

# Inhalt der Vorlesung „Computational Chemistry“

**Inhalt:** Beschreibung von Molekülen mit computerbasierten Methoden

→ insbesondere Moleküleigenschaften:

- **Struktur:** die Molekülgeometrie

*welche Gestalt haben Moleküle ?*

- **Konformationsraum** von Molekülen

*welche Anordnungen der Atome sind sinnvoll ?*

- zeitliche **Bewegung** von Molekülen

*wie finden Konformationsänderungen statt?*

- **Docking**

*wo bindet ein Ligand an ein Protein?*

- Berechnung von **Interaktionsenergien**

*wie stark bindet ein Ligand an ein Protein?*

Methoden

Kraftfelder bzw.

Molekülmechanik (MM)

Quantenmechanik (QM)

Energieminimierung

Samplingmethoden

Moleküldynamik (MD)

Kraftfelder

Scoringfunktionen

# Was diese Vorlesung nicht behandelt

**(1) Drug Design**

**(2) Medizinische und pharmazeutische Aspekte**



Spezielle Lehrveranstaltungen, zum Teil für den Masterstudiengang

**(3) Grundlagen aus der Chemie**

- Oktettregel
- Stöchiometrie

**(4) Grundlagen aus der Mathematik (Analysis):**

- Ableitungen
- Integrale

**(5) Homologiemodellierung** siehe Softwarewerkzeuge der Bioinformatik

# Was ist Computational Chemistry?

## Computational Chemistry:

Arbeitsgebiet an der Schnittstelle von

theoretischer Chemie (Moleküleigenschaften und -struktur),

*Molecular Modeling* (Docking) und

struktureller Bioinformatik (Homologiemodelle, Proteinfaltung).

## Haupteinsatzbereich von Computational Chemistry:

finde mittels numerischer Rechnungen Antworten auf chemische/biologische/pharmazeutische Probleme.

→ Vorhersage von Moleküleigenschaften

oftmals im Zusammenhang mit experimentellen Fragestellungen:

Welches Syntheseprodukt liegt vor?

Welche Reaktionstellen kommen in Betracht?

Wie kann ich vorhandene Moleküle hinsichtlich ihrer Eigenschaft verbessern?

# Geschichte der Computational Chemistry

**Geschichte** der Computational Chemistry:

entweder **recht lang** - wenn man ab der Entwicklung der Quantenmechanik in den 1920er Jahren als Ursprung der theoretischen Chemie rechnet, oder **recht jung**, da genaue Rechnungen an Molekülen mit vielen hundert Atomen erst seit der Entwicklung moderner, leistungsstarker Computer in den 1980er Jahren möglich sind.

Computerchemie gehörte stets zu den Wissenschaftsgebieten mit den größten Anforderungen an Rechenleistung.

Ca. 2/3 aller wissenschaftlich genutzten Rechenzeit wird für quantenchemische Applikationen und Moleküldynamik (MD)-Simulationen verwendet.

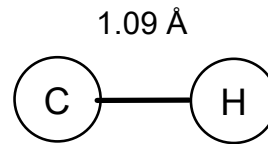
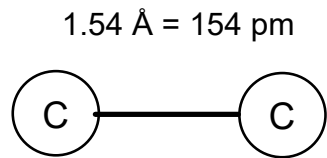


Erwin Schrödinger

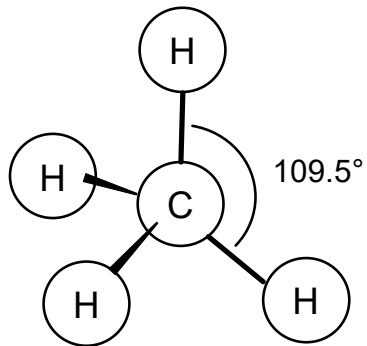


Michael J.S. Dewar

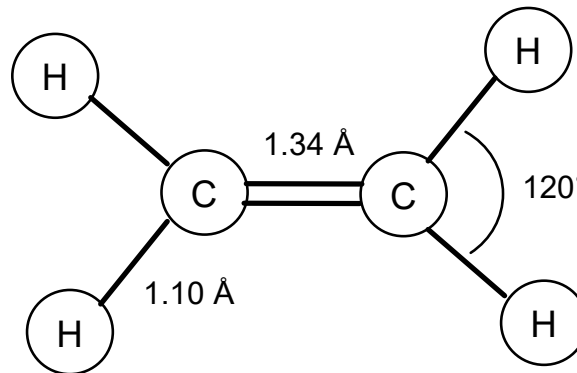
# Molekülgeometrie



Bindungslängen oder  
Bindungsabstände

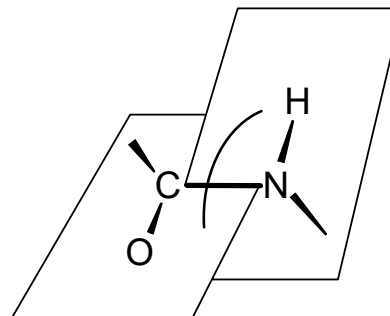
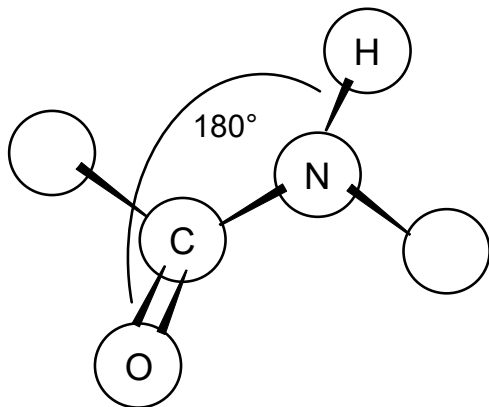


H-C-H tetraedrisch



H-C-H planar

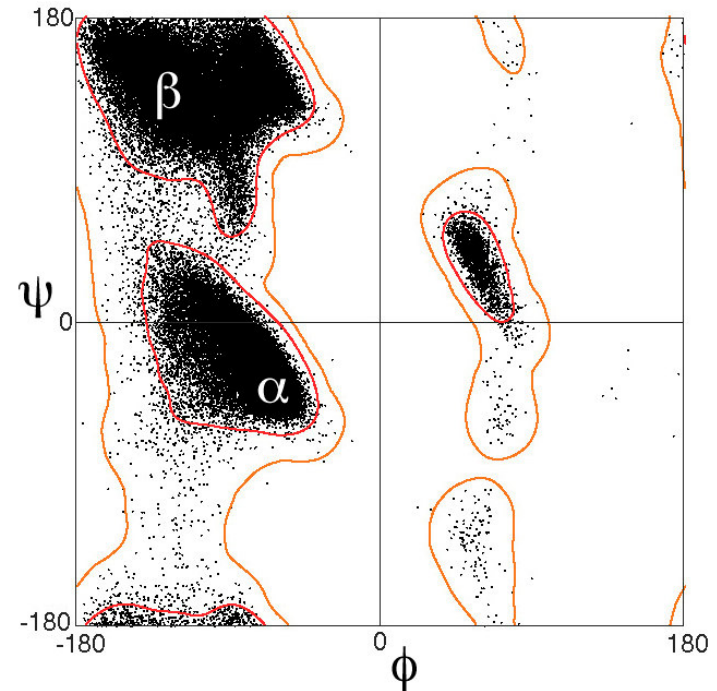
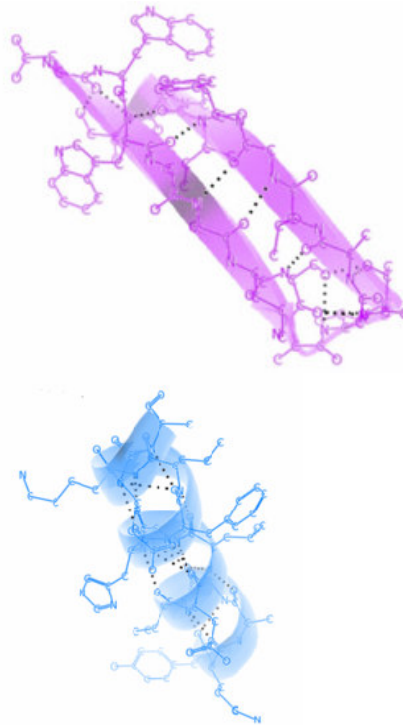
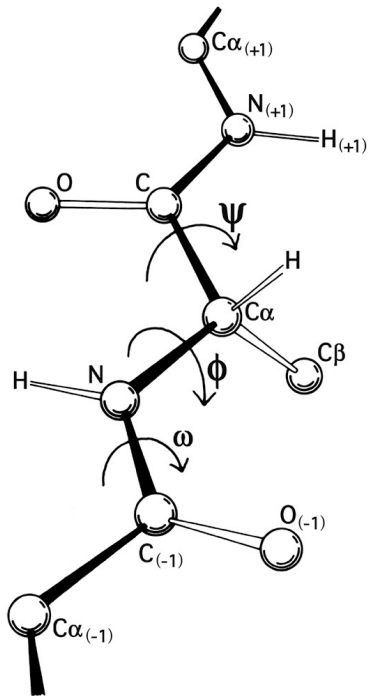
Bindungswinkel



