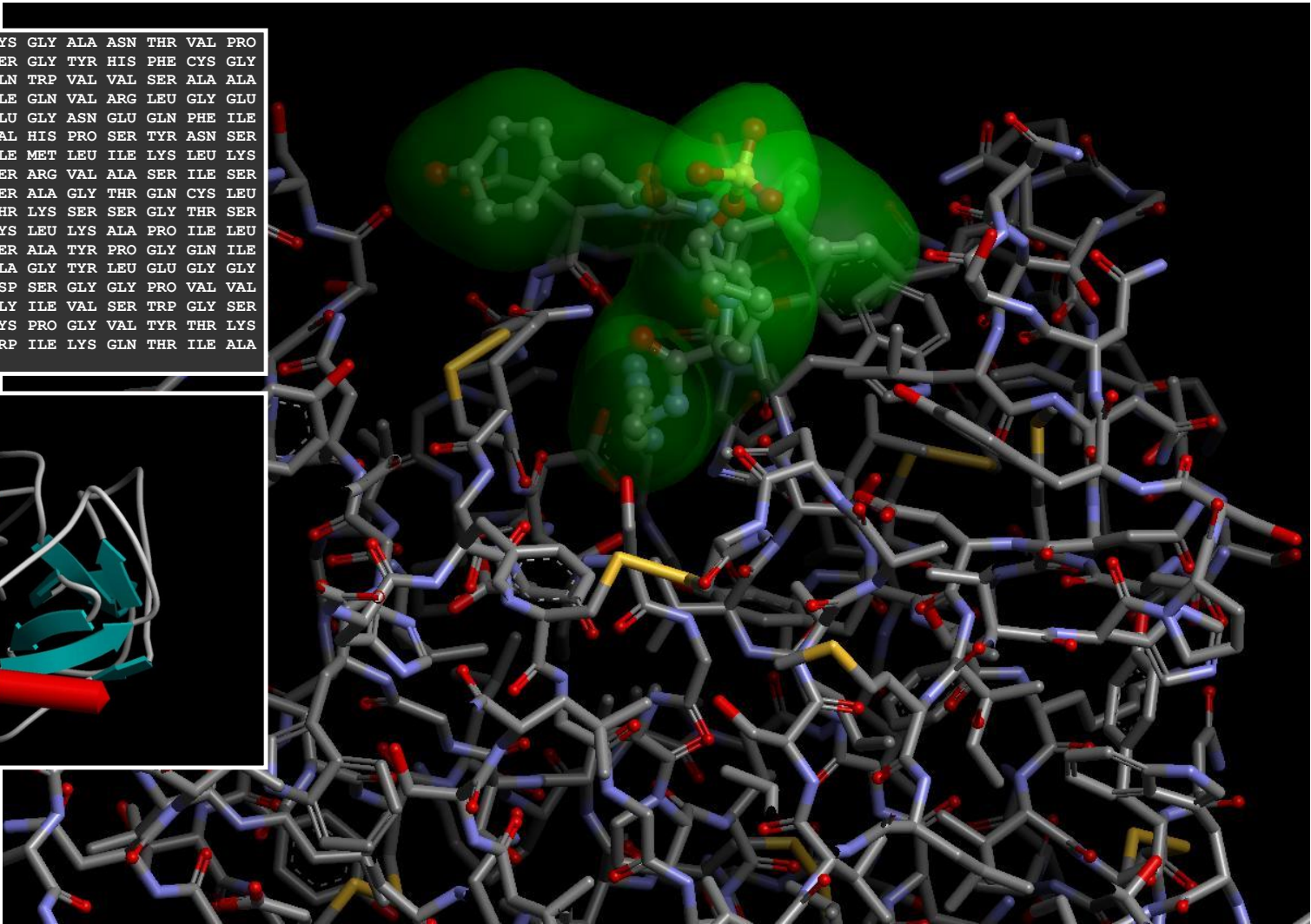
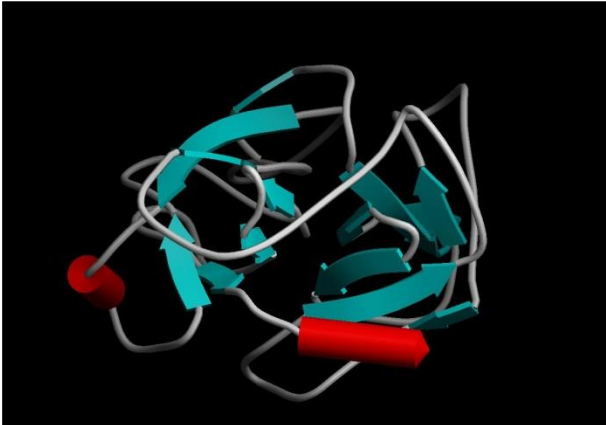
Tripos



Die Proteinstruktur

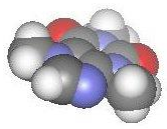
```

ILE VAL GLY GLY TYR THR CYS GLY ALA ASN THR VAL PRO
TYR GLN VAL SER LEU ASN SER GLY TYR HIS PHE CYS GLY
GLY SER LEU ILE ASN SER GLN TRP VAL VAL SER ALA ALA
HIS CYS TYR LYS SER GLY ILE GLN VAL ARG LEU GLY GLU
ASP ASN ILE ASN VAL VAL GLU GLY ASN GLU GLN PHE ILE
SER ALA SER LYS SER ILE VAL HIS PRO SER TYR ASN SER
ASN THR LEU ASN ASN ASP ILE MET LEU ILE LYS LEU LYS
SER ALA ALA SER LEU ASN SER ARG VAL ALA SER ILE SER
LEU PRO THR SER CYS ALA SER ALA GLY THR GLN CYS LEU
ILE SER GLY TRP GLY ASN THR LYS SER SER GLY THR SER
TYR PRO ASP VAL LEU LYS CYS LEU LYS ALA PRO ILE LEU
SER ASP SER SER CYS LYS SER ALA TYR PRO GLY GLN ILE
THR SER ASN MET PHE CYS ALA GLY TYR LEU GLU GLY GLY
LYS ASP SER CYS GLN GLY ASP SER GLY GLY PRO VAL VAL
CYS SER GLY LYS LEU GLN GLY ILE VAL SER TRP GLY SER
GLY CYS ALA GLN LYS ASN LYS PRO GLY VAL TYR THR LYS
VAL CYS ASN TYR VAL SER TRP ILE LYS GLN THR ILE ALA
SER ASN
  
```

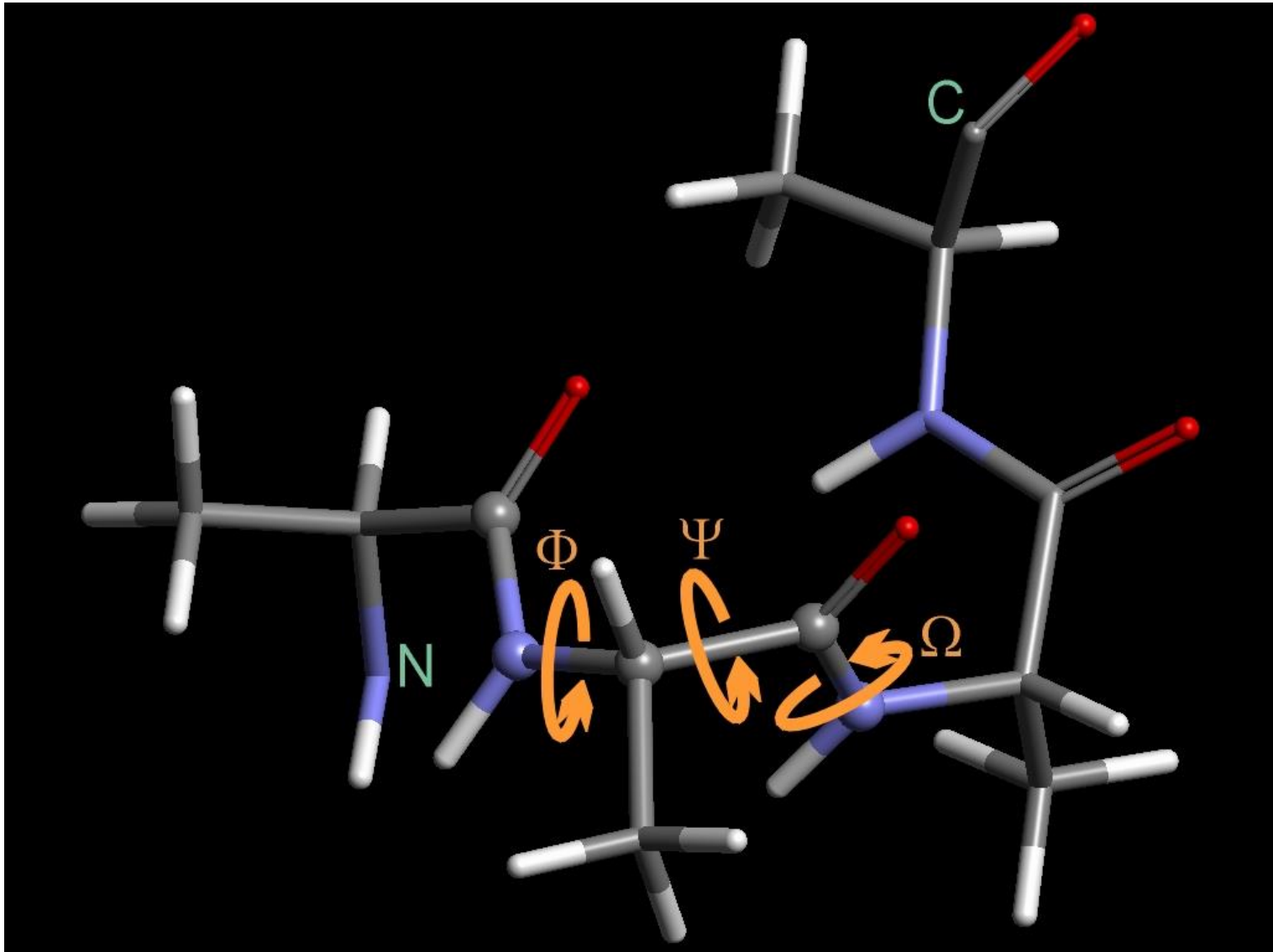


Trypsin

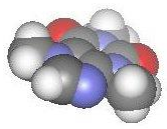




Das Protein Rückgrat (Backbone)

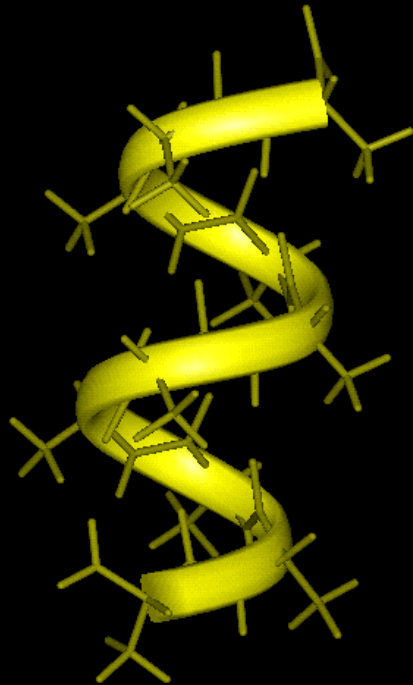


PhiPsiOmega



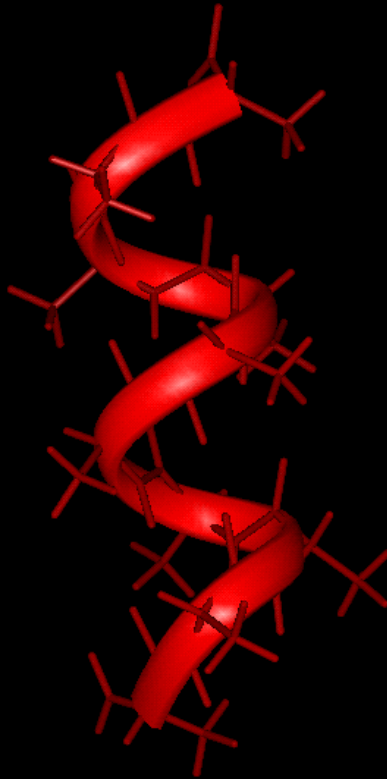
Rechtsgängige Helices in Proteinen

pi helix



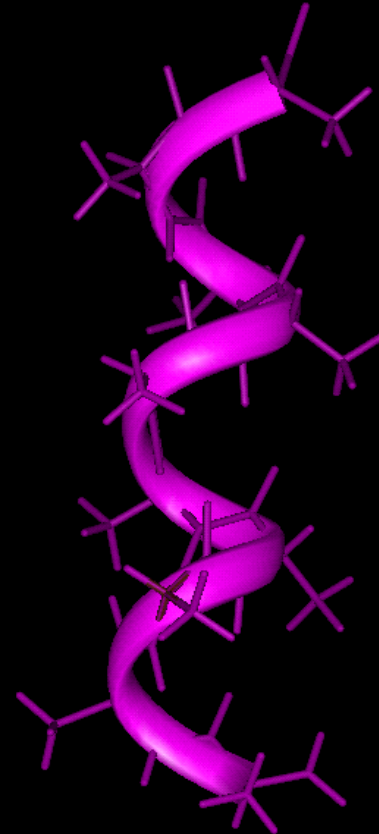
PHI: -30 deg | PSI: -90 deg | OMEGA: 180 deg

right-handed alpha-helix

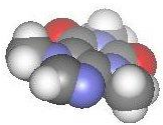


PHI: -65 deg | PSI: -40 deg | OMEGA: 180 deg

3₁₀ helix



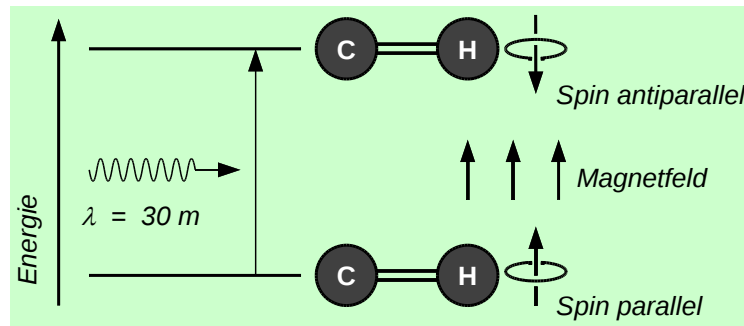
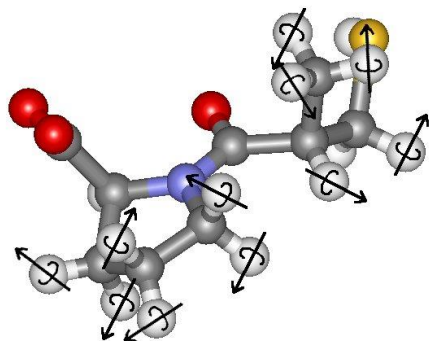
PHI: -60 deg | PSI: -30 deg | OMEGA: 180 deg



