

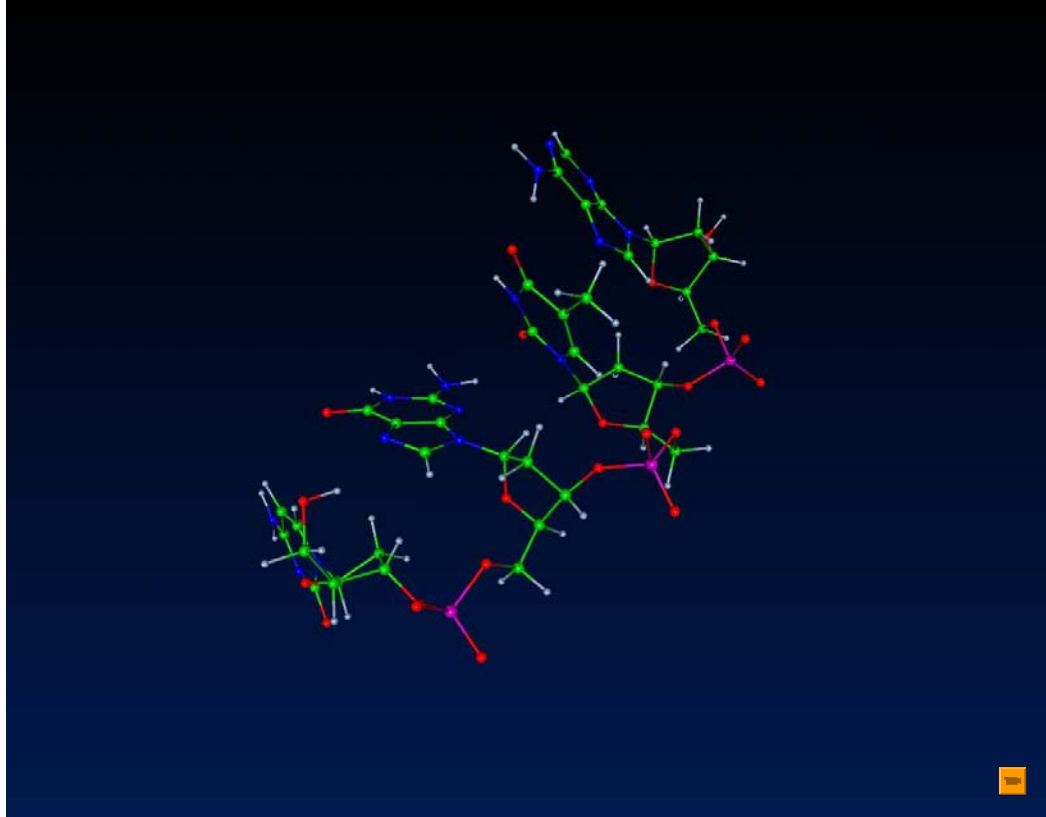
Moleküldynamik und Docking

Wilhelm Huisinga (FU)
und
Frank Cordes (ZIB)



Vorlesung Algorithmische Bioinformatik, WS 2002/03
27. Januar 2003

Das ist (klassische) Moleküldynamik





Fragen ?

Typische Fragestellungen

- **Faltung**: beliebige Ausgangsstruktur → Faltungszustand
(**Optimierungsprozess**)
- **Kinetik** des Faltungsprozesses (**Dynamik**)
- Bindungsaffinität beim **Docking** (Energiedifferenz)
Energieerwartungswerte (**Statistik**)
- **Konformationsdynamik**: Identifikation wesentlicher
3d-Strukturen, ihrer Lebensdauer sowie der
Übergangsraten (**Statistik & Dynamik**)
- ...

Wo kommen solche Fragestellungen vor?

- Unterstützung von High Throughput Screening (virtuelles Screening)
- künstlicher Medikamentenentwurf (drug design)
- Unterstützung experimenteller Strukturanalyse (simulated annealing)
- Bewertung von Reaktionswegen (Katalysatoren)
- Strukturvorhersage von Komplexen (docking)
- Strukturvorhersage von Biomolekülen (Homology Modelling / Threading)
- chemische Intuition

Was braucht man alles?

1. 3d **Struktur** des Moleküls
2. Mathematisch/physikalisches Modell für die **Dynamik**
3. Numerisches Verfahren zur **Simulation**

... und natürlich Algorithmen zur Auswertung der Daten.

3d Struktur des Moleküls

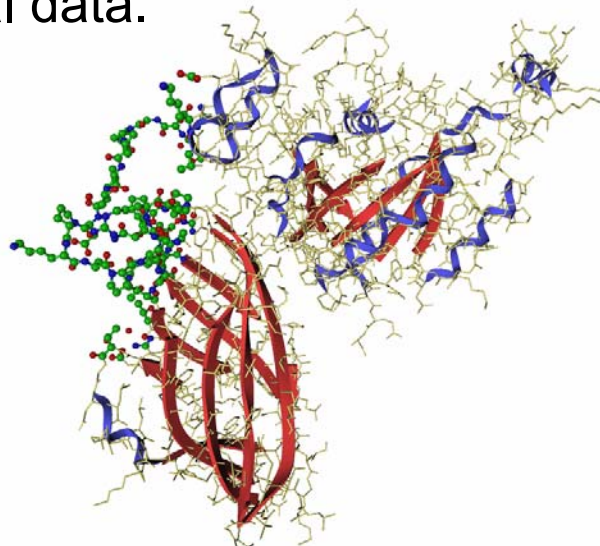
<http://www.rcsb.org/pdb/>

z.B. **PDB-Datenbank:**

„The Protein Data Bank (PDB) is an **archive of experimentally determined three-dimensional structures** of biological macromolecules.