Torsionswinkel oder  
Diederwinkel  
(*dihedral angle*)  
vgl. Peptidbindungen in  
Proteinen

# Bedeutung von Diederwinkeln in der Natur

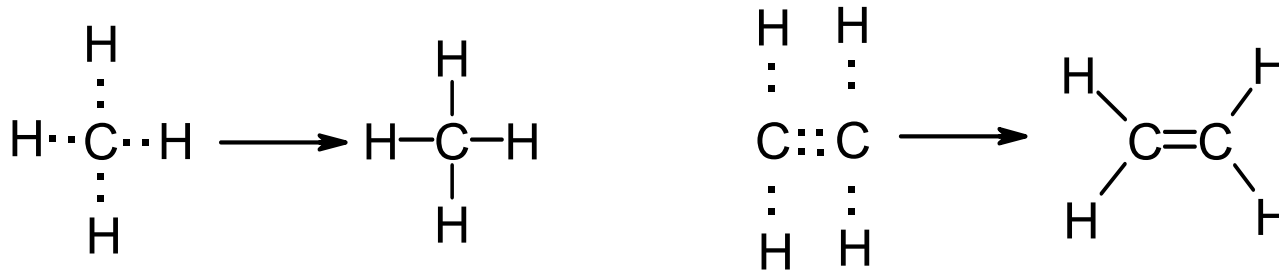
Proteine sind Polypeptide die durch sequenzielle Verknüpfung von Aminosäuren entstehen. Die Sekundärstrukturen werden dabei durch die jeweiligen Diederwinkel  $\phi$  (phi) und  $\psi$  (psi) bestimmt.



Im Ramachandran-Plot finden sich charakteristische Bereiche z.B. für  $\alpha$ -Helices und  $\beta$ -Faltblattstrukturen.

# Darstellung chemischer Strukturen (I)

Die Valenzelektronen der Atome werden paarweise zu Bindungen gruppiert



Diese Darstellung als **Lewis-Strukturen** gibt die kovalenten Bindungen zwischen den Atomen in einem Molekül wieder.

Die Anzahl der Valenz- und Rumpfelektronen kann man aus dem Periodensystem der Elemente entnehmen.

# Das Periodensystem der Elemente

Elemente der ...

Hauptgruppen		Nebengruppen										Hauptgruppen					Rumpf- elektronen	
H													He	0				
Li	Be											B	C	N	O	F	Ne	2
Na	Mg											Al	Si	P	S	Cl	Ar	10
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr	28
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xn	46
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn	
Valenz- elektronen	1	2	2.....2										3	4	5	6	7	

Beispiele:

Na• 10 Rumpf e. + 1 Valenz e.

Na<sup>+</sup> 10 Rumpf e. + 0 Valenz e. = abgeschlossene Schale von Neon

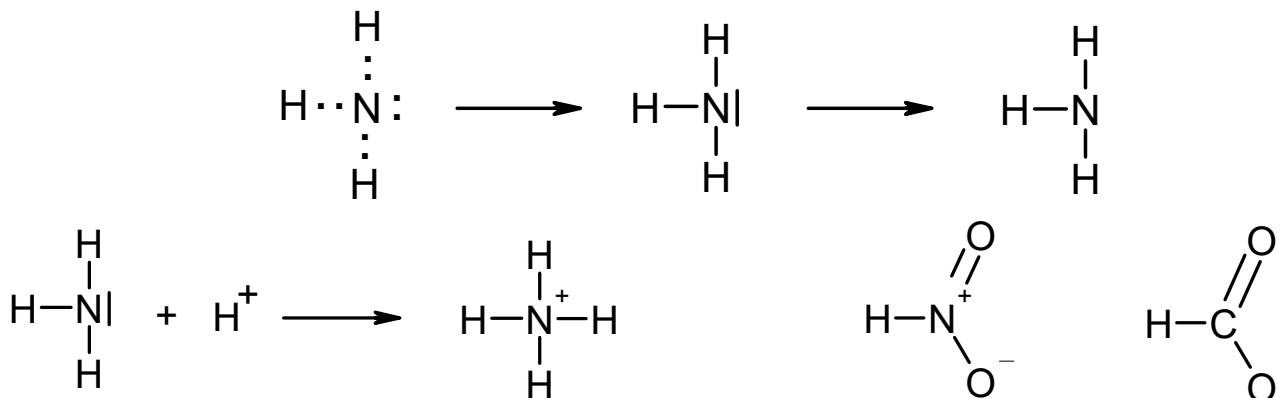
Cl<sup>-</sup> 10 Rumpf e. + 8 Valenz e. = abgeschlossene Schale von Argon

Vollständig aufgefüllte  
Schale bei Edelgasen



## Darstellung chemischer Strukturen (II)

Freie Elektronenpaare, die nicht an einer Bindung beteiligt sind, (engl.: *lone pairs*) werden der Übersichtlichkeit halber oft nicht gezeigt



Hypervalente Atome können mehr Valenzelektronen um sich haben, als nach der Oktettregel zu erwarten:



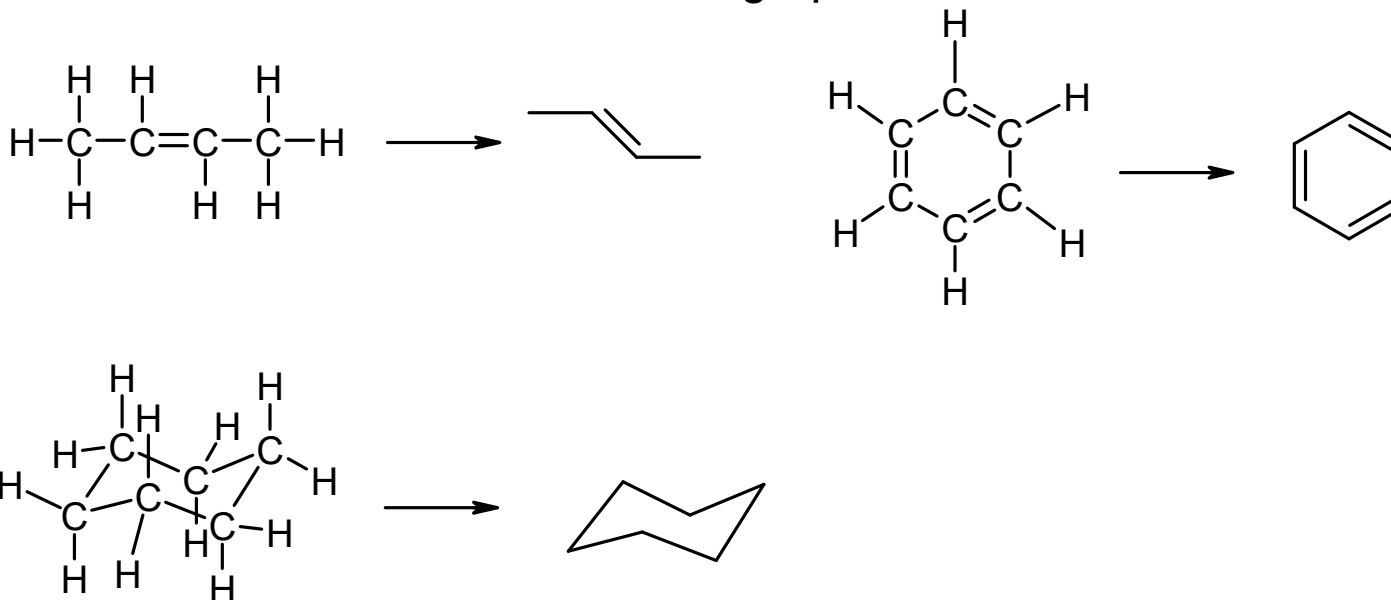
Problem der Lewis-Darstellung:  
Identische Bindungslängen trotz unterschiedlicher Darstellung!  
→ mesomere Grenzstrukturen

Welche Bedeutung haben freie Elektronenpaare für Wasserstoffbrückenbindungen?