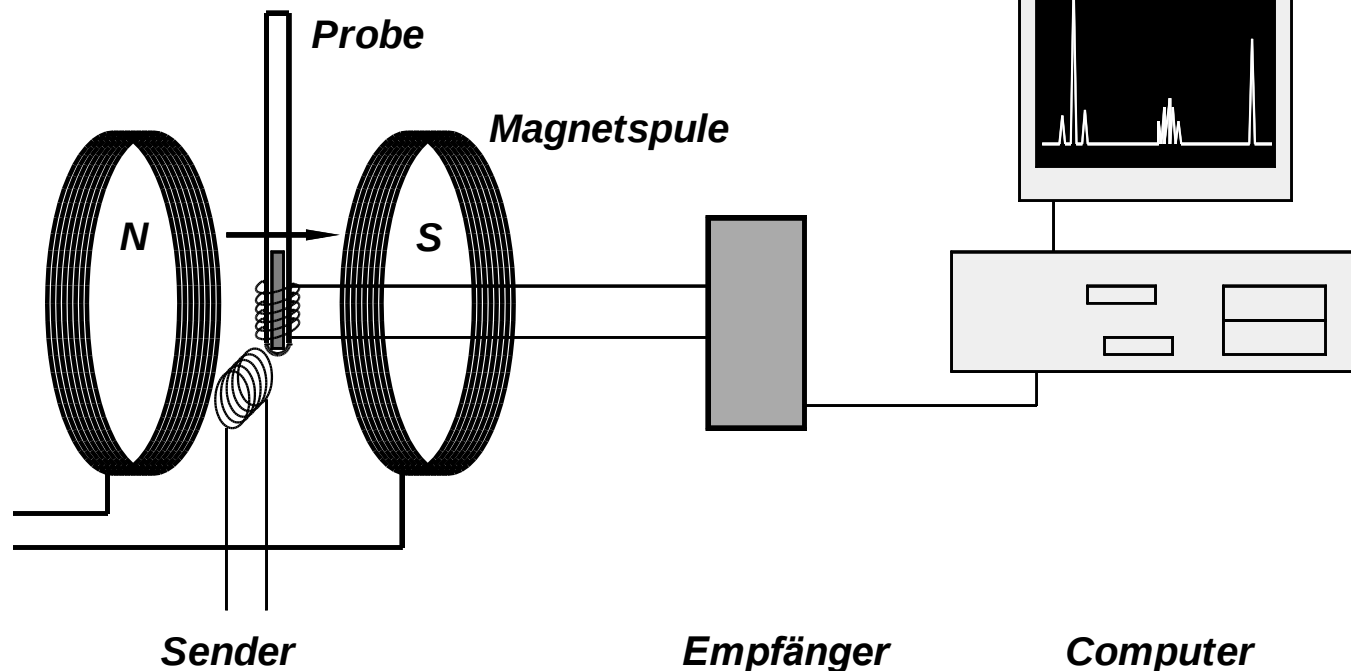
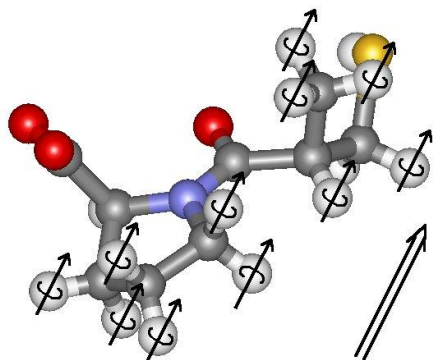
Kernresonanzspektroskopie (NMR)

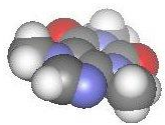
NMR-fähige Atomkerne: z.B. ^1H , ^{13}C (nicht ^{12}C !), ^{19}F , ^{31}P

ohne Magnetfeld:



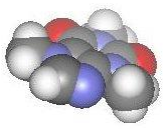
mit Magnetfeld:





Nuclear Magnetic Resonance (NMR)





- Biopolymerstrukturen gewonnen durch X-Ray, NMR und Elektronenmikroskopie
- Gegründet 1971 an den Brookhaven National Laboratories (BNL)
- Jetzt beim Research Collaboratory for Structural Bioinformatics (RCSB)
- 2019 ca. 160.000 Strukturen (+ ca. 15.000 / a)
Proteine und Peptide, Nucleinsäuren
- Datenformat PDB und mmCIF
- Webadresse:
 - www.rcsb.org



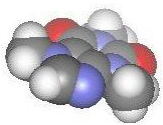
RCSB



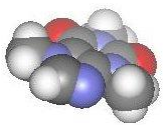
PDB



Trypsin



Einfaches Molecular Modeling



Was ist Molecular Modeling ?

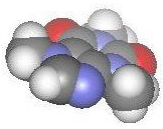
- Quantitatives, computerunterstütztes Hantieren mit mathematischen Computermodellen von Molekülen und Ensembles davon
- Messen von Strukturparametern
- Vorhersage von Molekülstrukturen
- Energieabschätzungen, z.B. Protein/Ligandwechselwirkungen in Docking-Experimenten
- Molekulardynamik
- u.v.m.



Collapse
Peptide



SN2
Substitution



■ Klassische Verfahren

■ Molecular Mechanics

- Mechanisches Modell mit parametrisierten "Federkräften"
- Beispiele : MM2/3, AMBER, OPLS, CHARMM
- keine Betrachtung der Bindungselektronen, Bindungen bleiben erhalten, neue Bindungen werden nicht gebildet, keine chemischen Reaktionen beobachtbar
- für kleine/mittlere/große Moleküle und Biopolymere

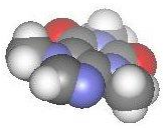
■ Quantenchemische Verfahren

■ Semi-empirical

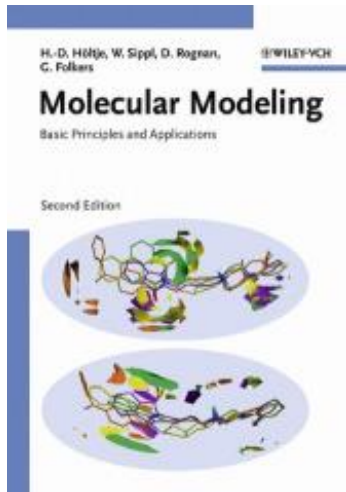
- Quantenchemie mit Orbitalen und Parametern
- Beispiele: MOPAC, MINDO/3, MNDO, AM1, PM3
- Alle Elektronen werden betrachtet. Bindungen ergeben sich durch "Elektronendichte", chemische Reaktionen sind beobachtbar. Ladungsverschiebungen werden sichtbar.
- für kleine und mittlere Molekülgrößen

■ Ab initio

- Quantenchemie ausschließlich auf Grundlage der Schrödingergleichung und von Naturkonstanten
- Beispiele : Gaussian 92/98, DFT (Density Function Theory)
- nur für kleine Moleküle



Literatur zum Molecular Modeling



Für Einsteiger:

Molecular Modeling

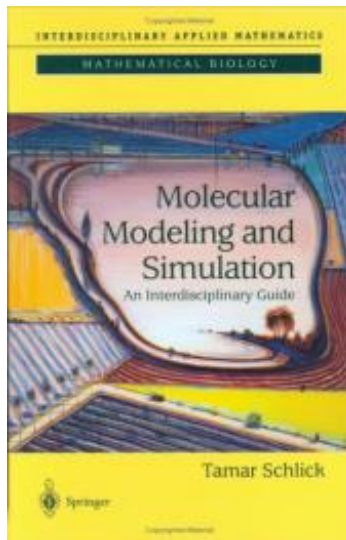
H.-D. Höltje, W. Sippl, D. Rognan, G. Folkers

Wiley-VCH, Weinheim, 2003

228 Seiten, englisch

ISBN 3527305890

€ 52,90



Für Fortgeschrittene:

Molecular Modeling and Simulation

T. Schlick

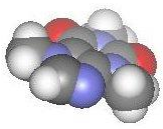
Springer-Verlag, New York, 2002

635 Seiten, englisch

ISBN 038795404X

€ 83,90

Enthält viel über Proteinstruktur
und einiges an Mathematik

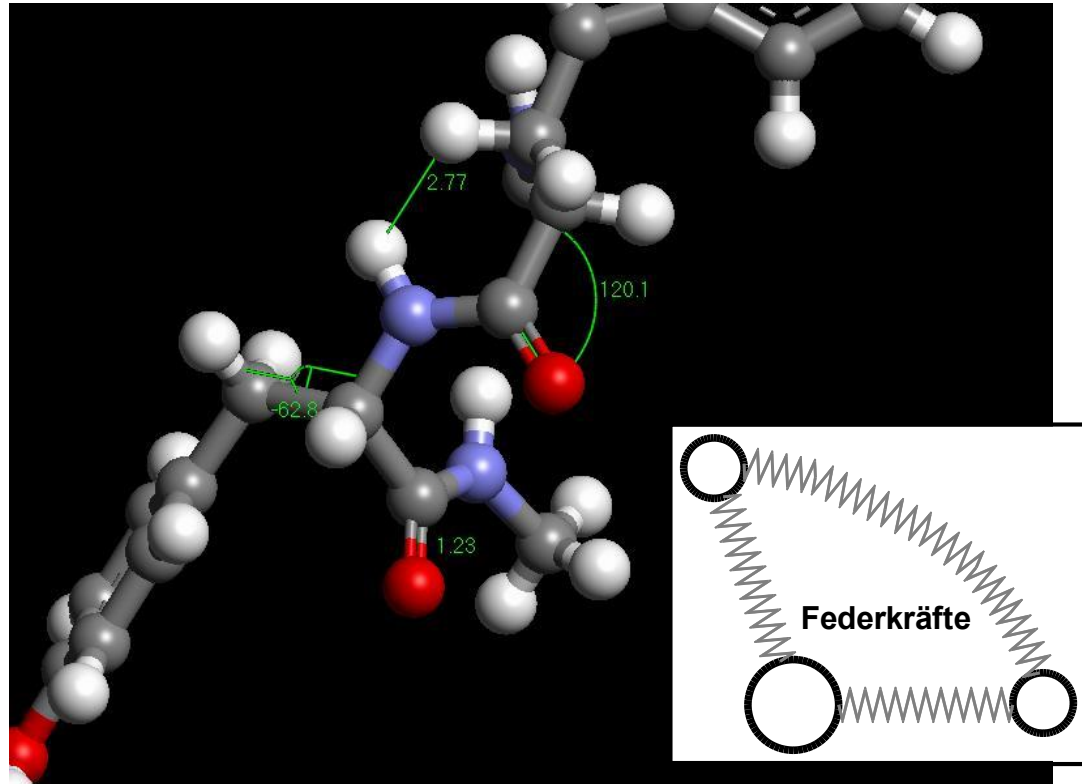


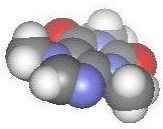
■ **Potentielle Energie $E_{\text{pot}} = f(\text{Molekülstruktur})$**

■ $E_{\text{pot}} = E_{\text{bnd}} + E_{\text{ang}} + E_{\text{tor}} + E_{\text{VDW}} + E_{\text{Hbd}} + E_{\text{ele}} + E_{\text{oop}}$

■ bnd	Bindungslängen
■ ang	Bindungswinkel
■ tor	Torsionswinkel
■ VDW	Van-der-Waals
■ Hbd	H-Brücken
■ ele	Elektrostatik
■ oop	Out-of-plane (Korrektur)

■ **Jede Veränderung auch nur einer Atomkoordinate verändert die potentielle Energie des Moleküls**





Bindungslängen E_{bnd}

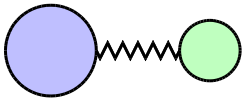
$$E_{\text{pot}} = E_{\text{bnd}} + E_{\text{ang}} + E_{\text{tor}} + E_{\text{VDW}} + E_{\text{Hbd}} + E_{\text{ele}} + E_{\text{oop}}$$



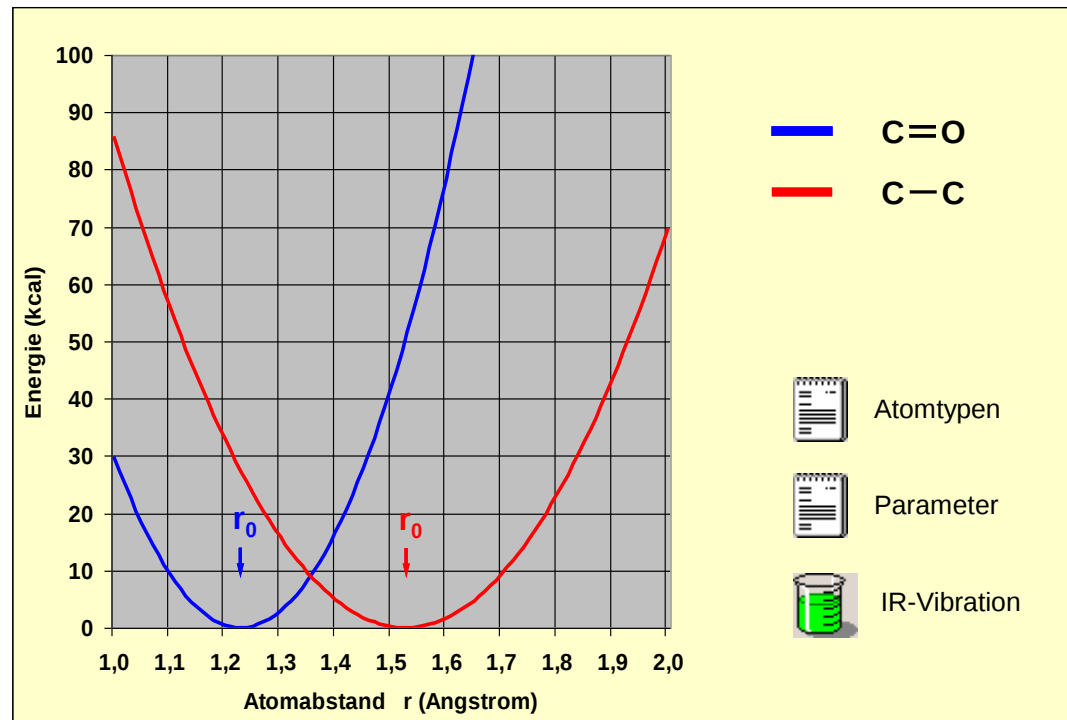
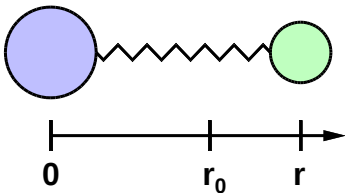
$$E_{\text{bnd}} = \sum_{\text{Bindungen}} K_r \cdot (r - r_0)^2$$

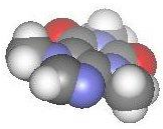
K_r Kraftkonstante der Bindung
 r_0 idealer Bindungsabstand (spannungsfrei)
 r aktueller Bindungsabstand

spannungsfrei:



gestreckt / gestaucht:





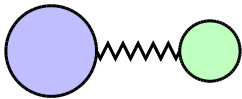
Morsepotential E_{mor}

$$E_{\text{pot}} = \boxed{E_{\text{bnd}}} + E_{\text{ang}} + E_{\text{tor}} + E_{\text{VDW}} + E_{\text{Hbd}} + E_{\text{ele}} + E_{\text{oop}}$$

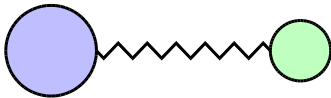
$$E_{\text{mor}} = \sum_{\text{Bindungen}} D_e \cdot \left(1 - e^{\sqrt{\frac{K_r}{2D_e}}(r-r_0)}\right)^2$$

K_r Kraftkonstante der Bindung
 D_e Dissoziationsenergie (Spaltung)
 r_0 idealer Bindungsabstand (spannungsfrei)
 r aktueller Bindungsabstand

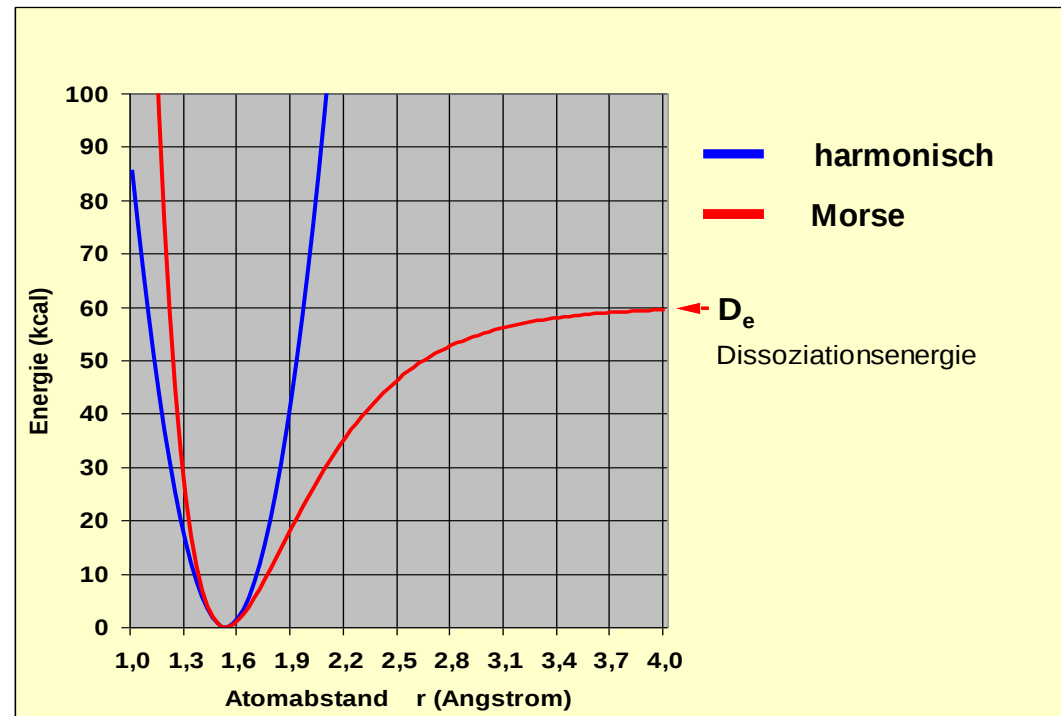
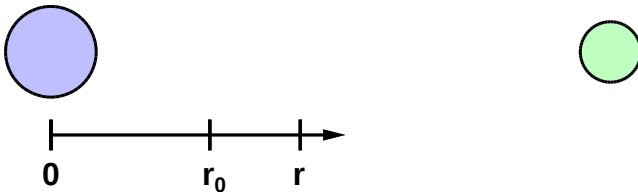
spannungsfrei:

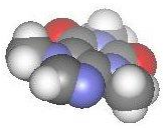


gestreckt / gestaucht:



getrennt:





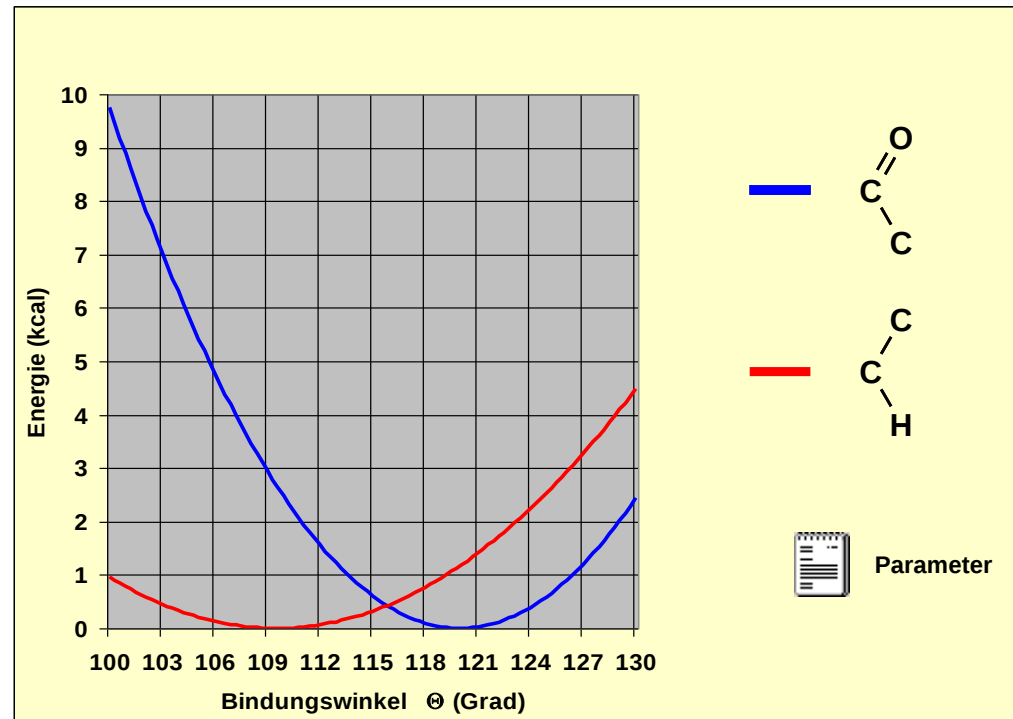
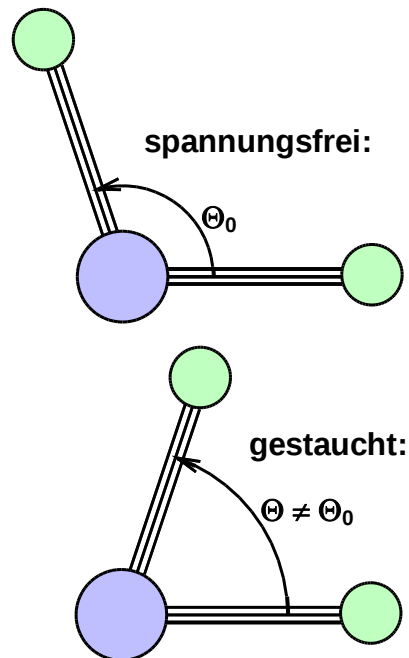
Bindungswinkel E_{ang}

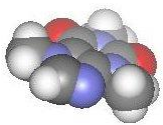
$$E_{pot} = E_{bnd} + E_{ang} + E_{tor} + E_{VDW} + E_{Hbd} + E_{ele} + E_{oop}$$



$$E_{ang} = \sum_{\text{Bindungswinkel}} K_{\Theta} \cdot (\Theta - \Theta_0)^2$$

K_{Θ} Kraftkonstante des Bindungswinkels
 Θ_0 idealer Bindungswinkel (spannungsfrei)
 Θ aktueller Bindungswinkel



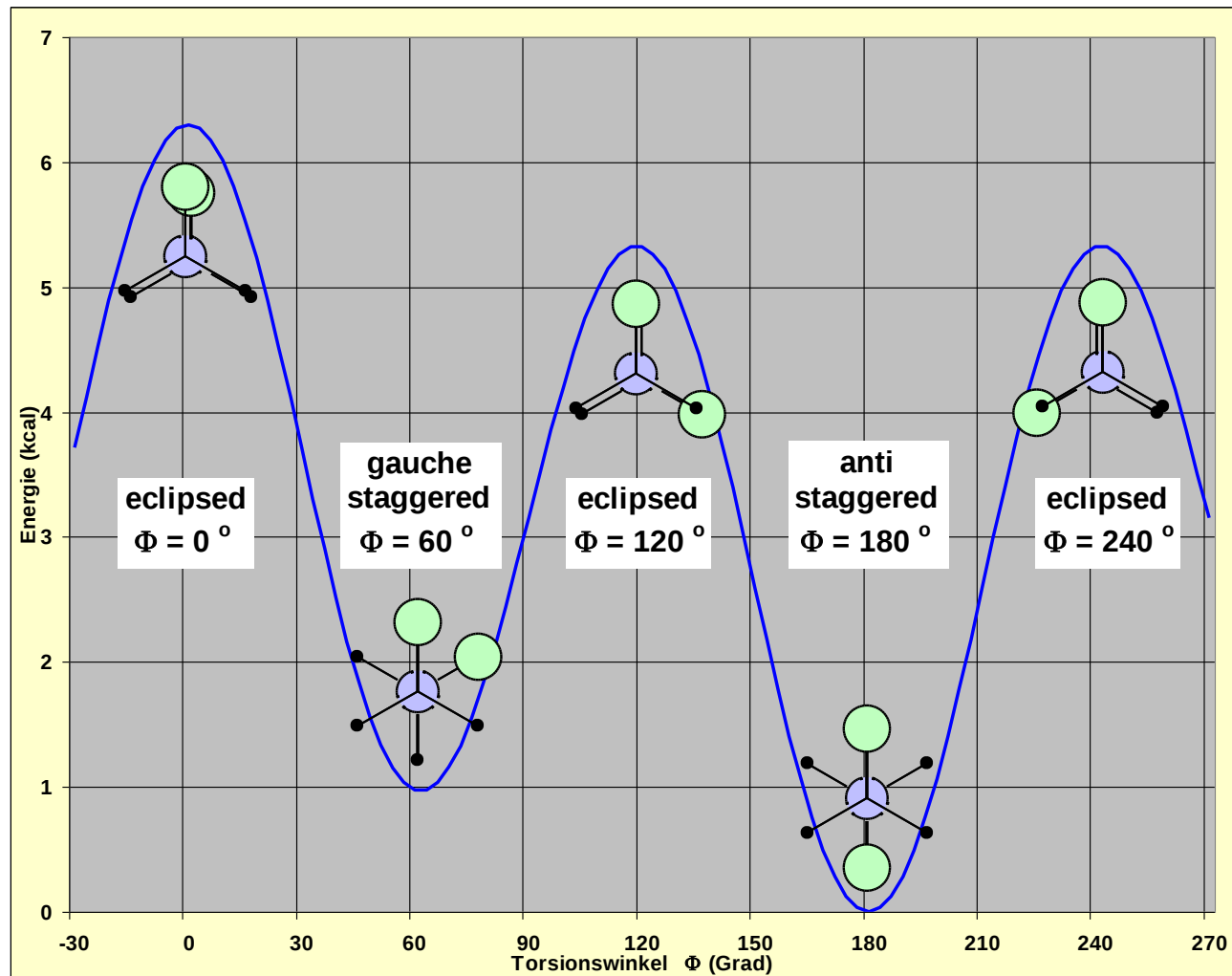
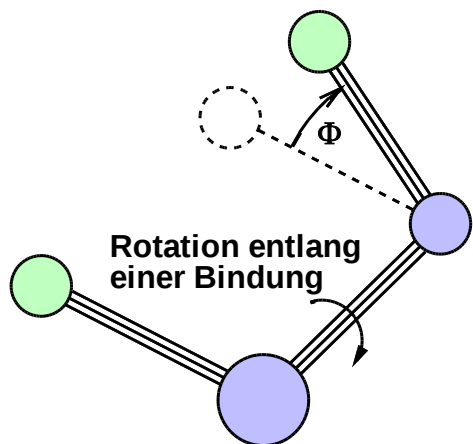


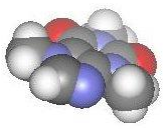
Torsionswinkel E_{tor}

$$E_{\text{pot}} = E_{\text{bnd}} + E_{\text{ang}} + E_{\text{tor}} + E_{\text{VDW}} + E_{\text{Hbd}} + E_{\text{ele}} + E_{\text{oop}}$$



Torsionswinkel Φ (Definition):





Torsionswinkel E_{tor}

$$E_{\text{pot}} = E_{\text{bnd}} + E_{\text{ang}} + E_{\text{tor}} + E_{\text{VDW}} + E_{\text{Hbd}} + E_{\text{ele}} + E_{\text{oop}}$$

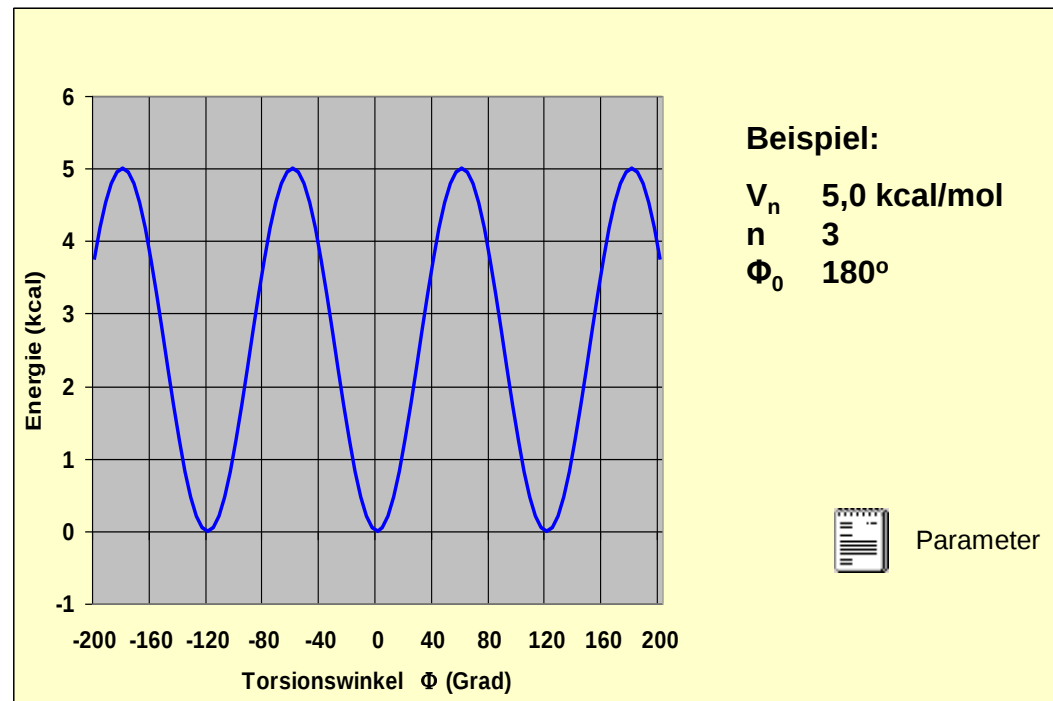
$$E_{\text{tor}} = \sum_{\text{Torsionswinkel}} V_n / 2 \cdot [1 + \cos(n\Phi - \Phi_0)] + \dots$$

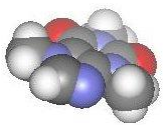
V_n Höhe der Rotationsbarriere
 n Periodizität (Anz.Barrieren)
 Φ_0 Phasenwinkel
 Φ aktueller Torsionswinkel
 $[1+\cos\dots]$ Fourierterm

Phasenwinkel Φ_0 :

0° Bei 0° liegt ein Maximum

180° Bei 0° liegt ein Minimum



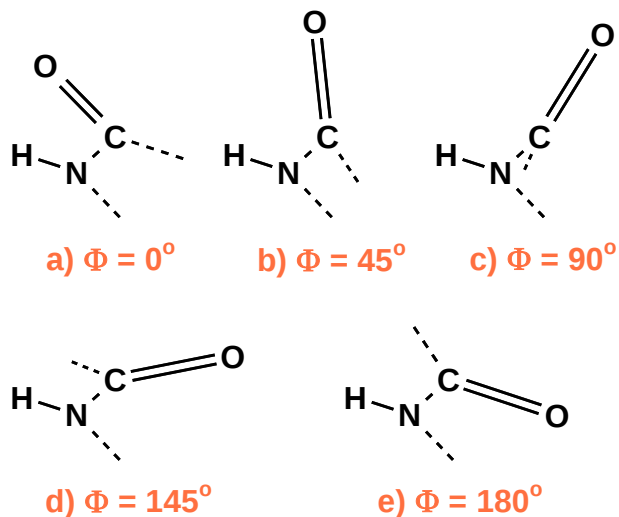


Beispiel: Torsionsenergie der Amidbindung

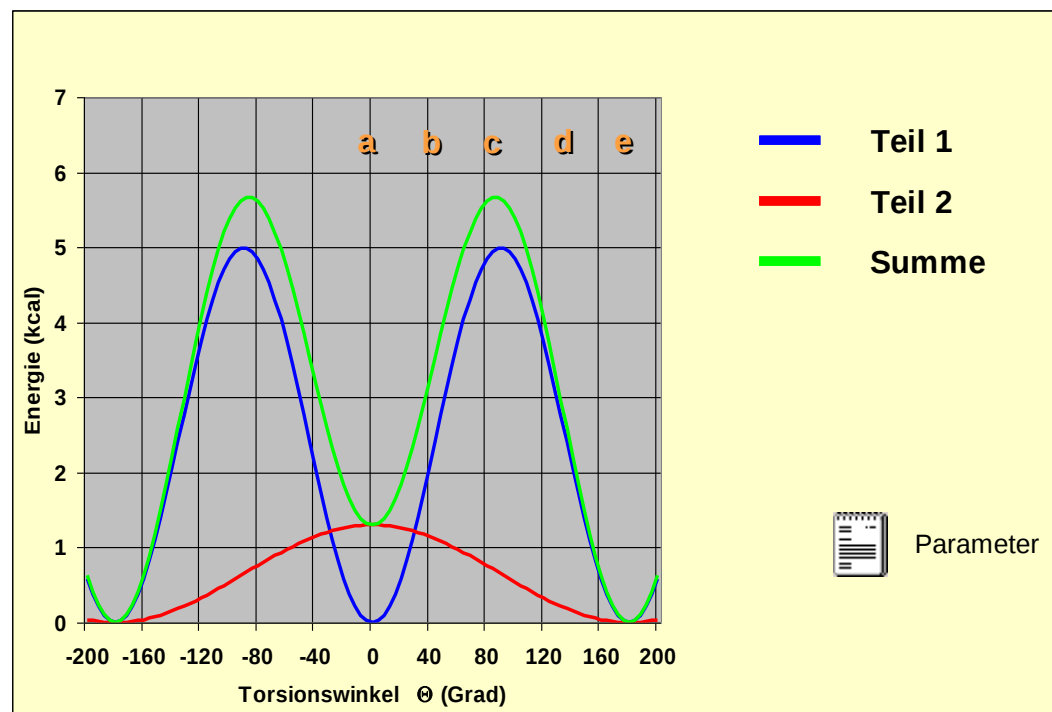
$$E_{\text{pot}} = E_{\text{bnd}} + E_{\text{ang}} + E_{\text{tor}} + E_{\text{VDW}} + E_{\text{Hbd}} + E_{\text{ele}} + E_{\text{oop}}$$