The archives contain **atomic coordinates**, bibliographic citations, **primary** and **secondary structure** information, as well as crystallographic structure factors and NMR experimental data.“

Zitat von der homepage



Was braucht man alles?

1. 3d **Struktur** des Moleküls
2. Mathematisch/physikalisches Modell für die **Dynamik**
3. Numerisches Verfahren zur **Simulation**

Einige Modelle für die Dynamik

Hamilton-Dynamik:

$$\dot{q} = p$$

$$\dot{p} = -DV(q)$$

die Vorhersagekraft und die Anwendungsbereiche der verschiedenen Modelle werden durchaus kontrovers diskutiert.

Langevin-Dynamik:

$$\dot{q} = p,$$

$$\dot{p} = -DV(q) - \gamma p + \sigma \dot{W}$$

Markov-Ketten-Monte-Carlo:

$$P(x_k, x_{k+1}) = Q(x_k, x_{k+1}) \min\left(1, \frac{\pi(x_{k+1})}{\pi(x_k)}\right)$$

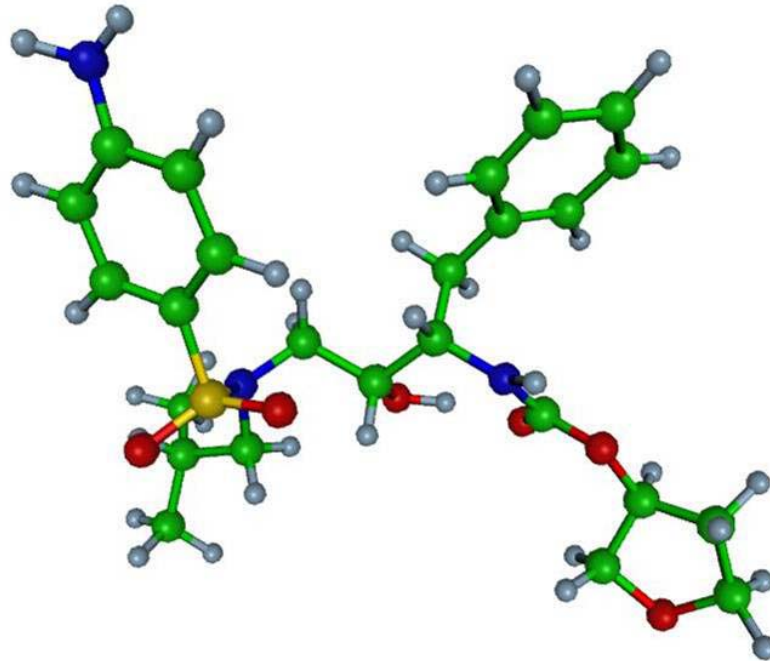
... und es gibt durchaus noch weitere Modelle.

Ein Modell etwas detaillierter

Hamilton-Dynamik:

$$\dot{q} = p$$

$$\dot{p} = -DV(q)$$



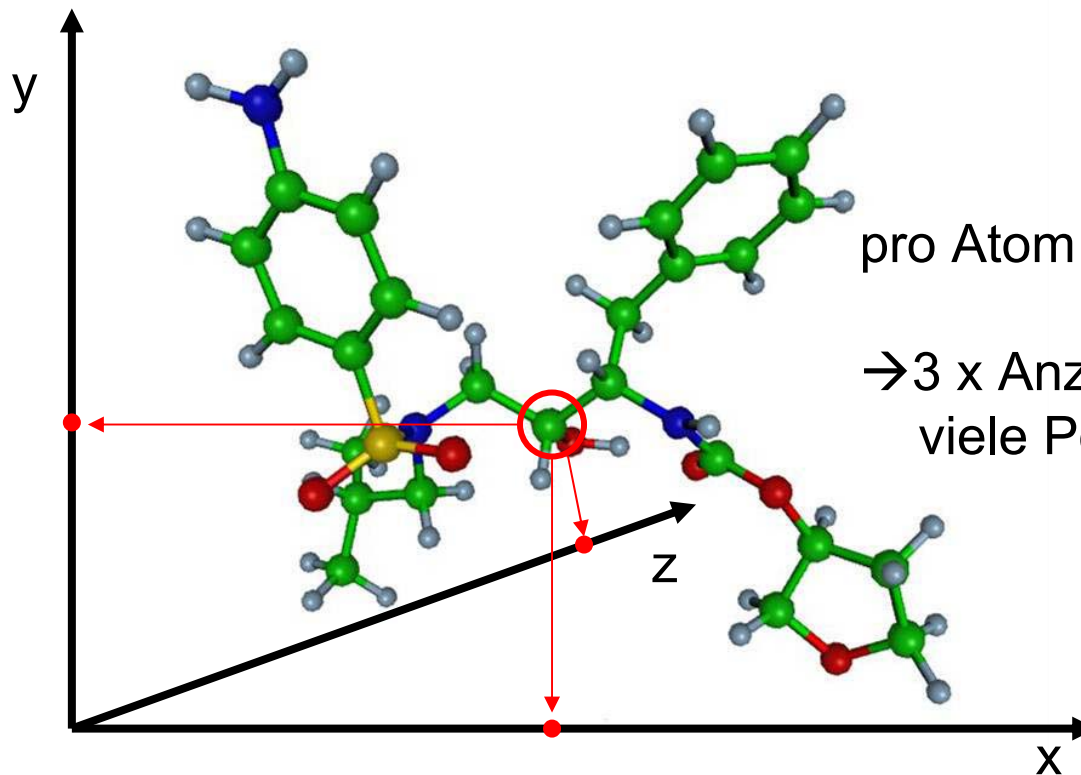
Ein Modell etwas detaillierter

Hamilton-Dynamik:

$$\dot{q} = p$$

$$\dot{p} = -DV(q)$$

Atompositionen (Orte)



pro Atom (x,y,z) Position

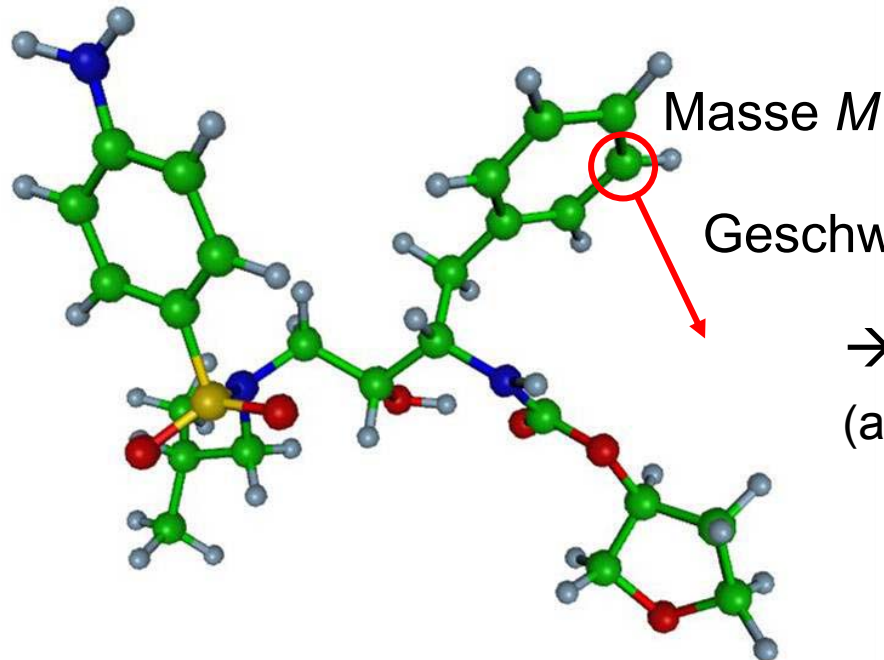
→ 3 x Anzahl Atome
viele Positionen

Ein Modell etwas detaillierter

Hamilton-Dynamik: Impulse (Geschwindigkeit, Masse)

$$\dot{q} = p$$

$$\dot{p} = -DV(q)$$



→ Impuls = $M v$

(auch hier:
3 x Anzahl Atome
viele Impulse)

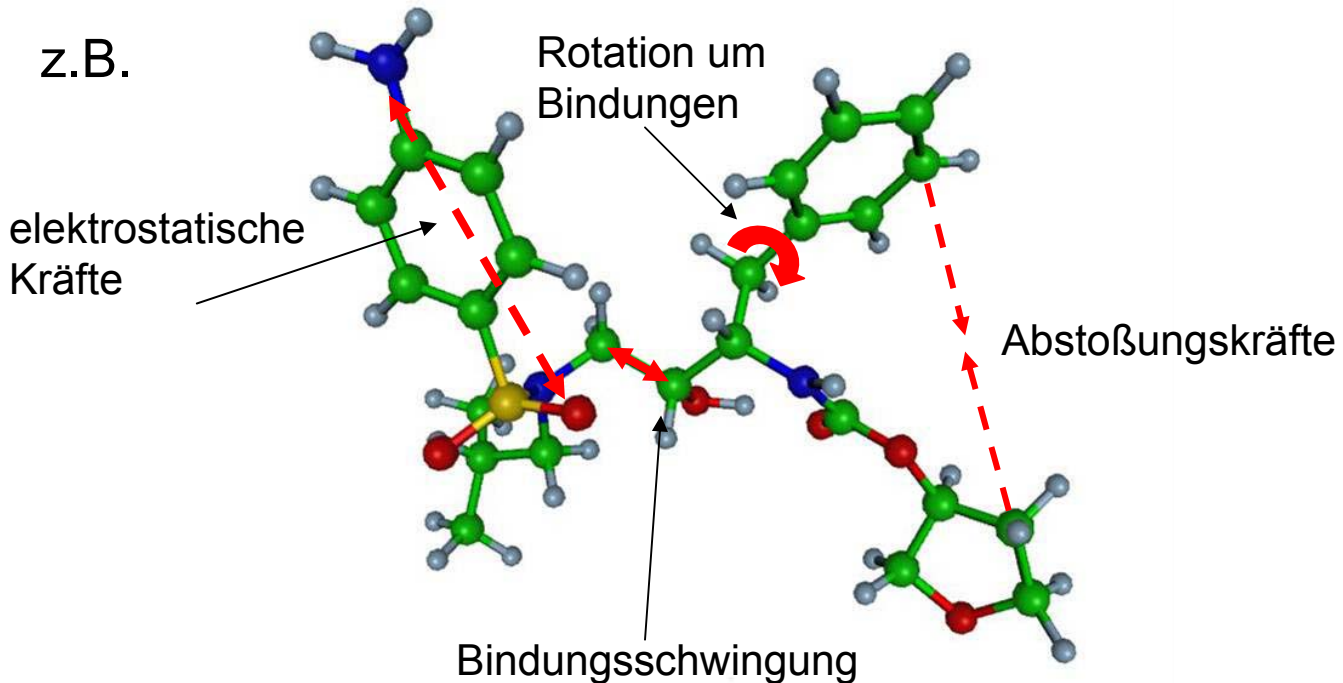
Ein Modell etwas detaillierter

Hamilton-Dynamik:

$$\dot{q} = p$$

$$\dot{p} = -D \boxed{V}(q)$$

Kraftfeld (Wechselwirkungspotential)