## Darstellung chemischer Strukturen (III)

Auch Kohlenstoffatome werden häufig weggelassen

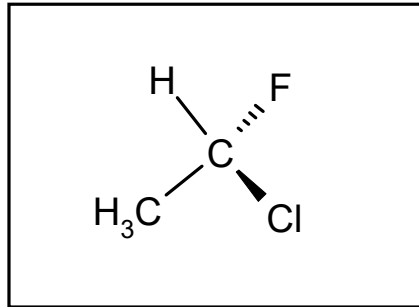
→ sind dann die Ecken des Molekülgraphen



Ecken und die Enden der Kanten stellen Kohlenstoffatome dar, die jeweils mit der entsprechenden Anzahl an Wasserstoffatomen abgesättigt werden, sodaß die Oktettregel erfüllt ist.

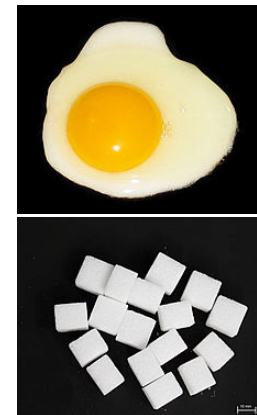
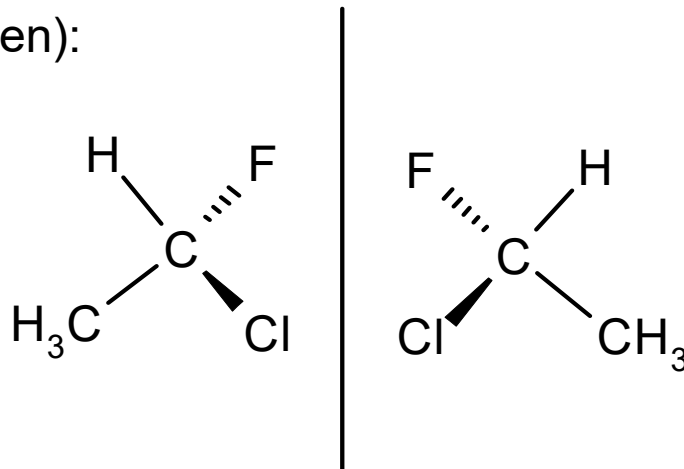
# Darstellung chemischer Strukturen (IV)

## Stereochemie



Keile markieren Bindungen zu Atomen, die aus der Ebene hervortreten; gestrichelte Keile solche, die nach hinten zeigen

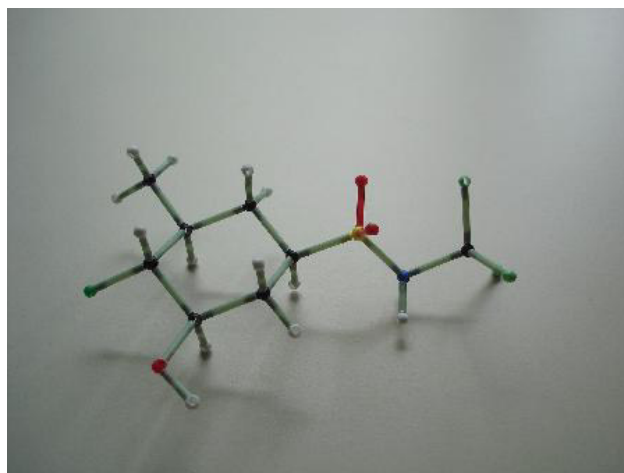
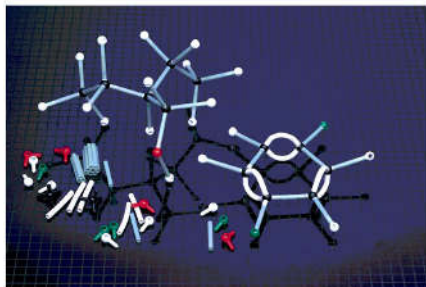
Vier verschiedene Substituenten an einem Kohlenstoffatom bewirken Chiralität (Bild und Spiegelbild lassen sich nicht zur Deckung bringen):



Welche chiralen Moleküle (aus der Natur) sind Ihnen bekannt?

## Zum Anfassen

Molekülbaukästen zur dreidimensionalen Veranschaulichung



Käuflich in verschiedenen Preisklassen erhältlich.

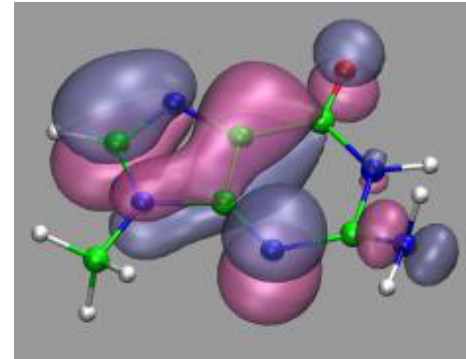
Inzwischen gibt es auch eine Reihe von Computerprogrammen mit denen sich Moleküle betrachten lassen, sodass durch interaktive Rotation ein Eindruck von deren 3-dimensionaler Gestalt gewonnen werden kann.

# Nützliche Software

Visualisierung des Outputs verschiedenster MM, MD, QM, und Docking -Programme, sowie von Protein- und DNA-Strukturen

## vmd

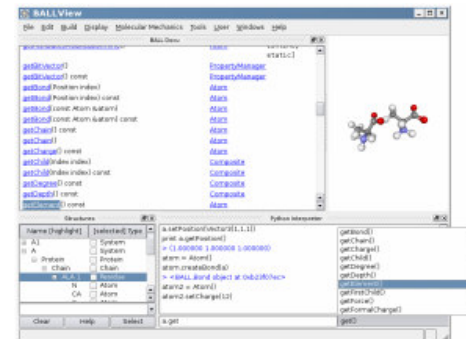
<http://www.ks.uiuc.edu/Research/vmd>



Visualisierung von Molekülen; implementiertes Kraftfelder

## Ball

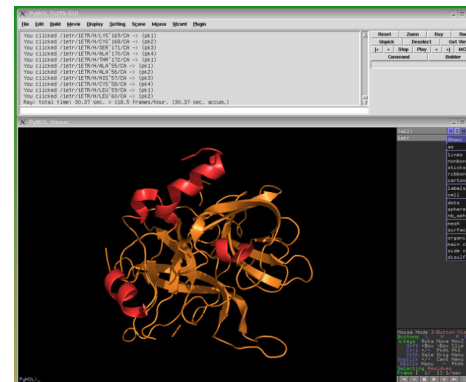
<http://www.bioinf.uni-sb.de/projects/ball.html>



## PyMOL, CHIMERA

[www.pymol.org](http://www.pymol.org)

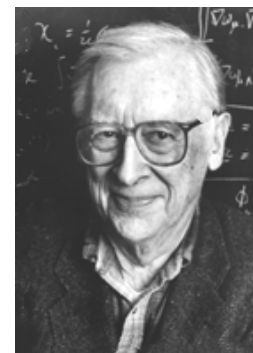
[www.cgl.ucsf.edu/chimera/](http://www.cgl.ucsf.edu/chimera/)



# Erhebung der Computational Chemistry „in den Adelsstand“

Der Ritterschlag für das Gebiet der **Computational Chemistry** war gewissermaßen der Nobelpreis für Chemie in 1998 an

- **John Pople** "for his development of computational methods in quantum chemistry"



- **Walther Kohn** "for his development of the density-functional theory"