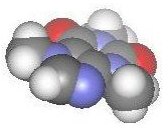


$$E_{\text{Amid}} = 5,0/2 \text{ kcal/mol} \cdot [1 + \cos(2\Phi - 180^\circ)] \text{ (Teil 1) Fourierterm 1} \\ + 1,3/2 \text{ kcal/mol} \cdot [1 + \cos(1\Phi - 0^\circ)] \text{ (Teil 2) Fourierterm 2}$$



Energie (a, cis) > Energie (e, trans)





Van-der-Waals-Wechselwirkungen E_{VDW}

$$E_{pot} = E_{bnd} + E_{ang} + E_{tor} + E_{VDW} + E_{Hbd} + E_{ele} + E_{oop}$$



$$E_{VDW} = \sum_{\text{Atompaae } i < j} \left[\frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^6} \right]$$

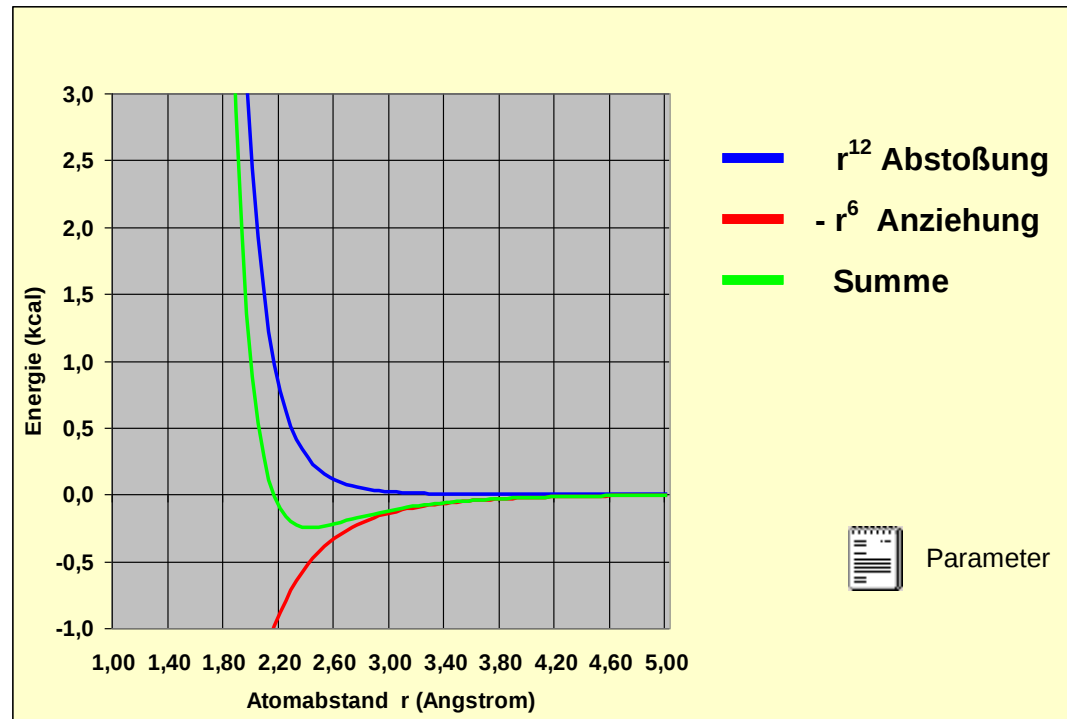
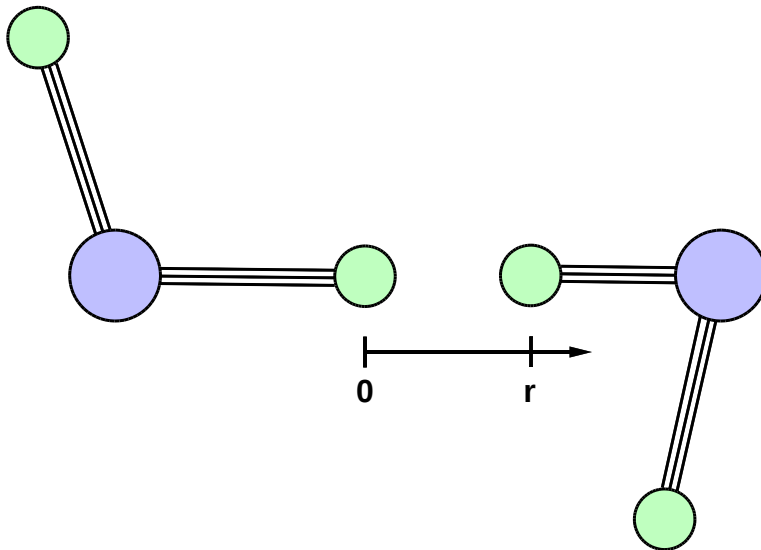
Lennard-Jones-Potential:

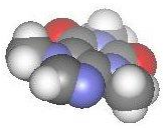
A_{ij} **Abstoßungsparameter** für Atompaar i und j

B_{ij} **Anziehungsparameter** für Atompaar i und j

r aktueller Atomabstand

Nicht-bindende Wechselwirkungen





Wasserstoffbrückenbindungen E_{Hbd}

$$E_{\text{pot}} = E_{\text{bnd}} + E_{\text{ang}} + E_{\text{tor}} + E_{\text{VDW}} + E_{\text{Hbd}} + E_{\text{ele}} + E_{\text{oop}}$$



$$E_{\text{Hbd}} = \sum_{\text{H-Brücken } i < j} \left[\frac{C_{ij}}{r^{12}} - \frac{D_{ij}}{r^{10}} \right]$$

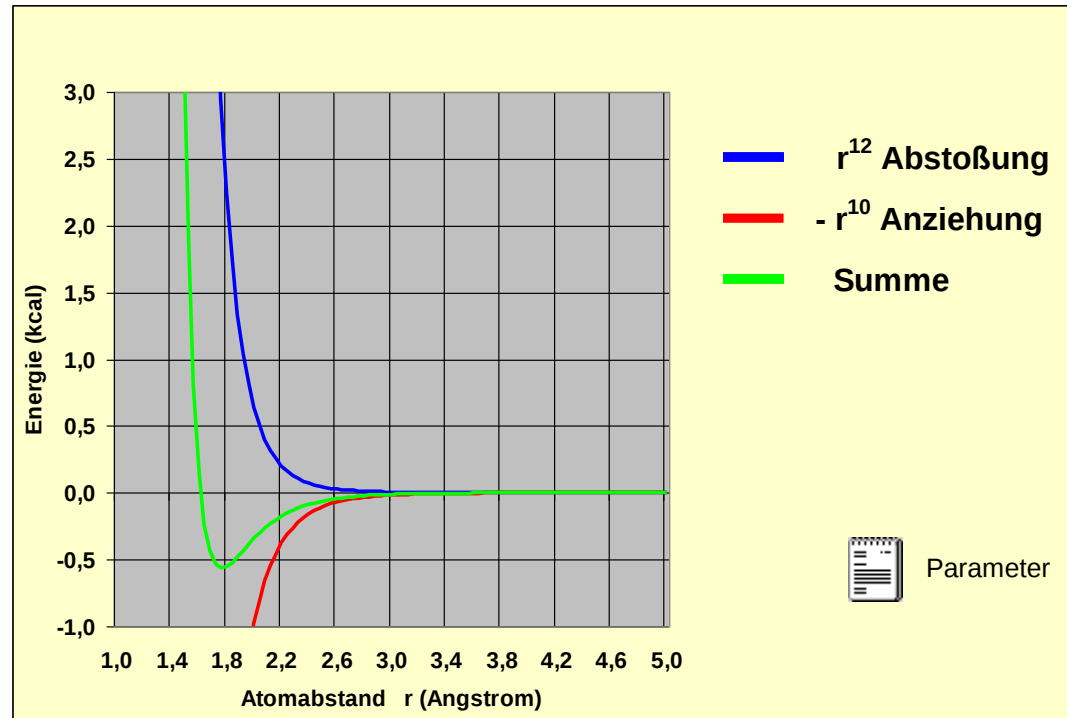
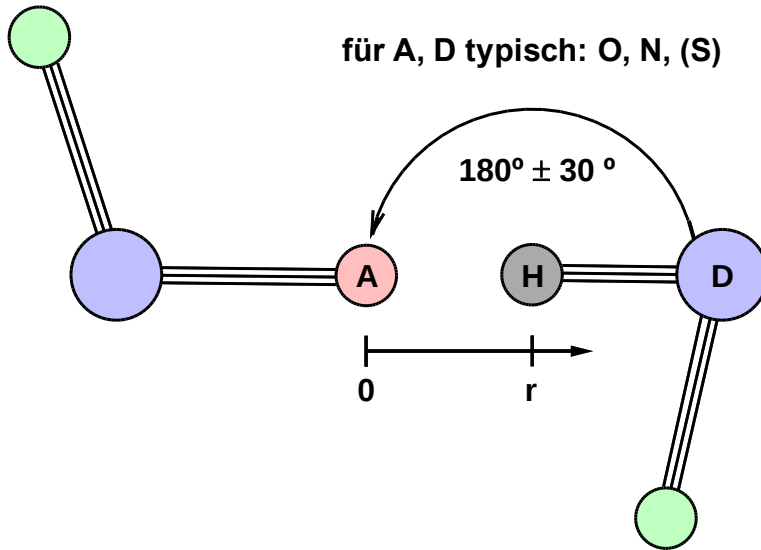
H-Brücken-Potential:

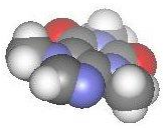
C_{ij} **Abstoßungsparameter** für Atompaar i und j

D_{ij} **Anziehungsparameter** für Atompaar i und j

r aktueller Atomabstand

H-Brücken zwischen H-Donor und H-Acceptor





Coulomb-Wechselwirkungen E_{ele}

$$E_{\text{pot}} = E_{\text{bnd}} + E_{\text{ang}} + E_{\text{tor}} + E_{\text{VDW}} + E_{\text{Hbd}} + \boxed{E_{\text{ele}}} + E_{\text{oop}}$$

$$E_{\text{ele}} = \sum_{\text{Atompaare } i < j} \frac{q_i \cdot q_j}{\epsilon \cdot r}$$

Coulomb-Potential:

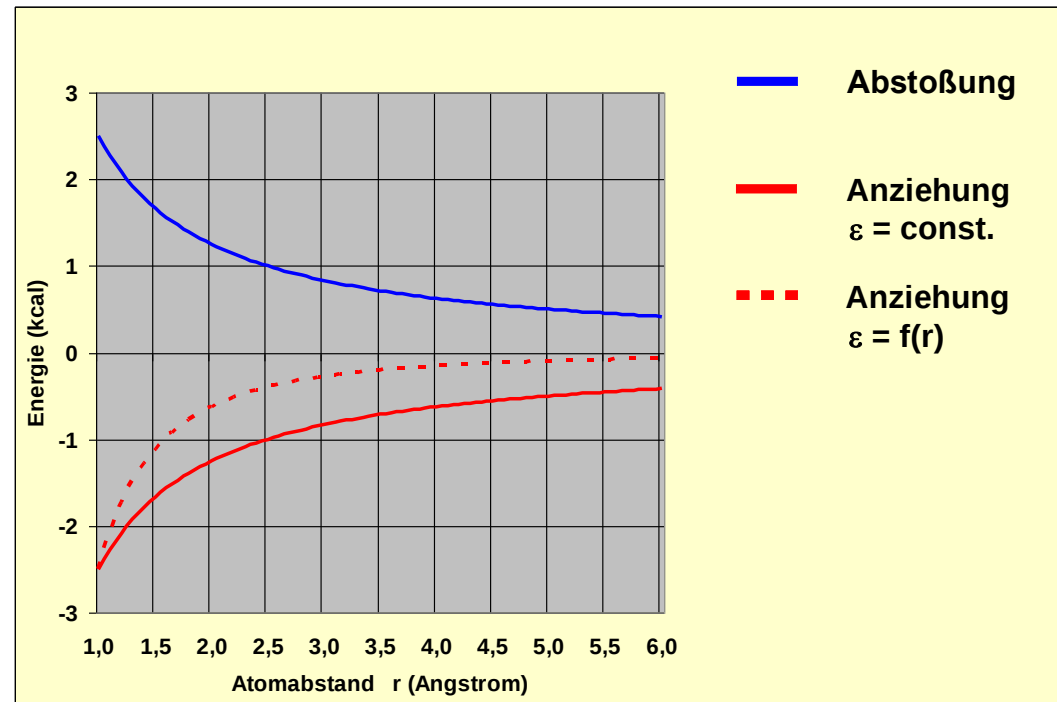
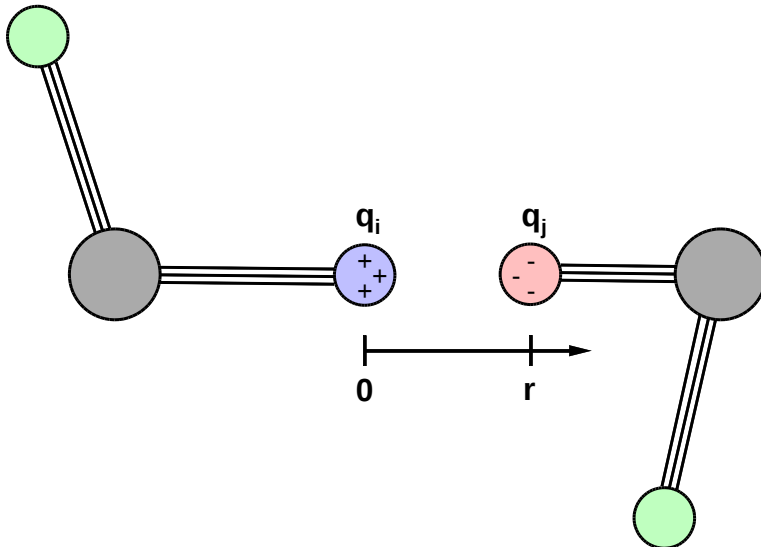
q Partialladungen der Atome i und j

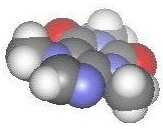
ϵ Dielektrizitätskonstante des Mediums

$e_{\text{Vak}} = 8,854 \cdot 10^{-12} \text{ C/Vm}$, $e_{\text{H}_2\text{O}} = e_{\text{Vak}} \cdot 80 \text{ (ca.)}$

r aktueller Atomabstand

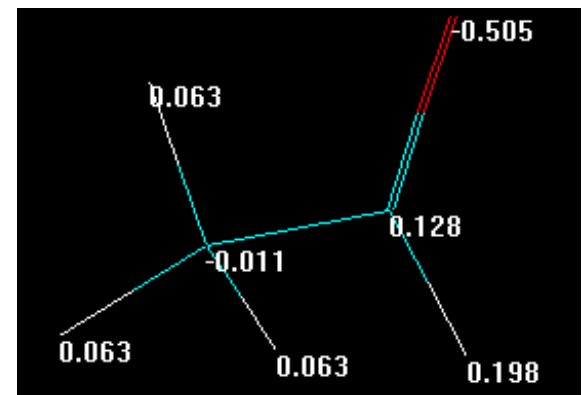
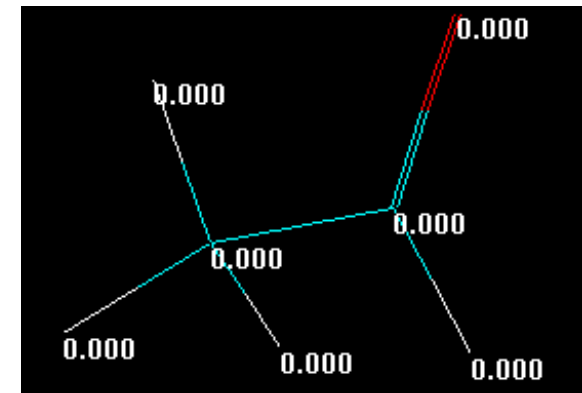
Elektrostatische Wechselwirkungen

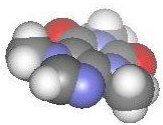




Wo kommen die Ladungen q_i und q_j her?