Ein Modell etwas detaillierter

Hamilton-Dynamik:

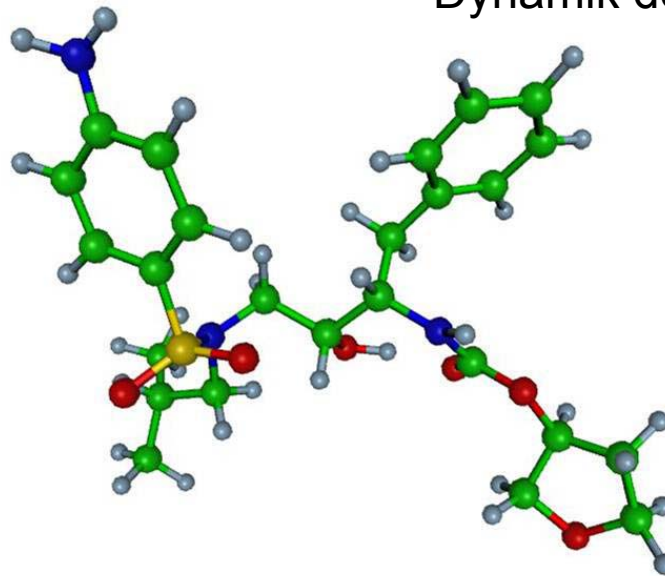
$$\dot{q} = p$$

$$\dot{p} = -DV(q)$$

→

$$q(t), p(t)$$

Lösung (Atompositionen und Impulse) in der Zeit t , also die Dynamik des Moleküls



Probleme:

- (1) Welche **Anfangswerte** wählt man?
- (2) Wie ist das **Kraftfeld** zu wählen?

Verbreitete Programmpakete und Kraftfelder

<http://msdlocal.ebi.ac.uk/docs/mmrefs.html>

Die Entwicklung brauchbarer Kraftfelder ist „eine Wissenschaft für sich“!

AMBER; Assisted Model Building and Energy Refinement)

CHARMM; Chemistry at HARvard Macromolecular Mechanics)

GROMOS (GROningen MOlecular Simulation package)

MMFF94, the **Merck** Molecular Force Field

Tripes, the force field of the Sybyl molecular modeling program

... um nur einige Programmpakete/Kraftfelder zu nennen.

Unterschiede: Modellierung von Wasserstoffbrückenbindungen, des Lösungsmittels, physikalische Terme, Anwendungsgebiete (Proteine, Nukleinsäuren, etc.)

Verbreitete Programmpakete und Kraftfelder

<http://msdlocal.ebi.ac.uk/docs/mmrefs.html>

GROMOS (GROningen MOlecular Simulation package)

→ **Software-Praktikum**

... um nur einige Programmpakete/Kraftfelder zu nennen.

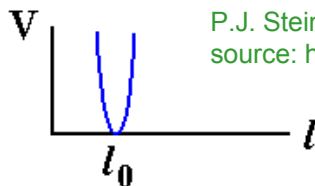
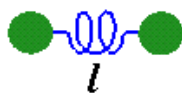
Unterschiede: Modellierung von Wasserstoffbrückenbindungen, des Lösungsmittels, physikalische Terme, Anwendungsgebiete (Proteine, Nukleinsäuren, etc.)



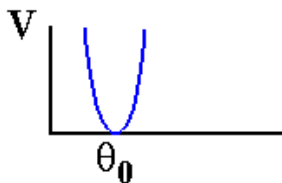
Bestandteile des Kraftfelds

P.J. Steinbach „Introduction to Macromolecular Simulation“
 source: http://cmm.info.nih.gov/intro_simulation/node15.html

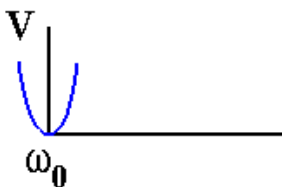
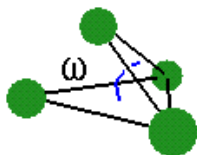
Bonds



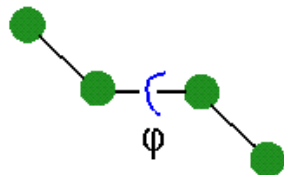
Angles



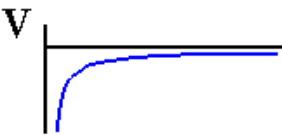
Improper
Dihedrals



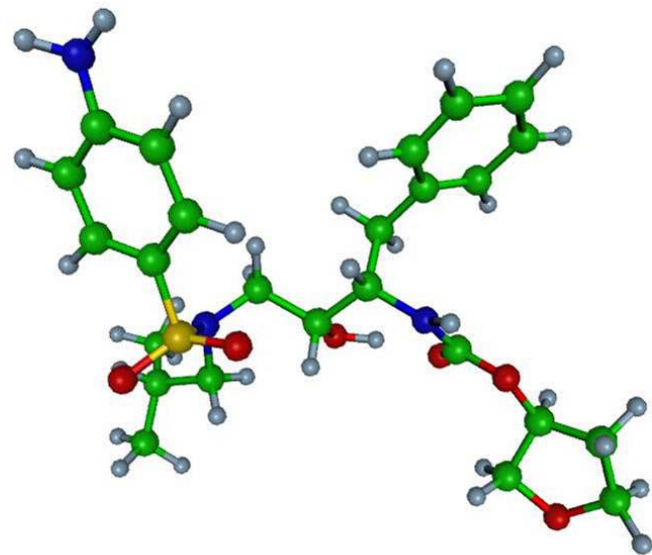
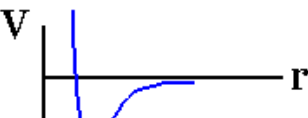
Torsions