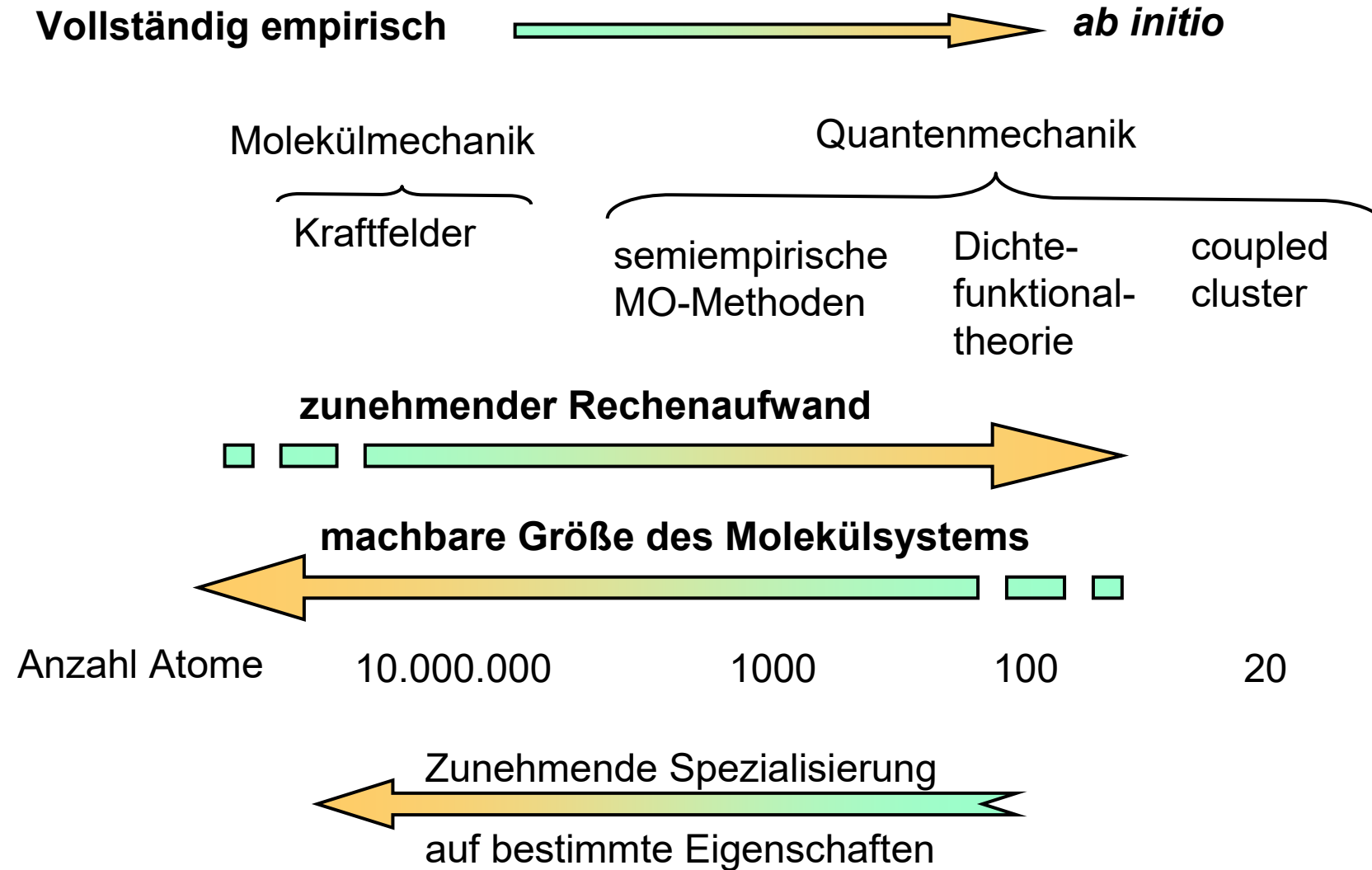


Diese Preise wurden in der Wissenschaftsgemeinde ("community") mit ungeheurer Befriedigung aufgenommen, nicht allein als Auszeichnung der beiden Forscher,

sondern als Auszeichnung des gesamten Gebiets.

Weiterhin: Nobelpreis für Chemie 2013 an **Martin Karplus**, **Michael Levitt** und **Arieh Warshel** "für die Entwicklung von multiskalen Modellen für komplexe chemische Systeme"

# Häufig verwendete Methoden



# Welche Methoden verwendet Computational Chemistry?

- Molekül-Mechanik (empirische Kraftfelder AMBER, OPLS, CHARMM, GROMOS, ...)
- Moleküldynamik (klassische Newton-Mechanik GROMACS,...)
- Semi-empirische Molekül-Orbital-Theorie (MNDO, AM1, PM3, OM2, MNDO/d, ...)
- Dichtefunktionaltheorie (B3LYP, PW91, BP86, LDA...)
- *ab Initio* Molekül-Orbital-Theorie (Hartree-Fock, Møller-Plesset, Coupled Cluster ...)

zur Computational Chemistry gehören ebenfalls:

- *Quantitative Structure-Activity Relationships* (QSAR)
- Chemoinformatik/*Cheminformatics*
- Docking (AutoDock, FlexX, DOCK, GLIDE, GOLD,...)
- Graphische Darstellung von Strukturen und Eigenschaften (VMD, Pymol,...)



# Wozu brauchen Bioinformatiker Computational Chemistry?

- Protein-Liganden Bindung
- Protein-Protein Bindung
- Proteinfaltung
- Docking (Konformationsanalyse)
- QSAR
- ...



<http://www.dell.com>  
Univ. Buffalo cluster



<http://www.aventis.com/>

→ Entwicklung von Medikamenten

# Überblick über den Inhalt der Vorlesung

## Molekül-Mechanik

- 1 Einleitung
- 2 Strukturen, molekulare Kräfte
- 3 Kraftfelder und Minimierung
- 4 Moleküldynamik-Simulationen
- 5 Statistische Mechanik und Sampling
- 6 Proteinstrukturen
- 7 Quantenchemische Grundlagen und Molekülorbital Theorie

## Quantenchemie

- 8 Quantenchemische Berechnungen
  - 9 Solvation, Docking
  - 10 Bindungsaffinitäten Enzym-Ligand und Scoringfunktionen
  - 11 Chemische Reaktionen
  - 12 Moleküleigenschaften
- Klausur (voraussichtlich) in der letzten Vorlesungswoche. Umfrage zu passendem Termin folgt...

# Schein

Es wird jede Woche in der Vorlesung 1 Übungsblatt ausgegeben, also insgesamt etwa 8-9 Übungsblätter/Programieraufgaben.

Die Aufgaben werden im Tutorium (Montags 11:15 – 12:45) in der darauf folgenden Woche besprochen.

In diesem Semester ist die Zulassung zur **Abschlussklausur** nicht an das Erreichen einer Mindestpunktzahl gekoppelt.

Einen **Übungsschein** über die erfolgreiche Teilnahme an der Vorlesung (6 LP) gibt es bei erfolgreicher Teilnahme an der **Abschlussklausur** und/oder der Nachklausur.

Die Note des Übungsscheins entspricht der besseren Note aus beiden Klausuren.

**Sprechstunde:** nach Vereinbarung (oder per e-mail)

# Literatur - Quantenchemie

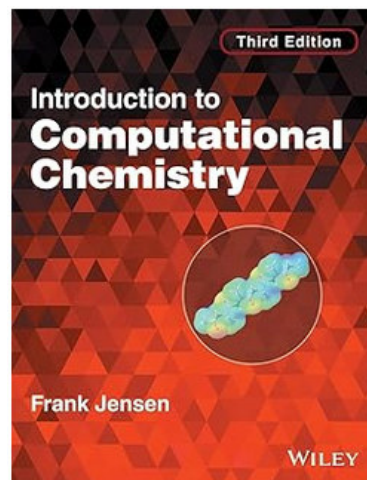
Vorlesungsfolien und Übungsaufgaben finden Sie auf unserer Webseite

<https://www-cbi.cs.uni-saarland.de/teaching/>

Introduction to Computational Chemistry

Frank Jensen, Wiley, ca. €87

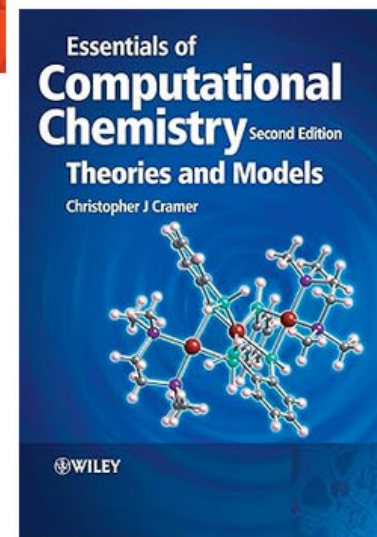
(ältere Ausgabe im Semesterapparat)