- Im klassischen Modell der Molekülmechanik tauchen keine Elektronen auf
(weder in den Bindungen, noch als freie Elektronen)
- Die formalen und partiellen Ladungen müssen separat hinzugefügt werden
- Tabellen
 - Jedes Atom in Tabellen suchen und Ladung der ähnlichsten Atomtypen übernehmen
 - Problem bei nicht gefundenen Atomgruppierungen
 - Beispiel AMBER (für Aminosäuren und einfache Zucker)
- Empirische Berechnungen
 - Entwickelt an der TUM (Gasteiger/Marsili)
 - Partieller Ausgleich der Unterschiede in den Orbital-Elektro negativitäten (PEOE)
 - Unabhängig von der räumlichen Atomanordnung ($\Rightarrow$ schnell und allgemein anwendbar)
 - In manchen Fällen nur mäßig genau
- Quantenmechanische Berechnungen
 - Zeitlich sehr aufwändig bis unmöglich
 - Ergebnis ist abhängig von der räumlichen Atomanordnung





Out-of-Plane-Korrektur E_{oop}

$$E_{\text{pot}} = E_{\text{bnd}} + E_{\text{ang}} + E_{\text{tor}} + E_{\text{VDW}} + E_{\text{Hbd}} + E_{\text{ele}} + E_{\text{oop}}$$

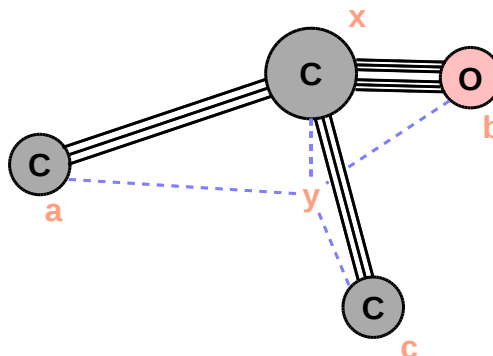
Korrektur nicht-planarer sp^2 -Zentren durch:

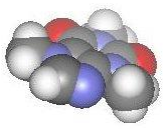
■ zusätzliche E_{ang} mit Projektion:

- z.B. Winkel ayc statt axc , zusätzlich xay mit $\Theta_0 = 0^\circ$

■ zusätzliche "Improper Torsions" E_{tor} :

- z.B. Torsionswinkel acxb , mit Minimum bei 180°





Die Potentialhyperfläche

■ Energie = f (Struktur)

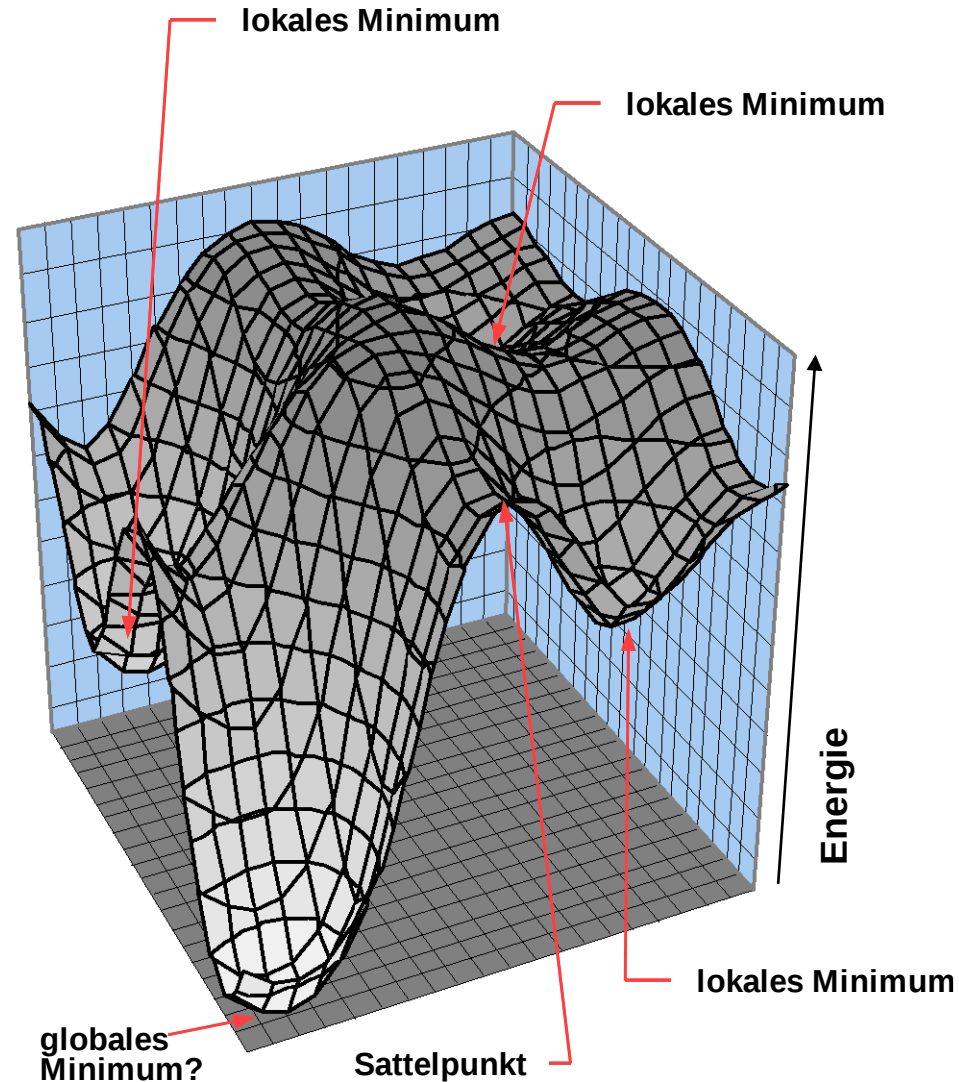
- $E_{\text{pot}} = E_{\text{bnd}} + E_{\text{ang}} + \dots$
- $E_{\text{pot}} = f(\text{allen Atomkoordinaten})$

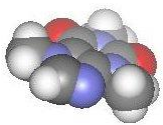
■ innere Freiheitsgrade

- je Atom 3 Koordinaten (x,y,z)
- bei N Atomen: 3N Koordinaten
- - 3 für Gesamttranslation
- - 3 für Gesamtrotation
- $\Rightarrow 3N - 6$ Freiheitsgrade

■ Potentialhyperfläche

- $E_{\text{pot}} = f(3N-6 \text{ Koordinaten})$
- Oberfläche mit Minima, Maxima im 3N-5-dimensionalen Hyperraum
- vorstellbar nur bei f (2 Koordinaten) im 3-dimensionalen Raum (siehe rechts)





Optimierungsmethoden

- steepest descent
- conjugate gradient (Polak-Ribière)
- block diagonal (Newton Raphson)

Gradient **g**

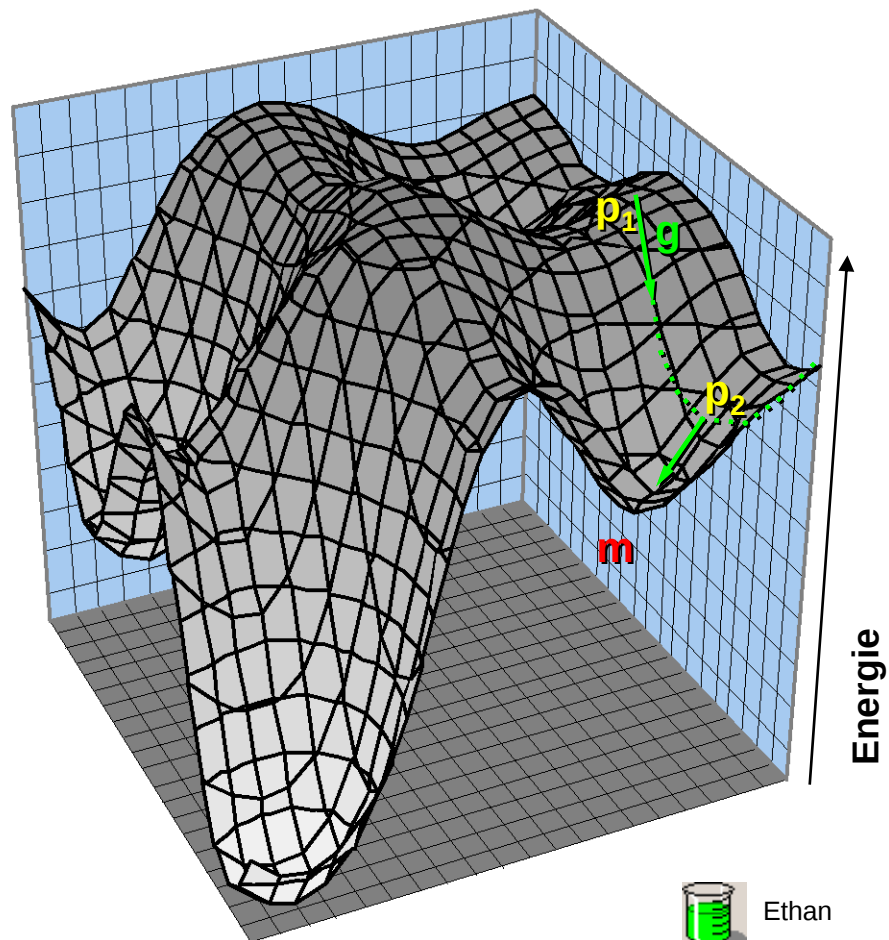
- Vektor zeigt an einem Punkt der Oberfläche "nach unten"
- $g = -\text{grad}(E_{\text{pot}})$

$$g = - \left(\frac{\delta E}{\delta x_1}, \frac{\delta E}{\delta y_1}, \frac{\delta E}{\delta z_1}, \frac{\delta E}{\delta x_2}, \frac{\delta E}{\delta y_2}, \frac{\delta E}{\delta z_2}, \dots \right)$$

Minimum **m**

- Der Gradient "verschwindet"
- $\text{RMS}(g) \approx 0$

$$\text{RMS}(g) = \frac{1}{3N} \left(\sum_{\text{Atome}} \left(\frac{\delta E}{\delta x_A} \right)^2 + \left(\frac{\delta E}{\delta y_A} \right)^2 + \left(\frac{\delta E}{\delta z_A} \right)^2 \right)^{1/2}$$

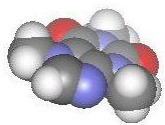


Ethan

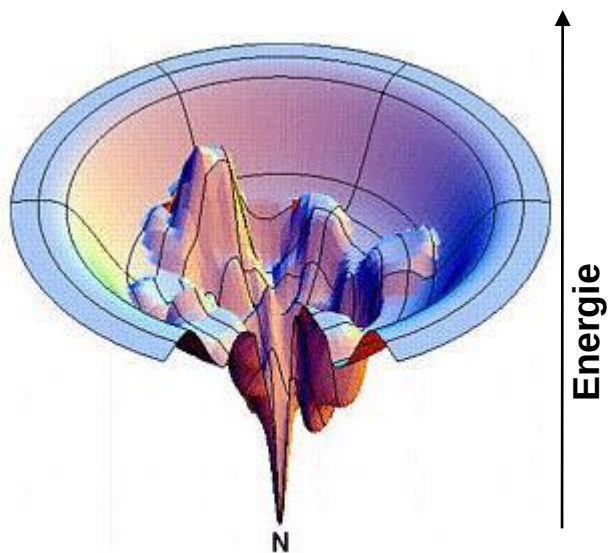
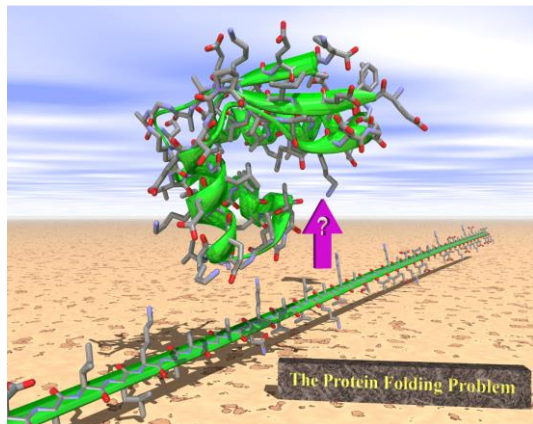


Ethan_step

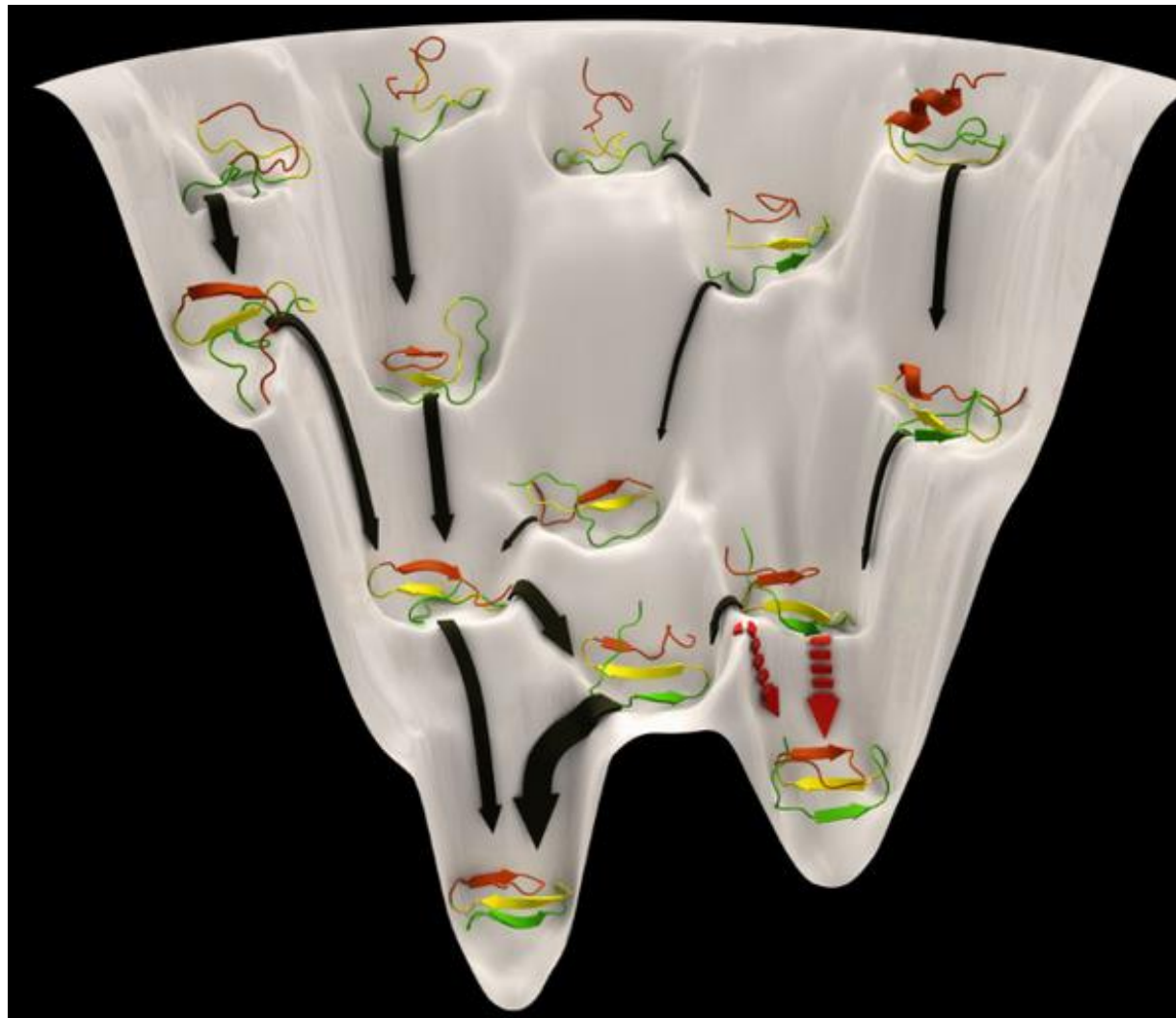


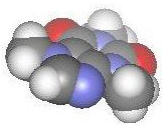


Protein Folding Funnel



N: gefaltetes (natives) Protein



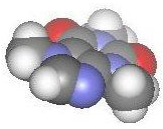


- Bei der Molekulardynamik werden die Atome eines Ensembles (ein Molekül oder mehrere Moleküle) auf der Grundlage der Maxwell-Boltzmann-Verteilung, der Energieerhaltung und der Newtonschen Bewegungsgesetze bei einer definierten Temperatur bewegt.