Electrostatics



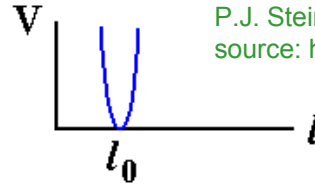
van der Waals



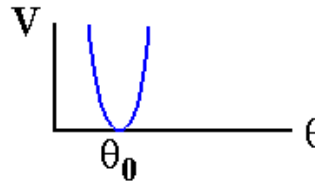
Bestandteile des Kraftfelds

P.J. Steinbach „Introduction to Macromolecular Simulation“
 source: http://cmm.info.nih.gov/intro_simulation/node15.html

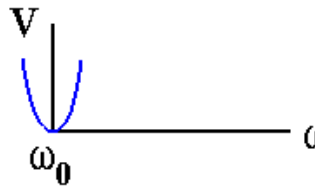
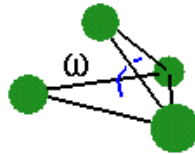
Bonds



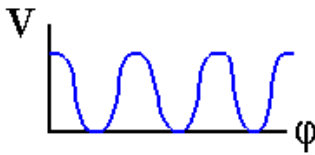
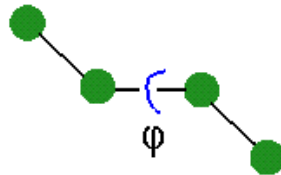
Angles



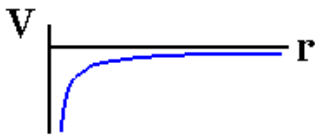
Improper Dihedrals



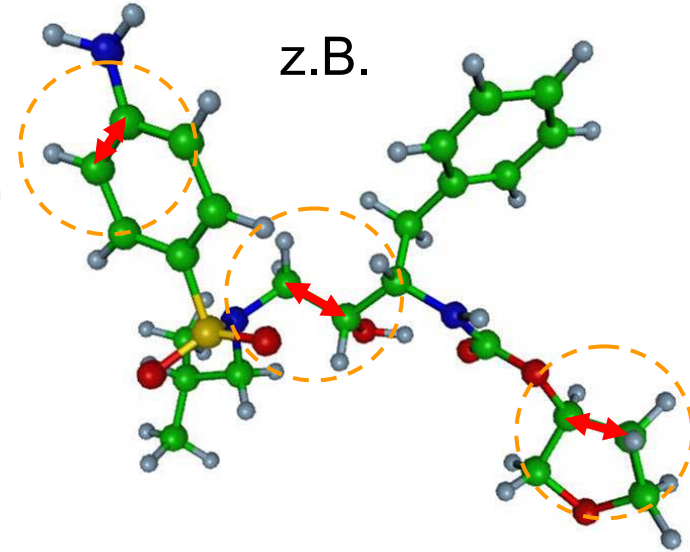
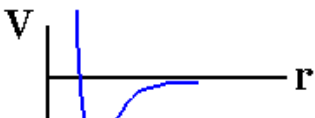
Torsions



Electrostatics



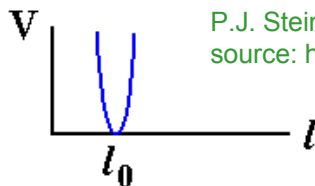
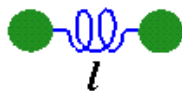
van der Waals



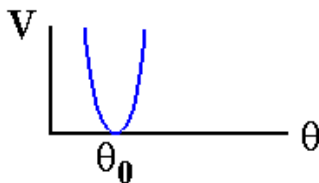
Bestandteile des Kraftfelds

P.J. Steinbach „Introduction to Macromolecular Simulation“
 source: http://cmm.info.nih.gov/intro_simulation/node15.html

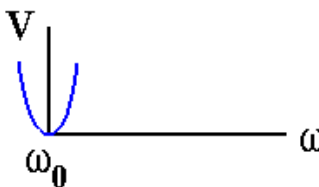
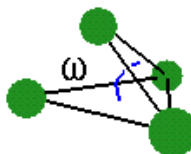
Bonds



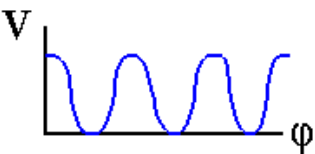
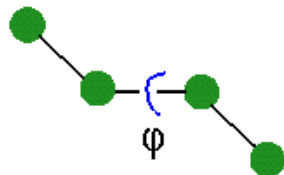
Angles



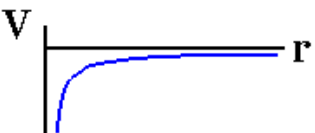
Improper
Dihedrals



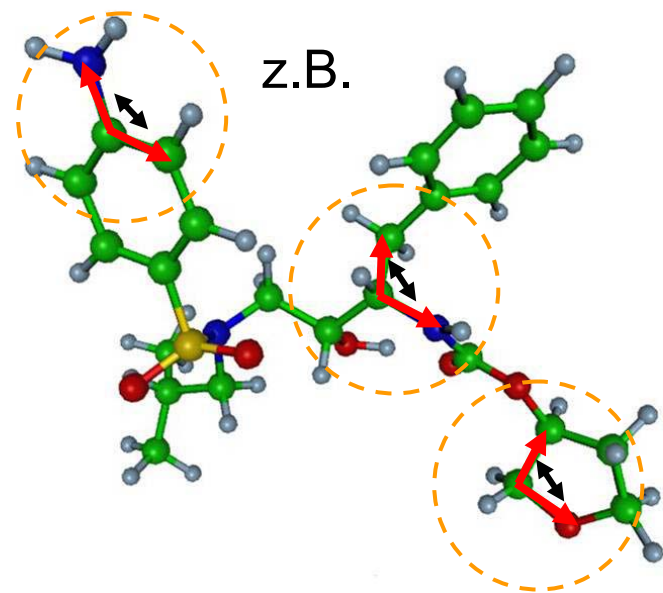
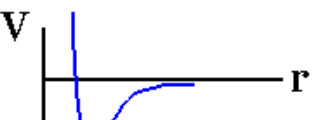
Torsions