Essentials of Computational Chemistry

Christopher J. Cramer, Wiley, ca. €56

(ältere Ausgabe im Semesterapparat)

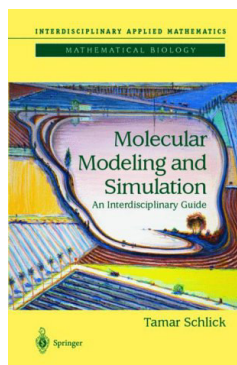


# Literatur – Molekülmechanik/Simulationen

Molecular Modeling and Simulation

Tamar Schlick, Springer, ca. €70

(in Info-Bibliothek, Semesterapparat)



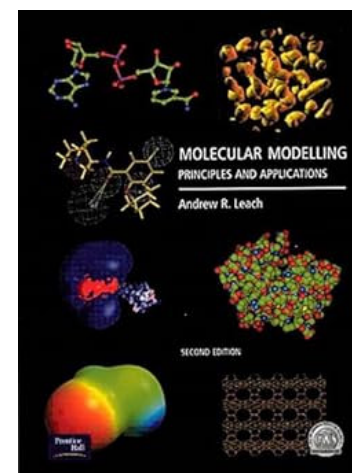
Molecular Modelling. Principles and Applications

2nd ed 2001, Andrew R. Leach,

Prentice Hall, (vergriffen)

(in Semesterapparat

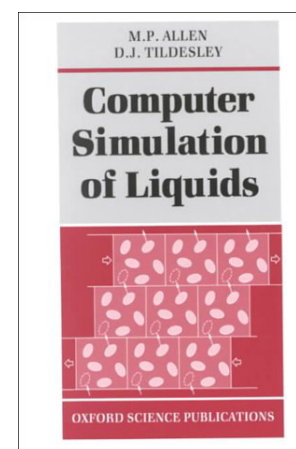
und Lehrbuchsammlung)



Computer Simulation of Liquids

M.P. Allen & D.J Tildesley, Oxford Science, ca €66

(im Semesterapparat)

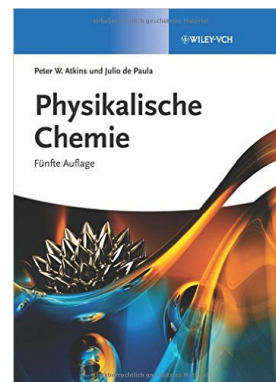


# Literatur – physikalische Chemie

Physikalische Chemie

Peter W. Atkins, Wiley-VCH, ca. €86

(siehe SULB)



Kurzlehrbuch der physikalischen Chemie

Peter W. Atkins, Wiley-VCH, ca. €60

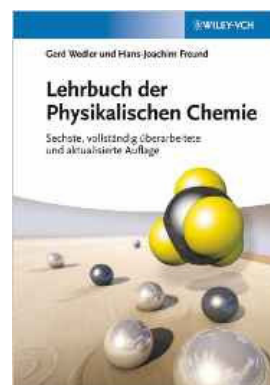
(siehe SULB)



Lehrbuch der physikalischen Chemie

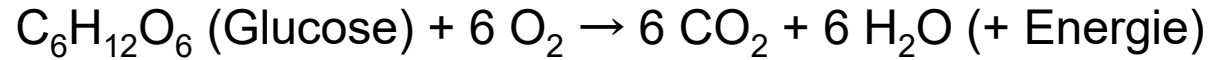
Gerd Wedler, Wiley-VCH, ca. €90

(siehe SULB)

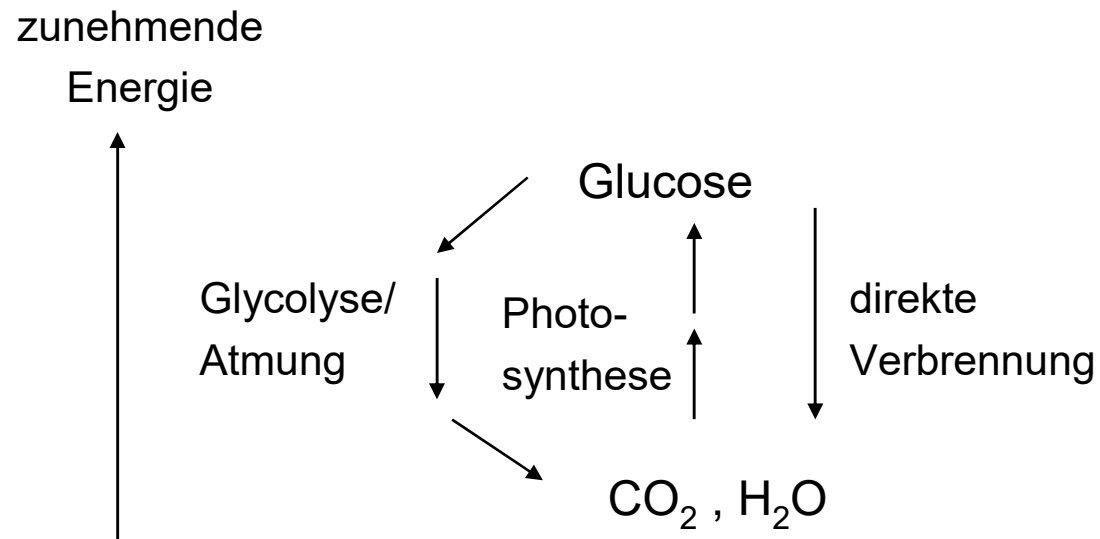


Zurückliegende Auflagen sind ebenfalls geeignet.

# Stöchiometrie und Energieerhaltung

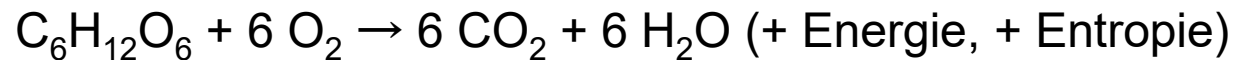
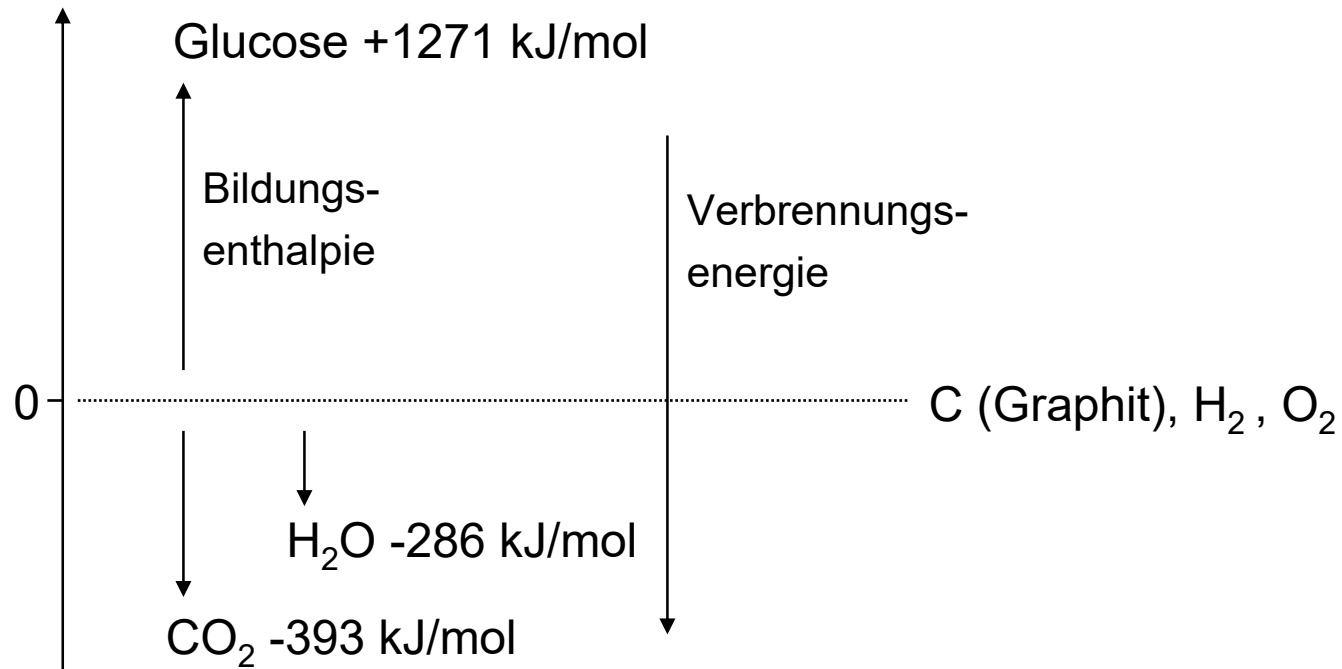


Gleiche Anzahl von Atomen (6\*C, 12\*H, 18\*O) auf beiden Seiten der Reaktionsgleichung.