- Möglichkeiten und Konsequenzen:
 - Die Flexibilität der Moleküle wird sichtbar
 - Überwinden von Potentialbarrieren → Konformationsübergänge sind möglich
 - Suche nach energiearmen Konformationen wird möglich
 - Berechnung thermodynamischer Eigenschaften durch die Generierung statistischer Ensembles



Aspartame
dynamic



- Nach der Maxwell-Boltzmann-Verteilung ist die mittlere kinetische Energie der Teilchen eines Ensembles nach folgender Gleichung mit der Temperatur des Ensembles verknüpft:

$$E_{\text{kin}} = \sum_{\text{Atome}} \frac{m_i}{2} v_i^2 = \frac{3}{2} N k T$$

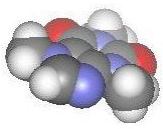
E_{kin} mittlere kinetische Energie der Atome [J]
 m_i Masse des Atoms i [kg]
 v_i Geschwindigkeit des Atoms [m/s]
 N Anzahl der Atome
 k Boltzmann-Konstante ($1,3807 \cdot 10^{-23}$ J/K)
 T abs. Temperatur [K]

- Nach dem Energieerhaltungsgesetz ist die Gesamtenergie des Ensembles als Summe der potentiellen Energie (Spannung) und der kinetischen Energie (Temperatur) konstant.

$$E_{\text{ges}} = E_{\text{pot}} + E_{\text{kin}} = \text{const.}$$

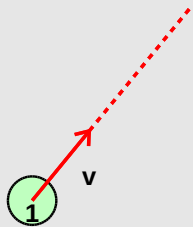


Aspartame
dynamic
non constant temp



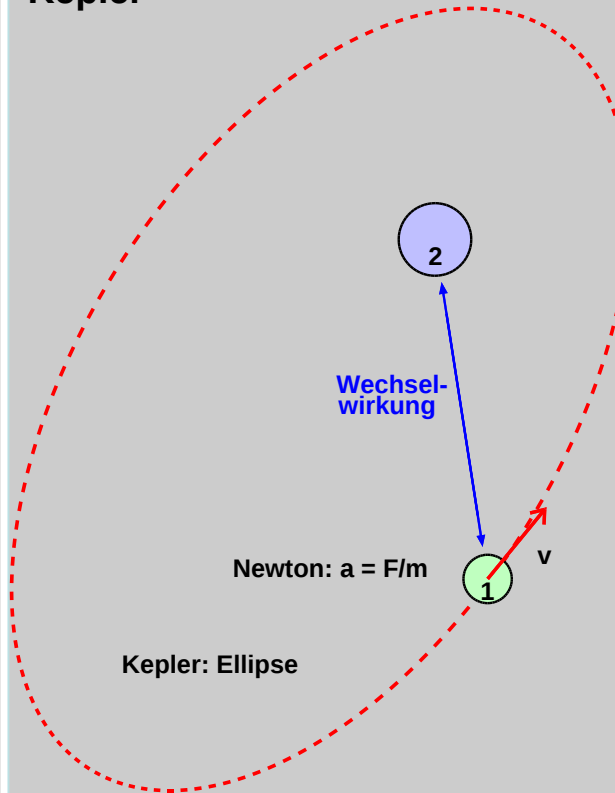
Teilchenbewegung: Newton – Kepler- Chaos

Newton



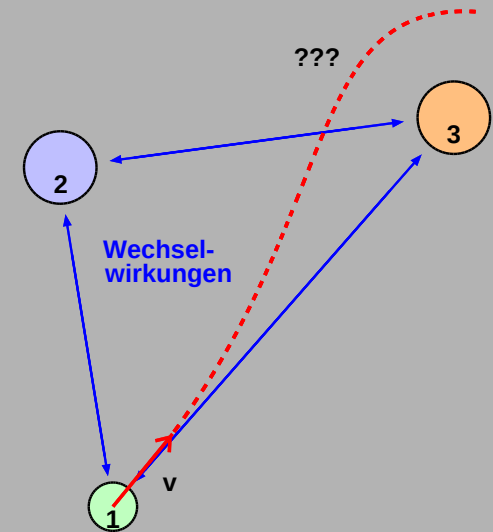
1 Körper:
„Bahnkurve“ ist mathematisch
geschlossen berechenbar

Kepler

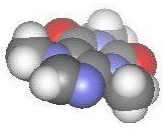


2 Körper:
Bahnkurve ist mathematisch
geschlossen berechenbar

Chaos

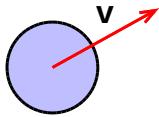


3 Körper (und mehr):
Bahnkurve ist mathematisch
NICHT geschlossen berechenbar,
sondern muss in kleinen
Zeitschritten Δt iteriert werden

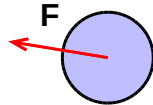


Molekulardynamik - Newton

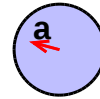
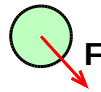
- Durch Anwendung der Newtonschen Bewegungsgleichungen werden die Ortskoordinaten und die Geschwindigkeiten der Atome bei ihrer Bewegung auf der Potentialhyperfläche berechnet. Simulationszeiten ps bis wenige ms.



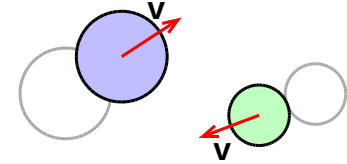
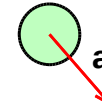
Orte/Geschw.
zum Zeitpunkt t



Kräfte durch
das Potential



Beschleunigungen



Orte/Geschw.
zum Zeitpunkt t+Δt

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \vec{a} \cdot \Delta t$$

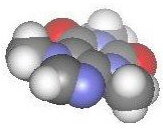
$$\vec{a} = \frac{\vec{F}}{m}$$

$$\vec{F} = -\text{grad}(E_{\text{pot}}) = -\frac{\delta E_{\text{pot}}}{\delta x \delta y \delta z}$$

t	Zeit [s]
Δt	Zeitinkrement ($\approx 10^{-15}$ s)
v	Geschwindigkeit des Atoms
a	Beschleunigung des Atoms innerhalb des Zeitraums Δt
F	Kraft auf das Atom
m	Masse des Atoms
E _{pot}	Potentielle Energie für das Atom
x...	Ortskoordinaten des Atoms

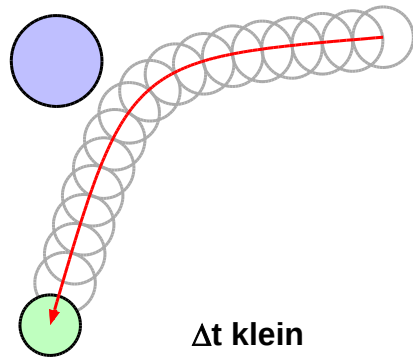


Aspartame
dynamic

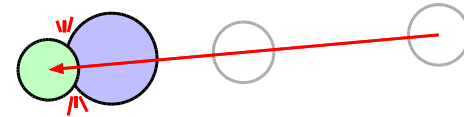


Bedeutung des Zeitinkrements Δt

- Die Wahl der Größe des Zeitinkrements Δt ist von entscheidender Bedeutung. Zu große Zeitinkremente (rechts im Bild) bilden die Trajektorien ("Flugbahnen") der Teilchen zu ungenau ab und führen u.U. zu Instabilität der Rechnungen aufgrund unnatürlich großer Annäherungen.



Δt klein

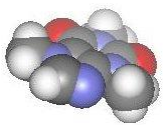


Δt groß

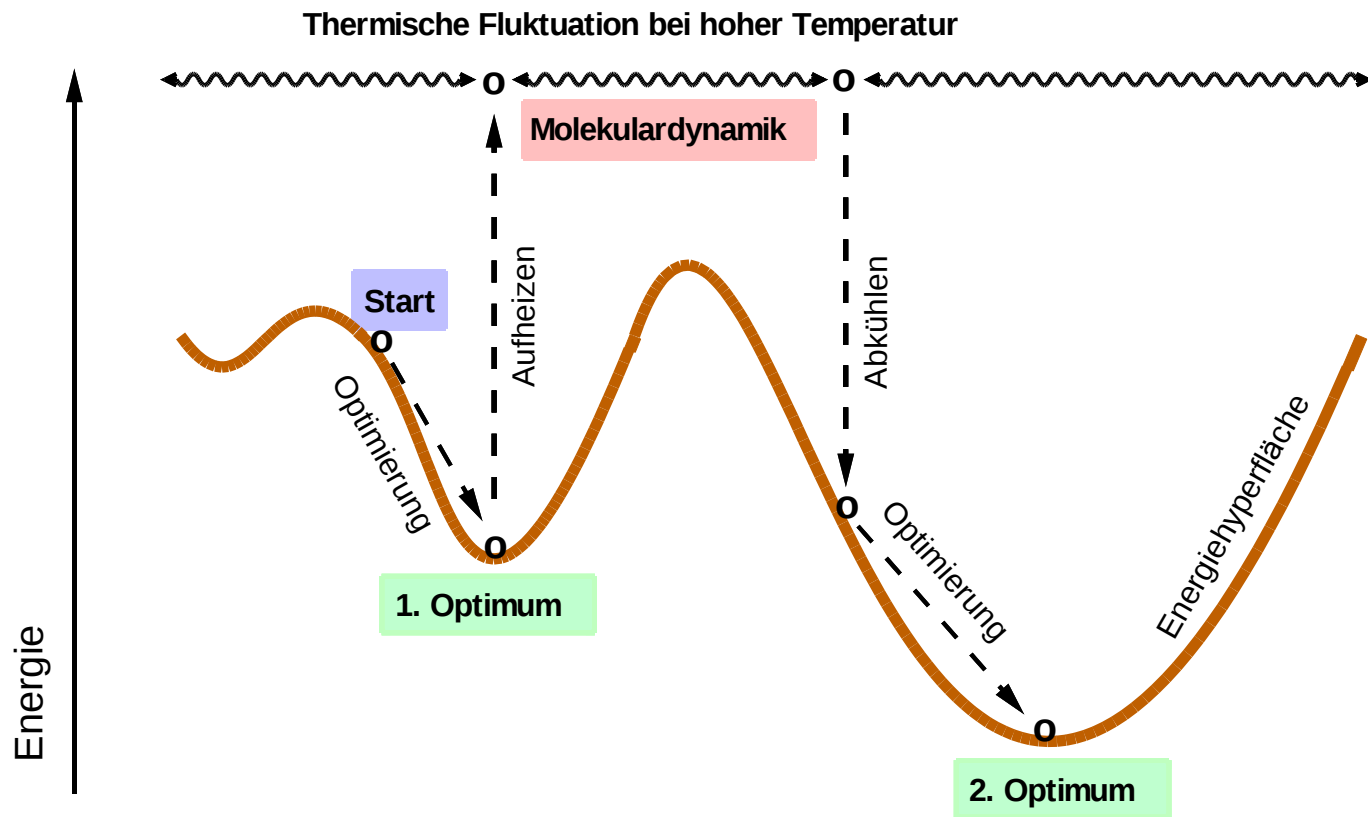
- Kleinere Zeitinkremente führen zu realistischen Trajektorien (links im Bild), vergrößern allerdings die benötigte Rechenzeit.



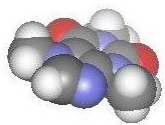
Aspartame
dynamic



Simulated Annealing

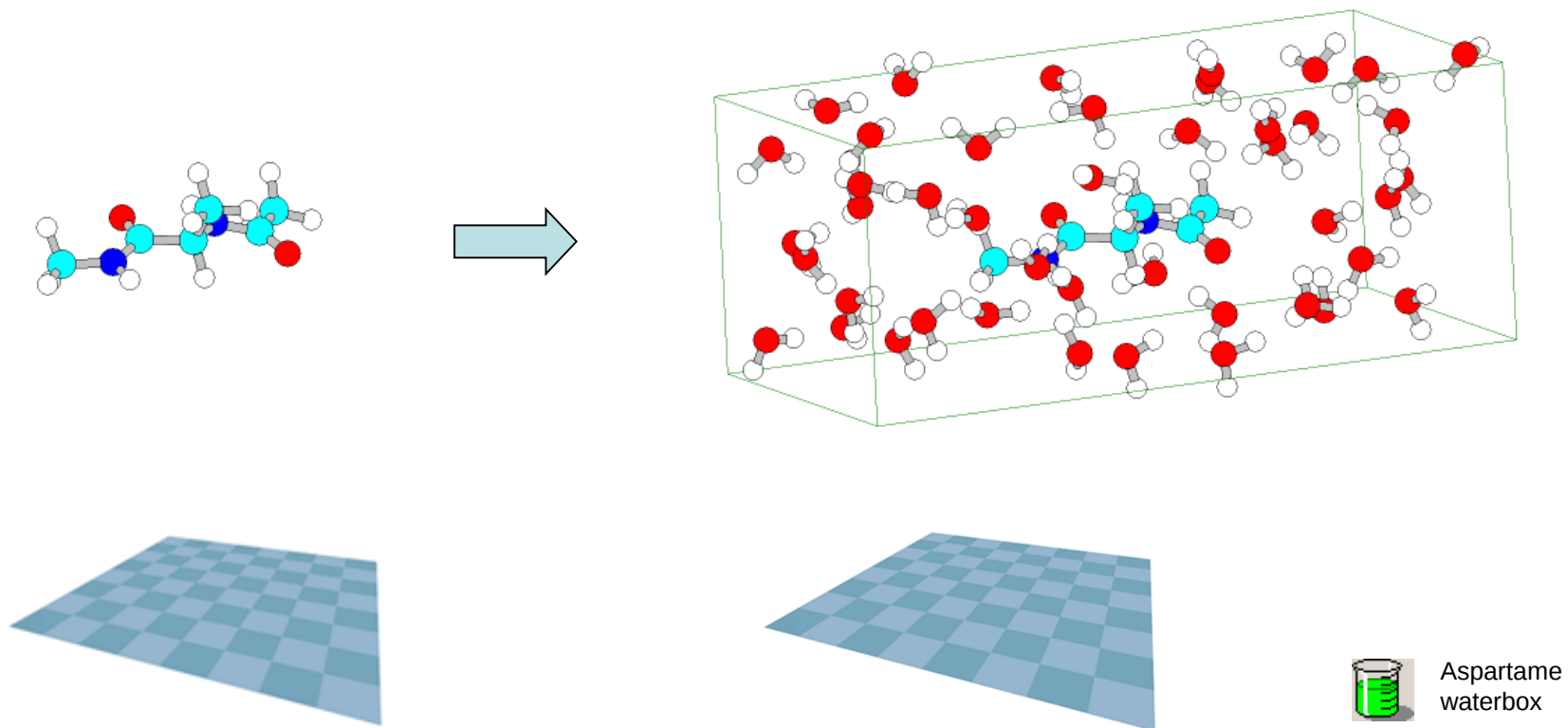


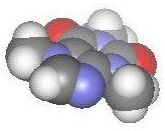
Aspartame
annealing



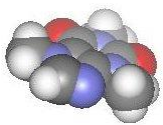
Wasserbox und periodische Randbedingungen

- Die Simulation einfacher Moleküle "im Vakuum" vernachlässigt die Wechselwirkungen mit dem Umfeld.
- Eine sich in allen Raumrichtungen selbst wiederholende Wasserbox gibt einem Molekül die "natürliche biologische" Umgebung.