Electrostatics



van der Waals

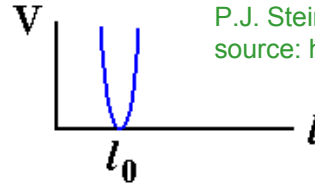




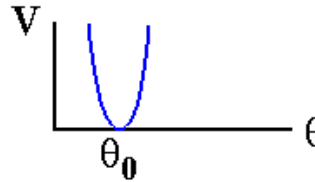
Bestandteile des Kraftfelds

P.J. Steinbach „Introduction to Macromolecular Simulation“
 source: http://cmm.info.nih.gov/intro_simulation/node15.html

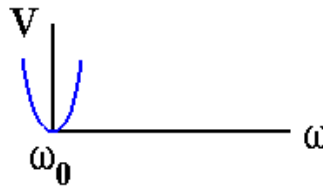
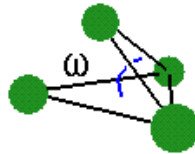
Bonds



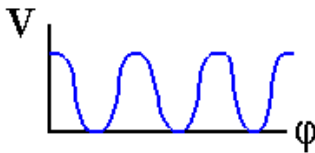
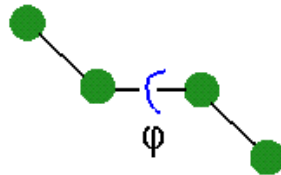
Angles



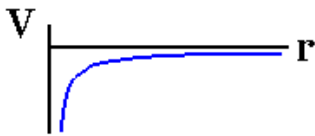
Improper
Dihedrals



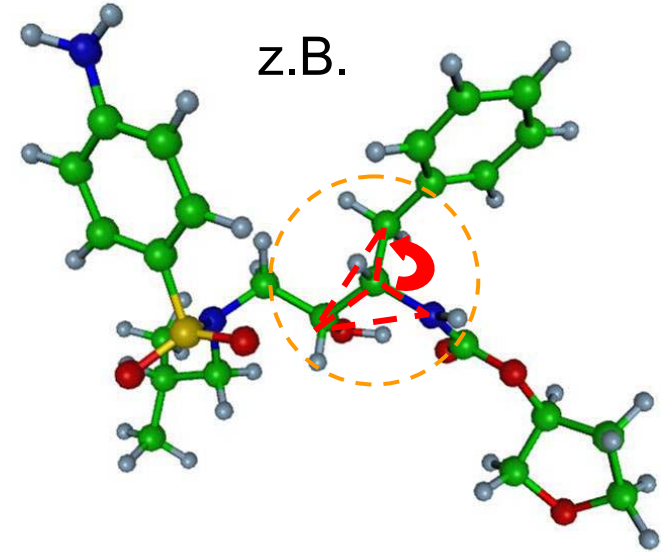
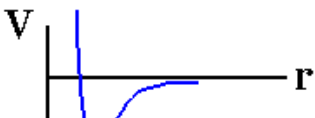
Torsions



Electrostatics



van der Waals

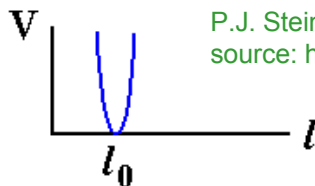
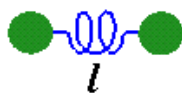




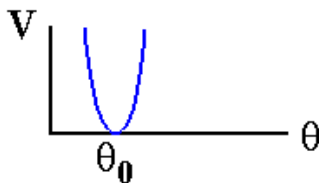
Bestandteile des Kraftfelds

P.J. Steinbach „Introduction to Macromolecular Simulation“
 source: http://cmm.info.nih.gov/intro_simulation/node15.html

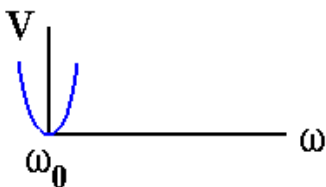
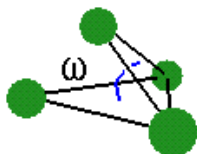
Bonds



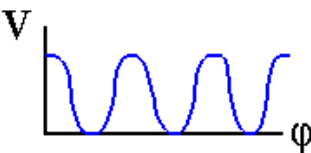
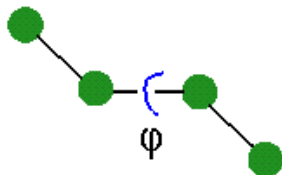
Angles



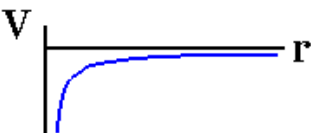
Improper Dihedrals



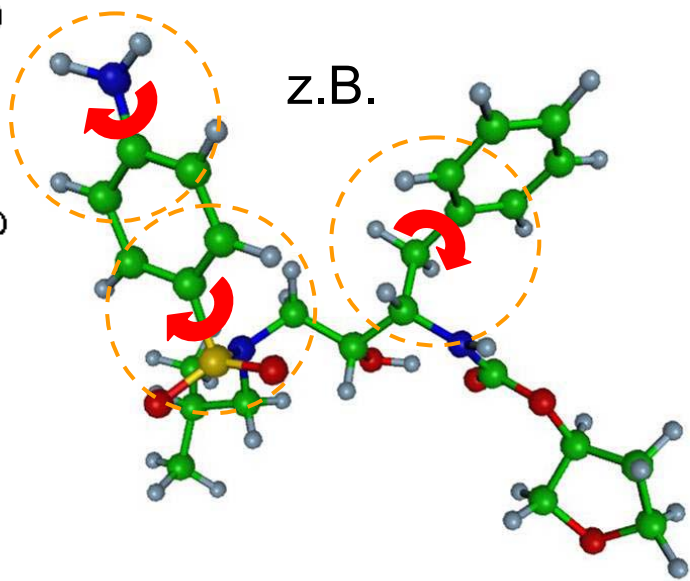
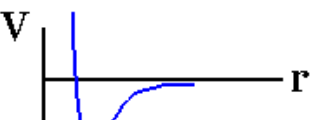
Torsions