# Verbrennungsenergie und Bildungsenthalpie

zunehmende  
Energie



$$\Delta G = \Delta H - T \cdot \Delta S \quad \text{Freie Enthalpie} = \text{Enthalpie} - \text{Temperatur} \cdot \text{Entropie}$$



# Stoffmenge und Konzentration

Um in Reaktionsgleichungen und anderen Situationen in der physikalischen Chemie die Menge an Atomen, Elektronen oder Molekülen quantifizieren zu können, bedarf es der Angabe der Stoffmenge:

$$1 \text{ mol} = 6.022 \cdot 10^{23} \text{ Teilchen} \quad (N_A = \text{Avogadrozahl})$$

Das vereinfacht auch Angaben zur Masse und Konzentration:

$$1 \text{ Kohlenstoffatom wiegt } 12.011 \text{ u} = 7.233 \cdot 10^{-26} \text{ kg}$$

$$1 \text{ mol Kohlenstoff} = 6.022 \cdot 10^{23} \cdot 12.011 \cdot 6.022 \cdot 10^{-27} \text{ kg} = 12.011 \text{ g}$$

Die Molmasse stimmt fast immer mit der Ordnungszahl des entsprechenden Elements überein (unterschiedlich schwere Isotope).

Ebenso wird die Stoffkonzentration in mol/l (Mol pro Liter) angegeben.

# Der Erste Hauptsatz der Thermodynamik

„Dem System ist egal, in welcher Form Energie übertragen wird.“

Es funktioniert ähnlich wie ein Bankkonto in Bezug auf Geld.

Alle Energiebeiträge die hineingesteckt werden sind positiv, alle Beiträge die das System leistet, negativ.

**Erster Hauptsatz:** verändert sich ein System von einem Zustand in einen anderen auf einem beliebigen adiabatischen Weg, so ist die geleistete Arbeit  $w_{ad}$  immer dieselbe, unabhängig von der angewandten Methode.

Der Wert von  $w_{ad}$  ist für alle Wege gleich und hängt nur vom Anfangs- und Endzustand ab.

$$w_{ad} = U_E - U_A$$

$U$  ist die **innere Energie** des Systems.

$U$  ist eine **Zustandsfunktion**, da sie den Zustand des Systems beschreibt.



# Energiebegriff (I)

Was ist Energie, was ist Arbeit?

Mechanik: Kraft • Weg

**Definition: Energie** ist die **Fähigkeit, Arbeit zu leisten.**

Energie und Arbeit sind äquivalent.

Enthalpie  $H = \text{Innere Energie } U + \text{Volumenarbeit } p \cdot V$

z.B. in Fällen bei denen sich das Volumen ändert, so wie bei der Verbrennung.



Formen von Energie:

Lageenergie (mechanisch)

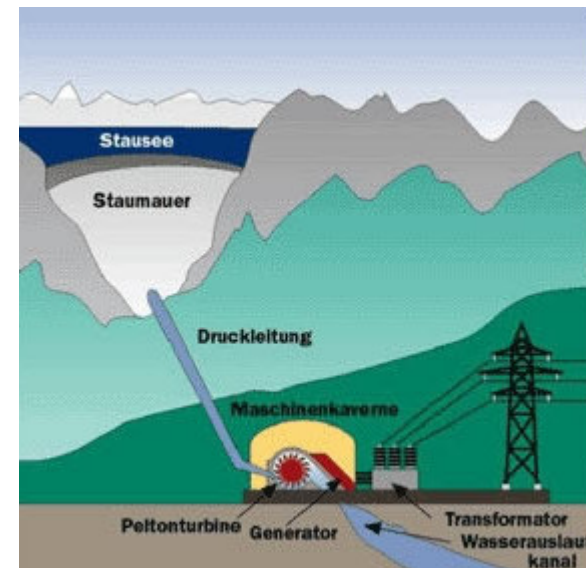
Elektrische Energie

Volumenarbeit

Chemische Energie



**Welche Energieformen kennen Sie noch?**



## Energiebegriff (II)

**System:** derjenige Teil der Welt, dem unser spezielles Interesse gilt.

Außerhalb des Systems befindet sich die **Umgebung**.

**Offene Systeme** erlauben den Austausch von Materie bzw. Wärme mit ihrer Umgebung.

**Abgeschlossene Systeme** haben mit der Umgebung weder mechanischen bzw. thermischen Kontakt, denn auch durch Wärmeaustausch kann Energie übertragen werden. Deshalb spricht man auch von *adiabatischen* Systemen.

Wenn wir an einem ansonsten isolierten System Arbeit leisten, nimmt seine Fähigkeit, selbst Arbeit zu leisten, zu, d.h. seine Energie nimmt zu. Wenn das System Arbeit leistet, so nimmt seine Energie ab.



# Newton'sche Gesetze

2. Gesetz

Die Beschleunigung  $a$  ist dem Verhältnis von Kraft  $F$  und Masse  $m$  proportional:

$$\begin{aligned} F &= m \cdot a \\ &= m \cdot \ddot{x} \\ &= m \frac{d^2 x}{dt^2} \end{aligned}$$

3. Gesetz:

Actio = Reactio



$$F_g = m \cdot g$$

## Energie - Kraft

$T$  : **kinetische Energie** eines Teilchens.

$$T(\dot{x}, \dot{y}, \dot{z}) = \frac{1}{2} m v^2$$

in kartesischen Koordinaten:

$$= \frac{1}{2} m (\dot{x}^2 + \dot{y}^2 + \dot{z}^2)$$

$U$  : **potentielle Energie** des Teilchens.

Meist hängt  $U$  nur von den Positionen ab, also  $U(x,y,z)$ .

Die **Newtonschen Bewegungsgleichungen** lauten damit

$$m\ddot{x}_i = - \left. \frac{\partial U}{\partial x_i} \right|_{x_i=x,y,z}$$

Für die Kraft  $F$  gilt:

$$F = - \frac{\partial U}{\partial x}$$

in einer Dimension

$$F = -\nabla U$$

in drei Dimensionen

mit dem Gradienten-

(Nabla-)Operator

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}$$

# Das mikrokanonische *NVE*-Ensemble

Gegeben: ein System mit Teilchenzahl  $N$  und Volumen  $V$ .

In einem idealisierten, von der Außenwelt abgeschlossenen, System ist die Gesamtenergie  $E$  konstant = **mikrokanonisches Ensemble**

Bsp.: harmonischer Oszillator, Schwingungsbewegung in einem harmonischen Potential

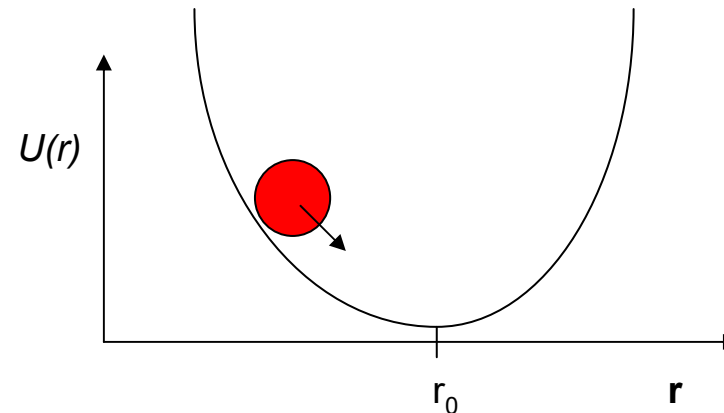
potentielle Energie  $U$ :

$D$ : z.B. Federkonstante

$$U(r) = \frac{1}{2} D (r - r_0)^2$$

Wenn Gesamtenergie  $E_0$  gegeben,

kinetische Energie = Gesamtenergie – potentielle Energie



# Was kann man mit Computational Chemistry berechnen?

(1) Was ist die energetisch beste Konformation eines Moleküls?