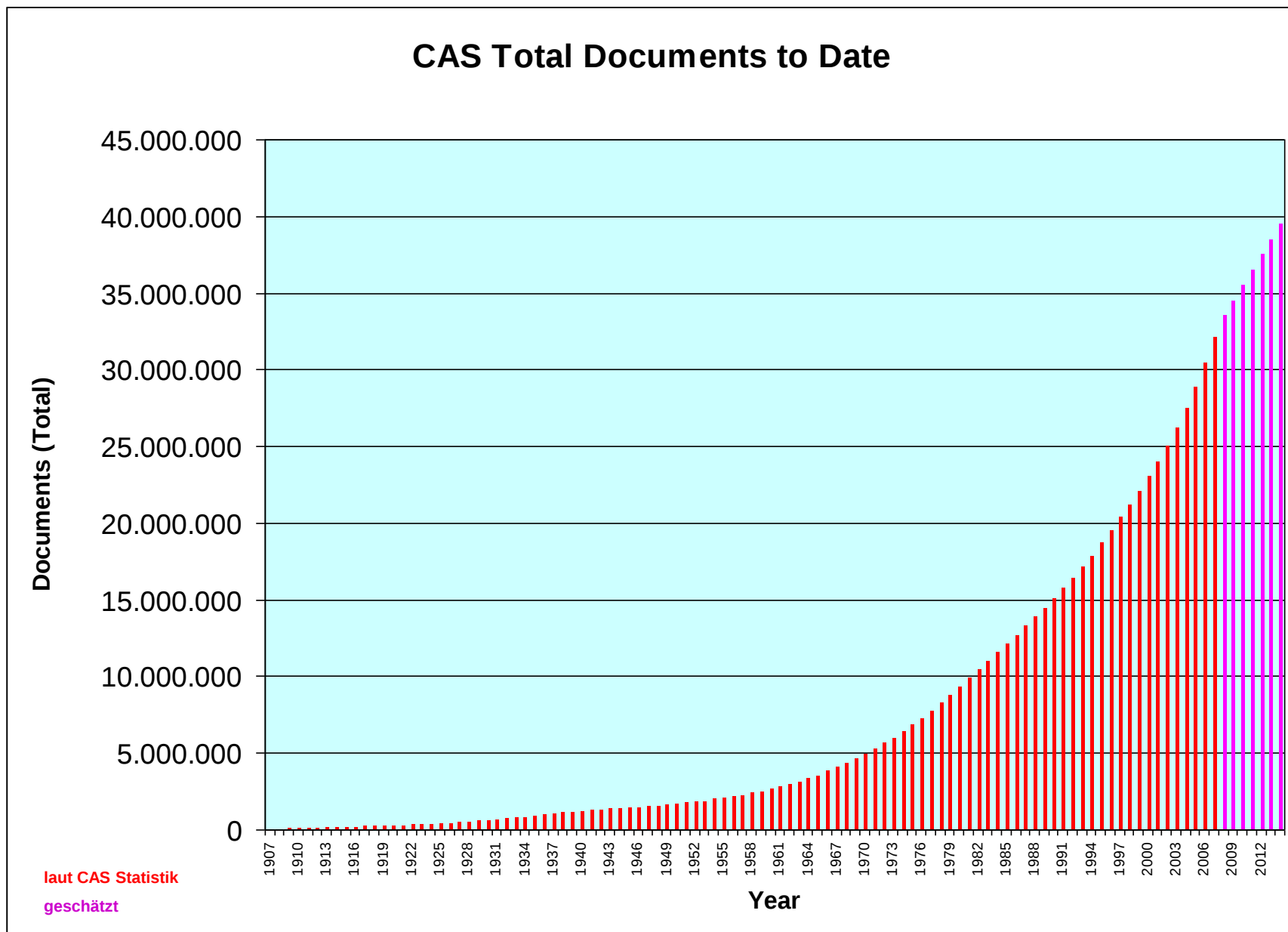


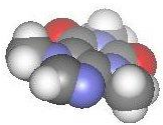


Literatur Online



Publikationen in der Chemie und Biochemie





Chemische Information ist komplex

■ Texte

- Namen, Charakterisierungen, Sachverhalte
- Eigenschaften, Beziehungen
- Literaturstellen etc.

■ Numerische Daten

- Schmelzpunkt, Siedepunkt, Dichte
- Spektren, Preis etc.

■ Strukturen

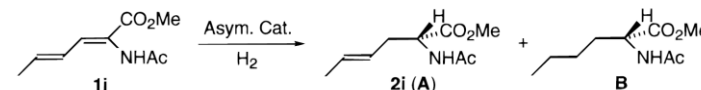
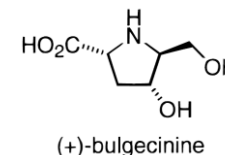
- Konstitution, Konfiguration und Konformation
- Koordinaten, Biopolymersequenz

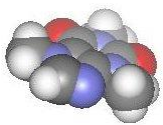
■ Reaktionen

- Synthesepläne, Reaktionsmechanismen

Horner–Emmons olefination¹⁰ (route B) served as an alternative to the cross-coupling. The olefination reaction between phosphonate **5** and a series of unsaturated aldehydes **6** proceeded in THF at $-78\text{ }^{\circ}\text{C} \rightarrow \text{rt}$ with tetramethylguanidine to give (*Z*)-dienamide esters in fair yields.¹¹ A greater diversity of functionalized dienamide esters could be made in this manner, but yields were generally lower when compared to the cross-couplings, especially in the preparation of γ -substituted enamido olefins ($R' \neq \text{H}$) such as **1f,h**. Overall the two methods for the preparation of the α,γ -dienamide esters were complementary and provided ready access to a variety of amino acid precursors. The results of both methods are summarized in Table 1.¹²

(2*R*,4*E*)-Methyl 2-Acetamido-6-((*tert*-butyldimethylsilyl)oxy)hex-4-enoate (2d**).** **General Procedure.** Methyl 2-acetamido-6-((*tert*-butyldimethylsilyl)oxy)hexa-2,4-dienoate (**1d**) (1.7 g, 5.4 mmol) and [(COD)Rh(*R,R*)-Et-DuPHOS]OTf (7.0 mg) were dissolved in MeOH (125 mL), and the vessel was charged with H₂ (90 psi). After 2 h the H₂ was vented and the solvent was evaporated. The crude material was passed through a short plug of silica with acetone before the optical purity was assayed by GC. Chromatography on silica gave pure **2d** (1.6 g, 95%): *R_f* 0.39 (1:2 system A); *t_R* 7.36 min (160 $^{\circ}\text{C}$, 30 psi, >99% ee); $[\alpha]_D^{20} -45.8^{\circ}$ ($c = 1.25$, CHCl₃); ¹H NMR δ 0.02 (s, 6H), 0.86 (s, 9H), 1.98 (s, 3H), 2.40–2.60 (m, 2H), 3.70 (s, 3H), 4.08 (dd, 2H, *J* = 1.2, 4.8 Hz), 4.60–4.70 (m, 1H), 5.40–5.60 (m, 2H), 5.60 (br d, 1H, *J* = 7.5 Hz); ¹³C NMR δ 18.3, 23.1, 25.9, 34.8, 51.8, 52.3, 63.2, 123.7, 134.1, 169.6, 172.3; HRMS calcd for C₁₅H₃₀NO₄Si (*M* + H⁺) *m/z* 316.1944, found 316.1943.





■ Warum wird publiziert?

- Kulturelles Erbe, Wissenserhalt und Fortschritt
- Reputation, Eitelkeit, Mittelvergabe

■ Wie wird publiziert?

■ Alchemie

- stark verschlüsselt
- oft bewusst irreführend

■ moderner Stil seit 1800-1840

- sachlich eindeutig
- geprüft
- formale Kriterien
- nachvollziehbar



N.Flamel
1330–1418



Ladenburg
1875



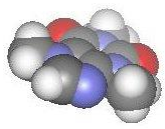
JACS
1998

■ Empfehlung zum Lesen

- Carl Djerassi "Cantor's Dilemma"

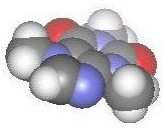


Djerassi

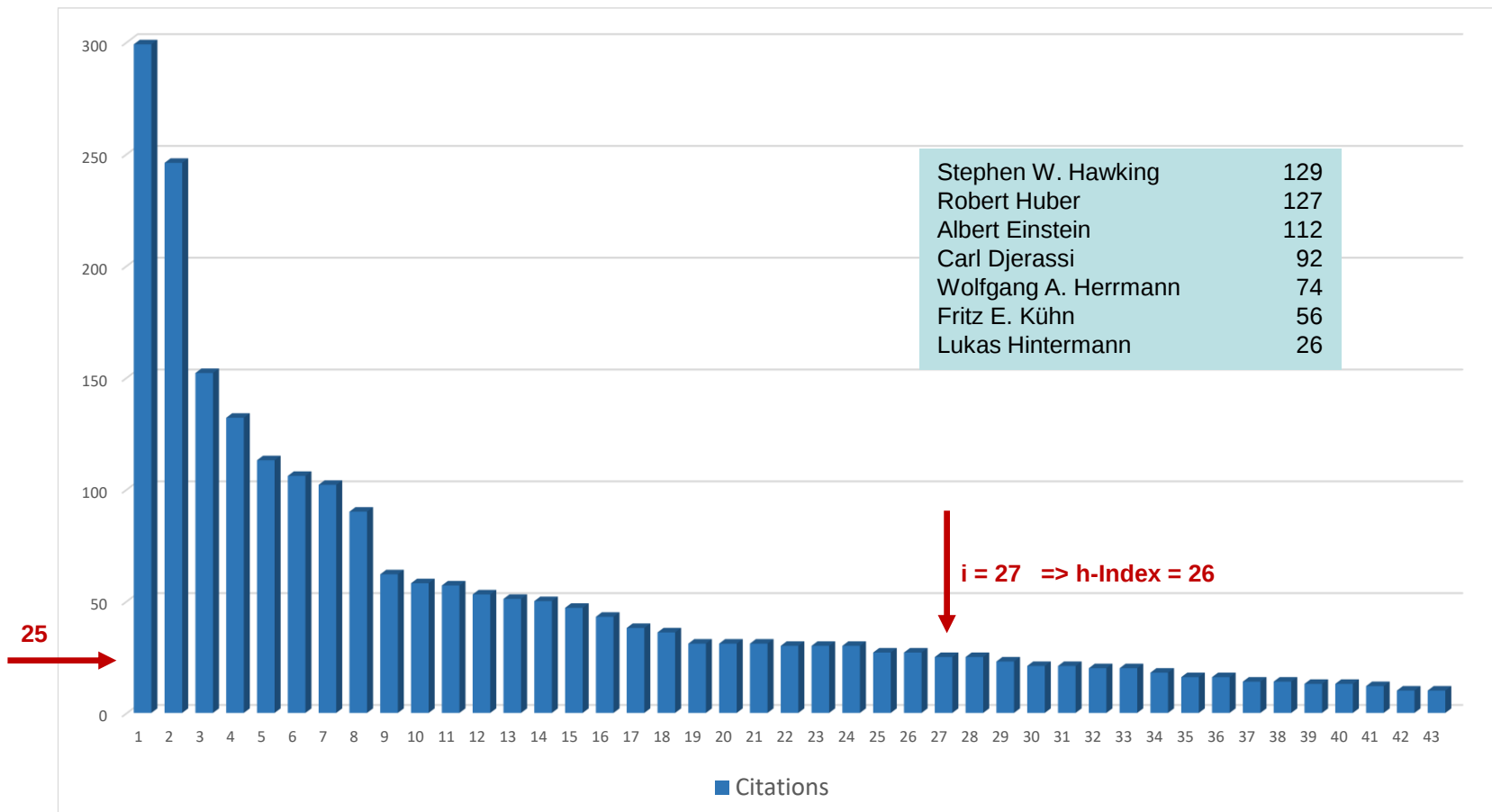


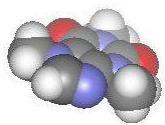
Ablauf einer wissenschaftlichen Publikation

- **1. Durchführung der wissensch. Arbeit**
- **2. Einigung auf Autorenliste**
 - Wer war wesentlich beteiligt?
 - Reihenfolge! Besonders erste und letzte Position: "J. Buchner et al."
- **3. Manuskript erstellen**
 - Nach den Vorgaben des Verlags
- **4. Bei einer (!) Zeitschrift einreichen**
 - Nicht parallel in mehreren!
- **5. Formale Prüfung durch den Verlag**
 - Sind alle wichtigen Sachen dabei?
- **6. Qualitätssicherung**
 - Gutachten erstellen (peer reviewing) durch konkurrierende Wissenschaftler
- **7. Änderungen und**
 - weiter bei 4. (Revising)
 - oder Rücknahme und ggf. Einreichung bei anderer Zeitschrift
- **8. Publikation der Arbeit**
 - Online bei Verlag und ggf. auf Papier
- **9. Aufnahme in Datenbanken**
 - Bibliografische Informationen zur Recherche
- **10. Publikation wird zitiert**
 - => Anstieg des impacts der Arbeitsgruppe
 - => Anstieg des impacts der Zeitschrift



- Die Betrachtung von Publikations- oder Citationszahl alleine ist nicht sinnvoll
- Der h-Index ist eine 2005 vom Physiker Jorge E. Hirsch vorgeschlagene Bewertung
 - Sortiere Publikationen nach fallender Citations-Zahl.
 - Durchzählen, bis die i-te Publikation weniger als i Citations hat. Dann ist $h\text{-Index} = i - 1$.
 - Der h-Index sollte allerdings kritisch betrachtet werden (wie alle bibliometrischen Bewertungen)!





Wer bezahlt ? => Open Access

■ Konzentration der Wissenschaftsverlage

- Früher: sehr viele Verlage
- Heute: Elsevier, Springer, Wiley-Blackwell, Taylor & Francis, Sage

■ Gewinnspannen bei 40% !

■ Preissteigerungen 10-15% pro Jahr !

■ Der Staat zahlt zweimal!

- bei der Forschungsförderung
- beim Literaturzugriff an Forschungseinrichtungen und Universitäten

■ Die Wissenschaft in armen Staaten hat keinen Zugriff auf die Literatur !



■ Aktuelle Situation

- The reader pays
- Folge: starke Engpässe an Forschungseinrichtungen und Universitäten
- Ungerechte Behandlung von Entwicklungsländern

■ Umbruch zu Open Access

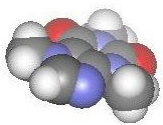
- The author pays
- Folge: weltweit frei verfügbare Literatur
- Vereine oder Forschungseinrichtungen unterstützen die Finanzierung

■ Mehrere Modelle

- Gold: Open Access Verlag übernimmt Veröffentlichung
- Green: Veröffentlichung über freie Publikation auf Doc-Server (z.B. eigene Website o.ä.)

■ Achtung! Neue illegale Pseudoverlage!