Electrostatics



van der Waals

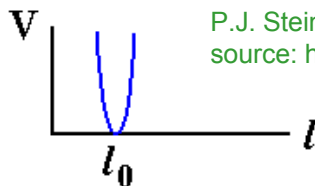
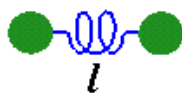




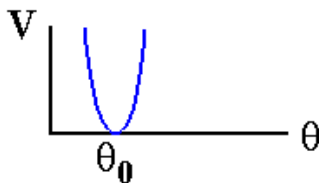
Bestandteile des Kraftfelds

P.J. Steinbach „Introduction to Macromolecular Simulation“
 source: http://cmm.info.nih.gov/intro_simulation/node15.html

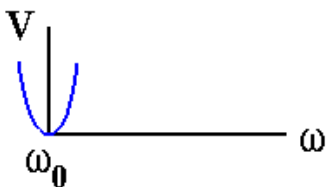
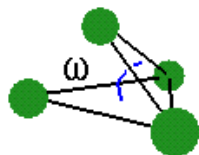
Bonds



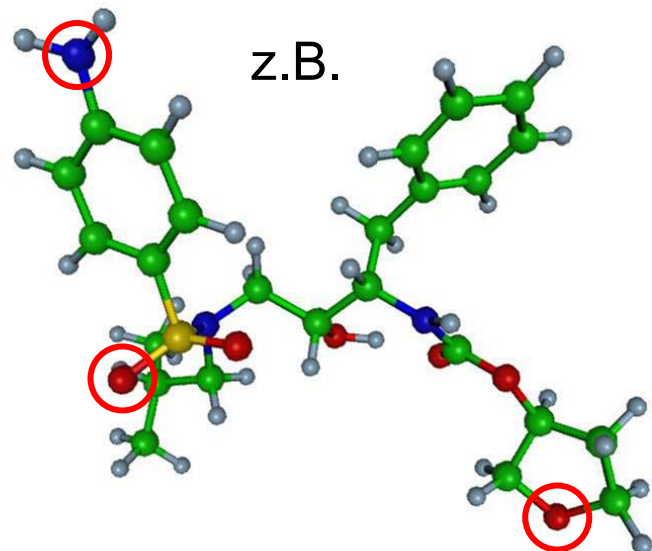
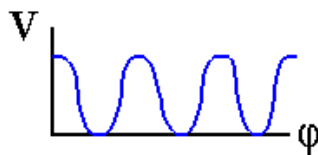
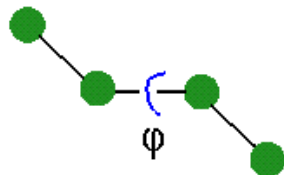
Angles



Improper
Dihedrals

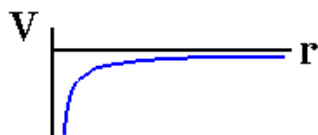


Torsions

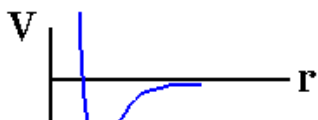


Wechselwirkung
geladener Atome

Electrostatics



van der Waals

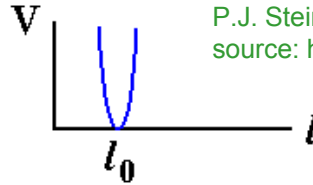




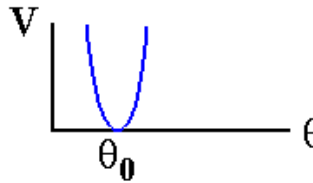
Bestandteile des Kraftfelds

P.J. Steinbach „Introduction to Macromolecular Simulation“
 source: http://cmm.info.nih.gov/intro_simulation/node15.html

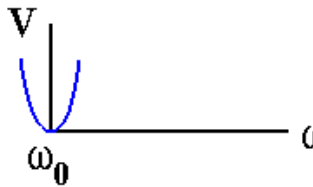
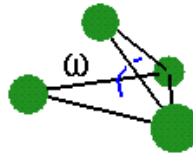
Bonds



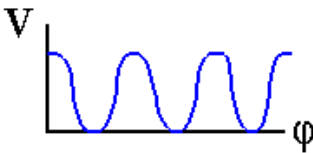
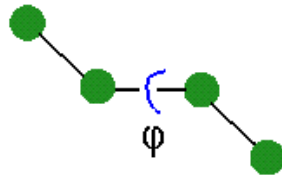
Angles