Ligand für sich, in der Bindungstasche eines Proteins  
Bindungsaffinität mittels energiebasierter Scoringfunktionen  
Proteinfaltung

(2) Was sind bei Raumtemperatur erreichbare andere Konformationen?

Konformationssampling des Moleküls (Boltzmann Statistik).

(3) Dynamik von Konformationsübergängen? Einfluss des Lösungsmittels?

Moleküldynamik-Simulationen  
Solvatationseffekte

(4) Was ist die energetisch stabilste Anordnung einer gegebenen Anzahl von Atomen und Elektronen (welches Strukturisomer ist am stabilsten)?

Welche Reaktionsprodukte sind denkbar?

Erfordert quantenmechanische Behandlung



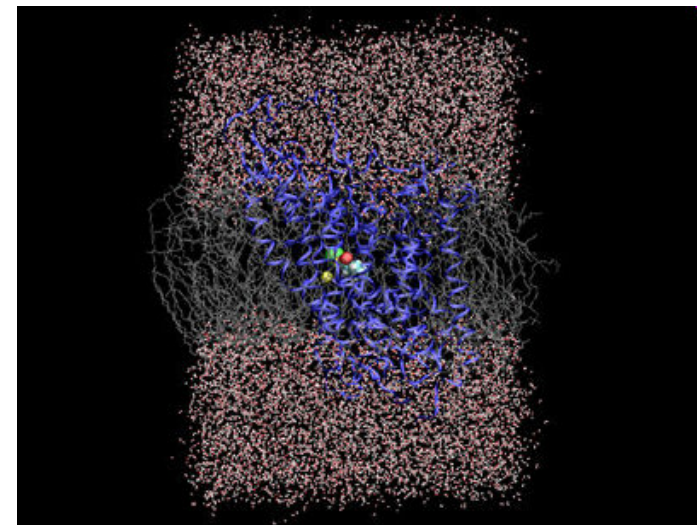
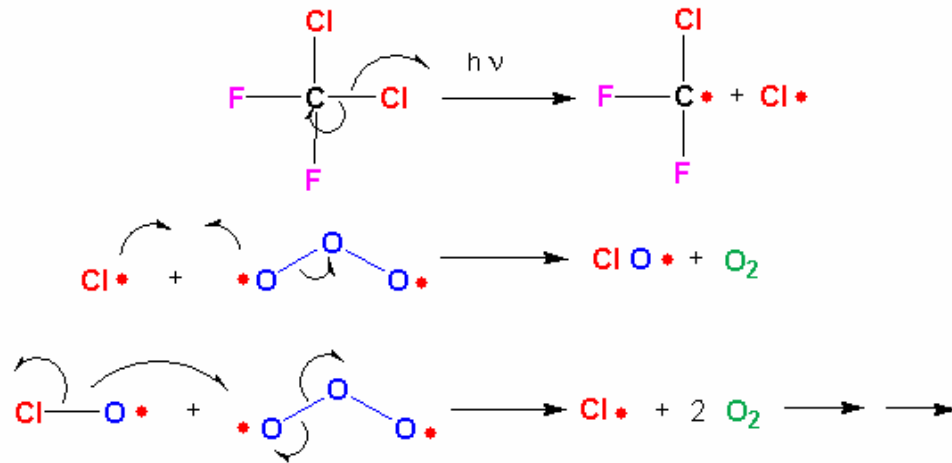
# Wie genau ist Computational Chemistry?

Durch Verwendung hochexakter Theorien wie der coupled-cluster-Methode können für kleine Moleküle (bis etwa 15 Atome) im Vakuum Eigenschaften genauer als im Experiment berechnet werden! Bsp: Ozonabbau durch FCKs

Bei Unstimmigkeiten müssen mittlerweile oft die experimentellen Daten korrigiert werden (z.B. Verbrennungsenergien)!

Für große Moleküle (Proteine) sind diese Verfahren jedoch nicht praktikabel.

Hier braucht man vereinfachte Verfahren wie Molekülmechanik (siehe Vorlesung 3).



## Anwendungsbeispiel

Anwendung z.B.: Berechnung von Reaktivitäten und Lösungseigenschaften von Aktiniden mittels relativistischer Quantenchemie am Pacific Northwest National Laboratory.

*„There are 177 underground waste storage tanks at Hanford. The tanks contain wastes collected over almost 50 years of plutonium production. The wastes include radioactive isotopes, toxic chemicals, corrosive liquids, organic solvents, and other dangerous and hazardous substances.“* <http://www.pnl.gov/tws/>

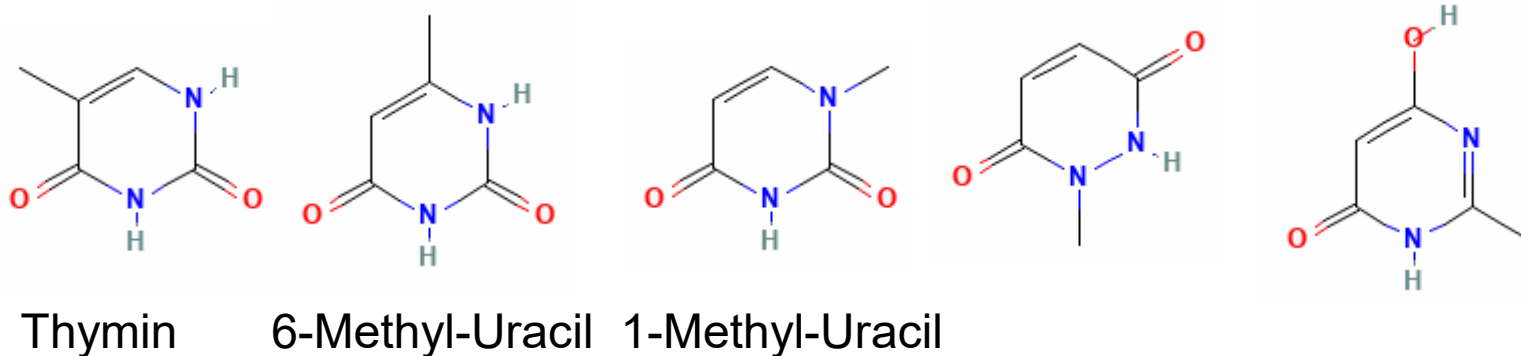
Problem hier: es fehlen experimentelle Daten,  
beispielsweise für die Löslichkeiten von  
Uran-Verbindungen wie  $\text{UF}_6$

Quantenchemische Rechenmethoden benötigen keine  
experimentellen Daten!



## Anwendungsbeispiel (2)

In Proben vom Asteroid (101955) Bennu wurden neben anderen organischen Verbindungen die Nukleinbasen Adenine, Guanin, Thymin, Cytosin und Uracil gefunden. Bilden diese sich bevorzugt, z.B. weil diese energetisch stabiler als ihre Isomere mit derselben Summenformel sind?



Die Suche in der Pubchem Datenbank liefert 1116 Isomere mit derselben Summenformel  $C_5H_6N_2O_2$ .

Mittels quantenchemischer Rechnungen lässt sich herausfinden, welches Isomer die günstigste Energie hat, ohne Experimente im Labor durchzuführen.

Bildquelle: <https://pubchem.ncbi.nlm.nih.gov>

# Zusammenfassung

- Computerchemie besitzt eine lange Geschichte.
- Bedeutung der Computerchemie wuchs stets parallel zur Entwicklung der Rechenleistung.
- Zwei wesentliche “Welten”: Quantenchemie  $\leftrightarrow$  Molekülmechanik
- Quantenchemie für sehr kleine Moleküle ist heutzutage hoch exakt, oft genauer als das Experiment
- bei großen Systemen (z.B. Proteinen) müssen jedoch starke Näherungen gemacht werden (z.B. Kraftfelder)

Das wesentliche **Lernziel** dieser Vorlesung ist zu **verstehen**, was die verschiedenen Methode leisten können und wo die Probleme liegen. Für die Praxis: Welche Methode ist für welchen Zweck geeignet, bzw. welche ist völlig ungeeignet.