- Kein peer-review QM-System!
- Organisation von fake-Konferenzen!
- Einkassierung der Gebühren ohne Gegenleistung!
- Robert Huber: „Das ist Betrug und muss verfolgt werden!“
- Beispiele: [WASET](#) oder [OMICS](#)



■ Primärliteratur

■ Originalliteratur (wissenschaftliche Aufsätze)

- Zeitschriften, Patentschriften, Konferenzberichte
- Monographien, Dissertationen
- Firmenschriften, Web-Seiten



JACS

■ Sekundärliteratur

■ Chronologisches Sammeln und Erschließen (Schlagwortregister, Indices)

■ Referieren ohne kritische Wertung

- Chemical Abstracts (CAS), PubMed, Medline
- Web of Science (WoS), Current Contents



CAS

■ Tertiärliteratur

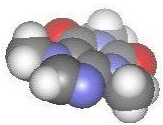
■ Didaktische Aufarbeitung mit sachlich-logischer Ordnung

■ Kritisch-wertende Sicht

- Übersichtsartikel (Reviews), Lehrbücher
- Reaxys (früher Beilstein + Gmelin), Houben-Weyl (jetzt: Science of Synthesis),
Theilheimer Datenbank, Römpp Lexikon

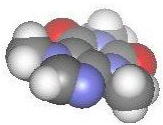


ChemRev



- **Online verfügbare Primärliteratur**
- **Ergänzung und teilweiser Ersatz der PaperWare in der Bibliothek**
- **Verfügbar im Hochschulnetz**
- **nötig: Web-Browser mit PDF-Reader**
- **Adresse:**
 - www.bibliothek.uni-regensburg.de/ezeit/ezb.phtml?bibid=TUM





■ Bibliographische Datenbanken

- Chemical Abstracts (CAS), PubMed, Medline
- Web of Science (WoS), Current Contents

■ Fakten-Datenbanken

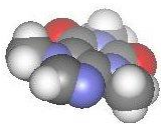
- Reaxys (Beilstein + Gmelin), SpecInfo

■ Struktur-Datenbanken

- Cambridge Crystallographic Data Center (CSD), Protein Database (PDB)

■ Bioinformatik-Datenbanken

- PROSITE, ExPASy, SWISS-PROT, Entrez (PubMed)



■ Weltweit größte chemische Datensammlung

■ Datenstand :

- 48 Mio. Abstracts von Artikeln, Patenten etc. ab 1907 aus ca. 10.000 Zeitschriften
- 143 Mio. Organische und Anorganische Verbindungen, Biosequenzen, Mineralien, Legierungen
- Zuwachs: ca. 4000 Einträge/d

■ physikalische oder chemische Substanzdaten verfügbar, aber nicht suchbar



■ Kontrollierte Vokabularien

■ Indices und Thesauri (z.B. CAS General Subject Index)

- bis 1956 alle 10 Jahre ein Decennial Index
- danach alle 5 Jahre ein Collective Index

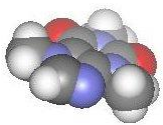
■ CAS-Registry Number (z.B. 6781-23-7) als einfache Substanzreferenz

■ CAS Source Index (CASSI) als anerkannte Journal-Referenz

■ Online über SciFinder

■ scifinder.cas.org





■ Geschichtliches

- Gründer: Friedrich Konrad Beilstein (geb. 1838 St. Petersburg)
- Originaltitel: „Beilsteins Handbuch der Organischen Chemie“
- Beilstein Institut für die Literatur der Organischen Chemie
- Beilstein Informationssysteme GmbH (1998 Elsevier Science)
- Beilstein Online (1988 STN, 1997 CrossFire)

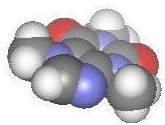
■ Charakteristika

- Stoffdatenbank mit physikalischen und chemischen Eigenschaften
- Ordnungskriterium ist die individuelle chemische Verbindung
- nicht-bibliographische Datensammlung
- Das Literaturmaterial wurde gesichtet und kritisch gewertet

■ Online

- Zusammen mit Gmelin (Anorganische Chemie) über Reaxys
www.reaxys.com





NCBI Databases (früher: Entrez)

- Enthält PubMed, PubChem, Proteine, Gene etc.
- Speziell für Mediziner und Biochemiker
- Adresse:
 - <http://www.ncbi.nlm.nih.gov/gquery>

NCBI

Entrez search and retrieval system

Search PubMed for [] Go Clear

☐ Limits Index History Clipboard

About Entrez

SITE MAP

PubMed Help
help documentation for PubMed

Entrez Help
help documentation for the rest of Entrez

The Entrez Databases
cross-references and information

Network Entrez
TCP/IP-based client-server version of WWW Entrez

Batch Entrez
retrieve a large number of sequences

Making WWW Links to Entrez
linking to PubMed and GenBank

bibliographic data from MEDLINE and other literature databases

sequences derived from the GenBank database

data sets of population sequences

sequences derived from the GenBank database

3-dimensional protein structures

assemblies of complete genomes

Entrez provides integrated access to biomedical literature, nucleotide and protein sequences, along with 3-D protein structures, complete genomes and population study data sets. Pre-computed similarity searches are available for most database records producing a list of related sequences, structure neighbors, as well as related articles.

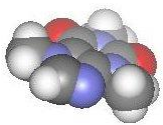
The Entrez search and retrieval system is provided by the [National Center for Biotechnology Information](#). NCBI also builds, maintains, and distributes the [GenBank](#) sequence database.

Revised: March 1, 2000.

[Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)

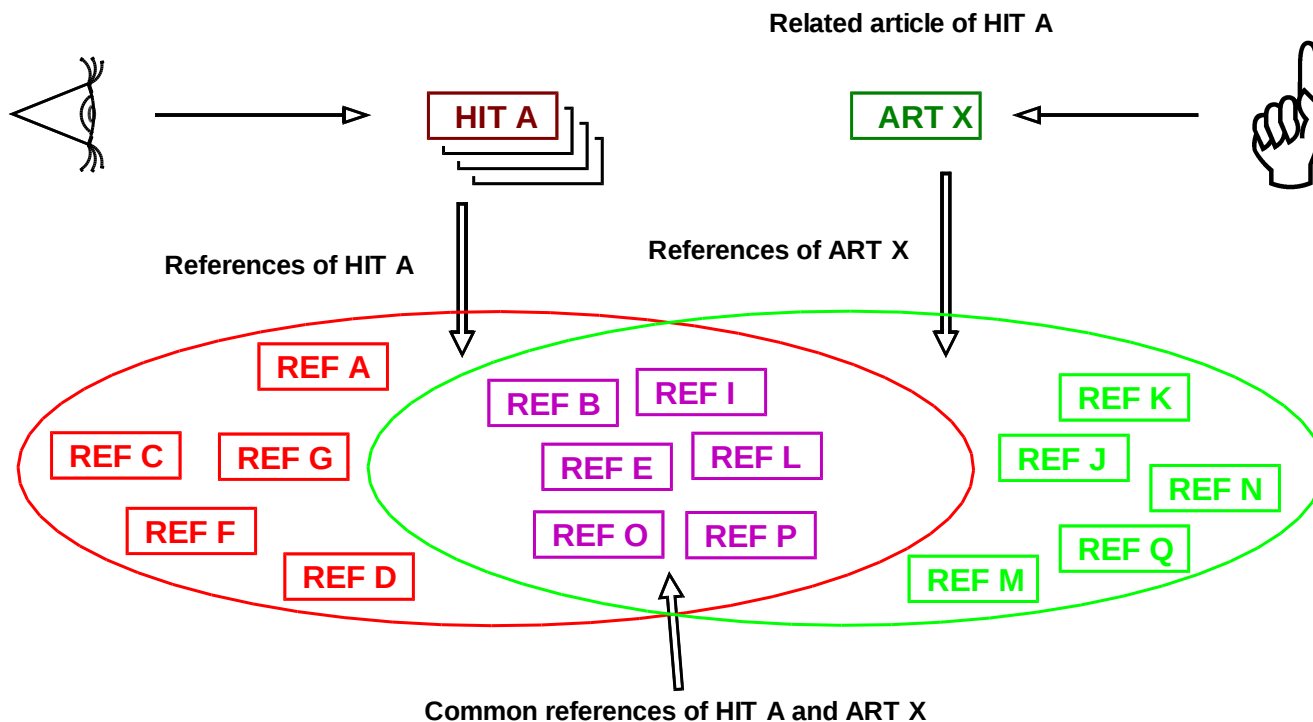


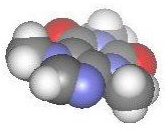
NCBI Databases



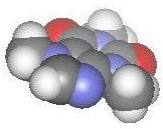
Mehr Hits aufgrund gemeinsamer Referenzen

■ Related Articles





Protein Datenbanken und Services



Hierarchische Ordnungen in der Proteinviefalt

■ nach der Enzym-Funktion

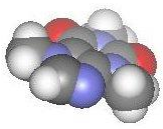
- E.C. Number (Enzyme Commission)

■ nach der Protein-Struktur

- CATH (Class, Architecture, Topology, Homologous Superfamily)
- alternativ: SCOP (Structural Classification of Proteins)

■ nach der Protein-Sequenz

- PROSITE (Consensus Patterns)



Proteinklassifizierung nach der Funktion

■ Enzyme Commission (E.C.) Number

■ Beispiel Lysozym E.C. 3.2.1.17

■ 1. Zahl : Hauptklasse

1 : Oxidoreduktasen (H,e)

2 : Transferasen (Gruppenübertrg.)

3 : Hydrolasen (Hydrol. Spaltung)

4 : Lyasen (Eliminierung)

5 : Isomerasen

6 : Ligasen (ATP, Kondensation)

7 : Translocasen (NEU 2018)
(Membrantransport)

■ 2. Zahl : Subklasse

3.1 : auf Esterbindungen

3.2 : auf Glycoside

3.4 : auf Peptide

3.7 : auf C-C-Bindungen

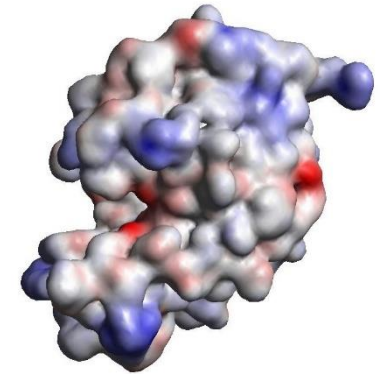
■ 3. Zahl : Sub-Subklasse

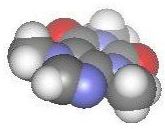
3.2.1 : auf O- und S-Glycoside

3.2.2 : auf N-Glycoside

■ 4. Zahl : Laufende Nummer

3.2.1.17 : Lysozym, 1,4-*N*-Acetylmuramidase





Proteinklassifizierung nach der Struktur

CATH - Klassifizierung

C : Class

- Vorherrschende Sekundärstruktur

A : Architecture

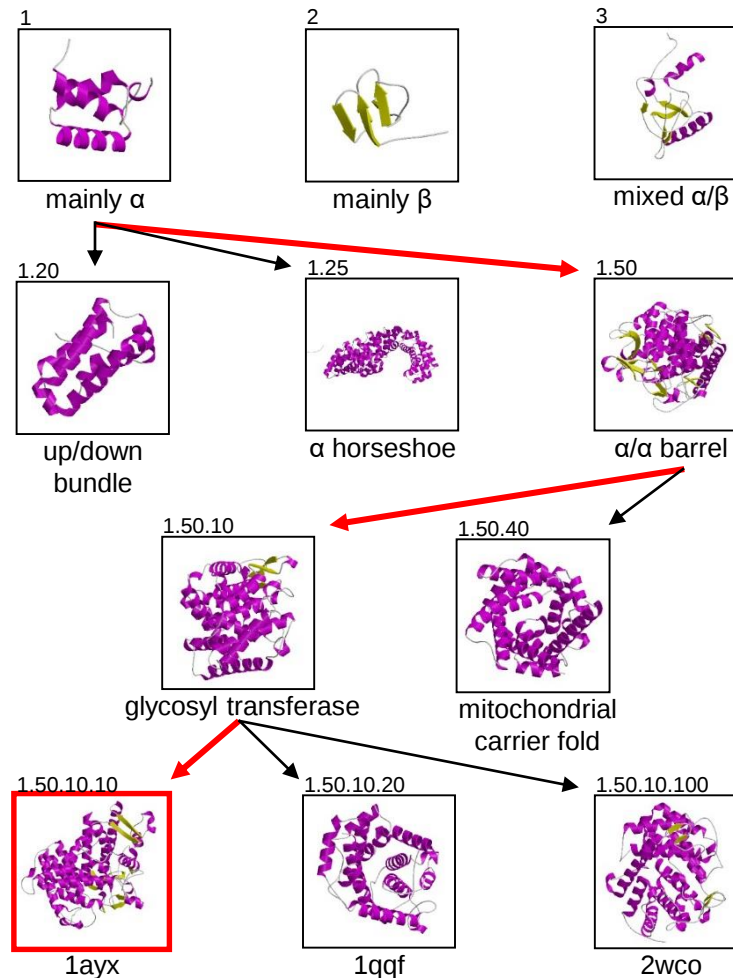
- 3D-Form durch Orientierung der Sekundärstrukturen (ohne deren Verknüpfung)

T : Topology

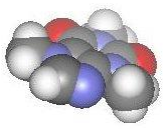
- 3D-Form und Verknüpfung der Sekundärstrukturen

H : Homologous Superfamily

- Verwandschaft durch gemeinsame Vorfahren (genetische Verwandschaft)

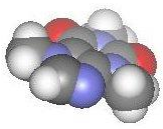


SCOP ist eine konkurrierende Klassifizierungsmethode



Die proteinogenen Aminosäuren

Arg R -4.5 Arginin						Asp D -3.5 Asparaginsäure
Lys K -3.9 Lysin	Ala A +1.8 Alanin	Met M +1.9 Methionin	Thr T -0.7 Threonin	Ser S -0.8 Serin	Gln Q -3.5 Glutamin	Glu E -3.5 Glutaminsäure
His H -3.2 Histidin	Phe F +2.8 Phenylalanin	Leu L +3.8 Leucin	Gly G -0.4 Glycin	Cys C +2.5 Cystein	Tyr Y -1.3 Tyrosin	Asn N -3.5 Asparagin
Pro P -1.6 Prolin	Trp W -0.9 Tryptophan	Val V +4.2 Valin	Ile I +4.5 Isoleucin	basisch unpolar (hydrophob) polar, ungeladen sauer -4.5 ... +4.5 : Hydropathie-Index (Kyte-Doolittle) außen innen		



Proteinklassifizierung nach der Sequenz

- Consensus Pattern in PROSITE Datenbanken
- Sequenzmuster in Proteinfamilien (Alignment)

*** Original Sequences ***

```
MLAATVLTLLGNAHACSKGTSHEAGIVCRITKPALLVLNHETAKVIQTAFQRASYPDITGEKAM
MALFGALFLALLAGAHAEFFPGCKIRVTSKALELVKQEGRLFLEQELETITIPDLRGKE
MMPGVYALLLLWGLATPCLGLLETVGTTLARIDKDELGKAIQNSLVGGPILQNVLTGTVTSVNQGLLGAGGL
MARGPDTARRWATLVVLAALGTAVTTTNP GIVARITQKGLDYACQQGVLT LQKELEKITIPNFSGNFKI
```

*** Optimal Multiple Alignment ***

```
--MLAATVLT LA-LLGNAHACSKGTS-HEAGIVCRITKPALLVLNHETAK---VIQTAFQRASYPD--ITGEKAM
--MALFGALFLA-LLAGAHA-----EFP GCKIRVTSKALELVKQEGLR---FLEQELETITIPD--LRGKE--
--MMPGVYALLL-LWGLATPCLGLL--ETVGTTLARIDKDELGKAIQNSLVGGPILQNVLTGTVTSVNQGLLGAGGL
MARGPDTARRWATLVVLAALGTAVTT-TNP GIVARITQKGLDYACQQGVL---TLQKELEKITIPN--FSGNFKI

.....o.#...#.....#...#o...#.....o.....o.....#....
```

- Beispiel für ein Consensus Pattern:

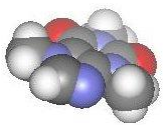
■ [NS] – G – [DNT] – [IMVY] – x – {STGDN} – [DN] – x(2) – [SGAP] – x(3,4) – G

Codes:

- Normale Einbuchstabencodes, getrennt durch "-"-Zeichen
- [] sind alternative AS an dieser Position
- { } sind an dieser Position nicht vorhanden
- x steht für jede AS
- (n,m) wiederholt den Vorgänger n-m mal



ScanProsite.txt



Biochemical Pathways Boehringer (Roche)

■ Biochemical Pathways (Gerhard Michael) by Roche (Boehringer Mannheim):

<http://www.roche.com/pathways>