## Arbeiten mit Strukturdaten (I)

Um Atomkoordinaten, Bindungen zwischen den Atomen und weitere Informationen zu speichern wurden im Lauf der Zeit zahlreiche Dateiformate entwickelt, die alle ihren jeweiligen Zweck erfüllen und den Stand der Technik reflektieren.



Beispiel einer Lochkarte mit 80 Spalten

## Arbeiten mit Strukturdaten (II)

Röntgenkristallstrukturdaten von Proteinen und DNA/RNA werden seit 1968 in der **Protein Data Bank** gespeichert. Seit 1971 ist die PDB sozusagen “open source”.

Strukturdaten die mittels Kryoelektronenmikroskopie oder NMR-Spektroskopie gewonnen wurden sind ebenfalls erfasst.



Webseiten:

<https://www.rcsb.org/>      <https://www.wwpdb.org/>

<https://www.ebi.ac.uk/pdbe/>

# Das pdb Dateiformat (I)

Das Format von .pdb Dateien reflektiert die Formatierungskonventionen der Programmiersprache FORTRAN, was die Aufteilung der Spalten betrifft, sowie die Limitierungen durch Lochkarten, z.B. maximal 80 Spalten.

Reihenfolge in einer pdb Datei:

Title (Header, Title, Author, Remark,...) u.a. auch die Auflösung

Primary Structure (Ddbref, Seqres,...) Aminosäuresequenz

Heterogen (Het, Hetnam, Formul,...) Andere Atome z.B. Ionen, Solvent

Secondary Structure (Helix, Sheet) Sekundärstrukturabschnitte

Connectivity Annotation (Ssbond, Link) Disulfidbrücken, zusätzliche Bindungen

Crystallographic & Transformation (Cryst, Origx, Scale; Mtrix) Elementarzelle

Coordinates (Model, Atom, Hetatm, Anisou, Endmdl, Ter) Atomkoordinaten

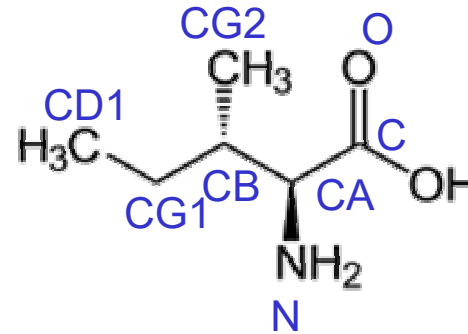
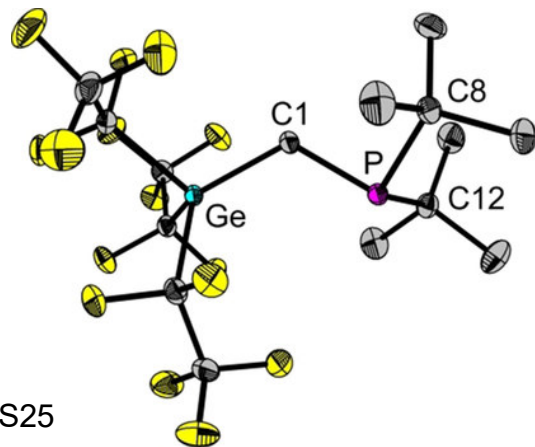
Connectivity (Conect) Bindungen zu und zwischen Heteroatomen

# Das pdb Dateiformat (II)

Anhand der Schlüsselworte in den Spalten 1-6 ergibt sich die Formatierung der restlichen Spalten in der jeweiligen Zeile:

1	2	3	4	5	5	7	8				
12345678901234567890123456789012345678901234567890123456789012345678901234567890											
ATOM	1	N	ILE A	1	-2.693	51.036	97.817	1.00	1.12		N
ATOM	2	CA	ILE A	1	-2.754	49.912	96.871	1.00	0.13		C
ATOM	3	C	ILE A	1	-3.331	48.677	97.571	1.00	0.79		C
ATOM	4	O	ILE A	1	-2.641	47.979	98.318	1.00	0.39		O
ATOM	5	CB	ILE A	1	-1.350	49.614	96.339	1.00	0.11		C
ATOM	6	CG1	ILE A	1	-0.729	50.840	95.663	1.00	0.35		C
ATOM	7	CG2	ILE A	1	-1.335	48.482	95.309	1.00	0.69		C
ATOM	8	CD1	ILE A	1	0.767	50.706	95.394	1.00	0.32		C
ATOM	9	N	GLY A	2	-4.607	48.430	97.315	1.00	0.29		N

chain resid x y z occupancy  
Temperaturfaktor





## Das pdb Dateiformat (III)

Die Abtrennung der Einträge innerhalb einer Spalte benötigt **keine** Leerzeichen!

1 – 6	s6	ATOM oder HETAM
7 – 11	i5	Atom serial number
13 – 16	i4	Atom name (CA, CB, 1HG, ZN, ...)
17	s1	alternative location index (A, B, C) fall occupancy <1.00
18 – 20	s3	residue name (ALA, CYS, HOH, WAT, ... 3ST, B4G)
21	s1	not used
22	s1	chain ID (A, B, C, ...)
23 – 26	i4	residue number (kann auch negativ sein)
27	s1	code for insertion of amino acids
31 – 38	f8.3	x
39 – 46	f8.3	y
47 – 54	f8.3	z
55 – 60	f6.2	occupancy (falls <1.00 sind Einträge in Spalte 17 vorhanden)
61 – 66	f6.2	Temperaturfaktor (üblicherweise der isotropische B-Faktor)
67 – 76		not used
77 – 78	s2	Element, rechtsaligniert (z.B. Zn, Mg, Fe)
79 – 80	i2	gegebenenfalls Ladung auf dem Atom

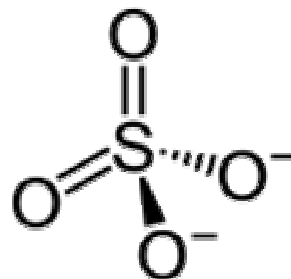
## Das pdb Dateiformat (IV)

Nach den eigentlichen Proteinatomen folgen die Heteroatome.

```
      1      2      3      4      5      6      7      8
1234567890123456789012345678901234567890123456789012345678901234567890
HETATM 8237 MG      MG A1001      13.872 -2.555 -29.045  1.00 27.36      MG
HETATM 3835 FE      HEM A  1      17.140  3.115  15.066  1.00 14.14      FE
HETATM 8238 S      SO4 A2001      10.885 -15.746 -14.404  1.00 47.84      S
HETATM 8239 O1     SO4 A2001      11.191 -14.833 -15.531  1.00 50.12      O
HETATM 8240 O2     SO4 A2001      9.576 -16.338 -14.706  1.00 48.55      O
HETATM 8241 O3     SO4 A2001      11.995 -16.703 -14.431  1.00 49.88      O
HETATM 8242 O4     SO4 A2001      10.932 -15.073 -13.100  1.00 49.91      O
```

Im Gegensatz zu den Aminosäuren bei denen die Bindungen durch die jeweilige Aminosäure definiert ist, werden die kovalenten Bindungen zu und zwischen den Heteroatomen explizit angegeben, und zwar redundant (doppelt):

```
      1      2      3      4      5      6      7      8
1234567890123456789012345678901234567890123456789012345678901234567890
CONNECT 8238 8239 8240 8241 8242
CONNECT 8239 8238
CONNECT 8240 8238
CONNECT 8241 8238
CONNECT 8242 8238
```



Maximal 5 Atome pro Zeile

Bildquelle: wikipedia.org

## Arbeiten mit Strukturdaten (III)

Typische Probleme bei der Verwendung von pdb Dateien:

- fehlende Atome und Residuen aufgrund der experimentellen Datenlage  
Rekonstruktion von Seitenketten und Loops
  - Alternative Konformationen/Locations von Seitenketten  
welches ist die bevorzugte Orientierung? (Rotamerbibliotheken)
  - Nichtideale Bindungsabstände und -Winkel, zu enge Kontakte zwischen nicht-verbundenen Atomen (steric clashes) (Energiminimierung)
  - Hinzufügen von Wasserstoffatomen (auch zu Wassermolekülen und Liganden). Protonierungszustände von Histidin, usw.
- Voraussetzungen für Protein-Ligand Docking

Teil der Übungsblätter sind auch Programmieraufgaben um Informationen aus pdb Dateien auszuwerten und bestehende Lösungsansätze für die oben genannten Probleme zu vergleichen.