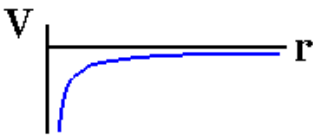
Improper Dihedrals



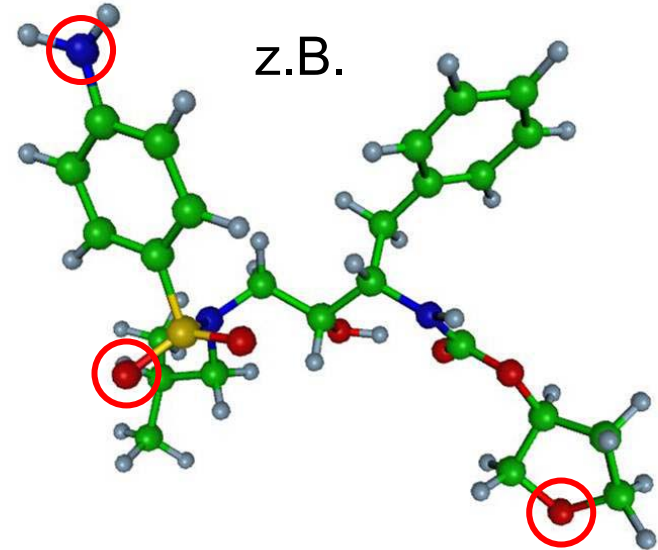
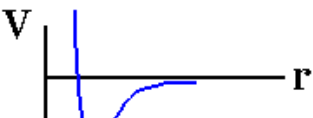
Torsions



Electrostatics



van der Waals



Wechselwirkung
 aller Atome

Bestandteile des Kraftfelds II

P.J. Steinbach „Introduction to Macromolecular Simulation“

... und so sieht's in Formeln gegossen mit $R=(q,p)$ aus:

$$V(\vec{R}) = V_{\text{bonded}}(\vec{R}) + V_{\text{nonbonded}}(\vec{R}),$$

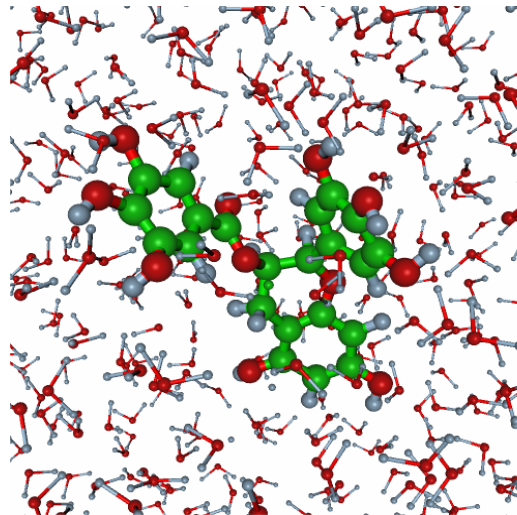
$$V_{\text{bonded}}(\vec{R}) = \sum_{\text{bonds}} k_l (l - l_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \\ + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{torsions}} A_n [1 + \cos(n\phi - \phi_0)]$$

$$V_{\text{nonbonded}}(\vec{R}) = \sum_{i < j} \left(\epsilon_{ij} \left[\left(\frac{r_{ij}^{\text{min}}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{\text{min}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_r \epsilon_0 r_{ij}} \right)$$

Zur Kraftfeldberechnung werden die kartesischen Atomkoordinaten in internen Freiheitsgrade/Variablen (Bindungslängen, Winkel, etc.) umgerechnet werden.

Das Lösungsmittel

Die **Struktur** und **Eigenschaften** eines Moleküls können entscheidend von dem **umgebenden Lösungsmittel** abhängen (z.B. kehren sich die hydrophilen Teile eines Proteins in Wasser nach außen).



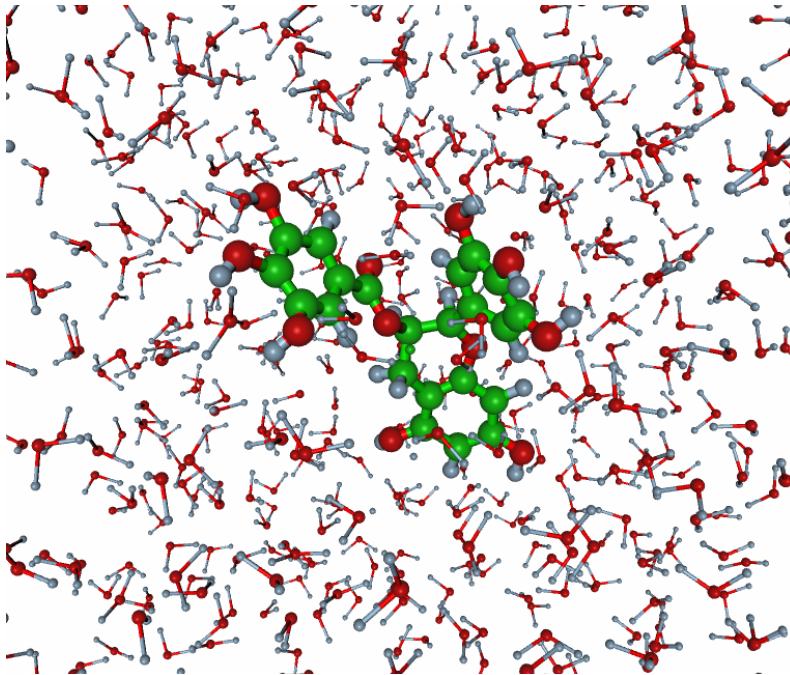
Modellierung:

1. explizit innerhalb des Kraftfeldes
→ Aufwand steigt
2. Ankopplung vereinfachter Lösungsmittelmodelle (z.B. Poisson-Boltzmann)
→ Kopplung & Genauigkeit?

Preisfrage: Wieviel Lösungsmittel simuliert man?

Das Lösungsmittel

Die **Struktur** und **Eigenschaften** eines Moleküls können entscheidend von dem **umgebenden Lösungsmittel** abhängen (z.B. kehren sich die hydrophilen Teile eines Proteins in Wasser nach außen).



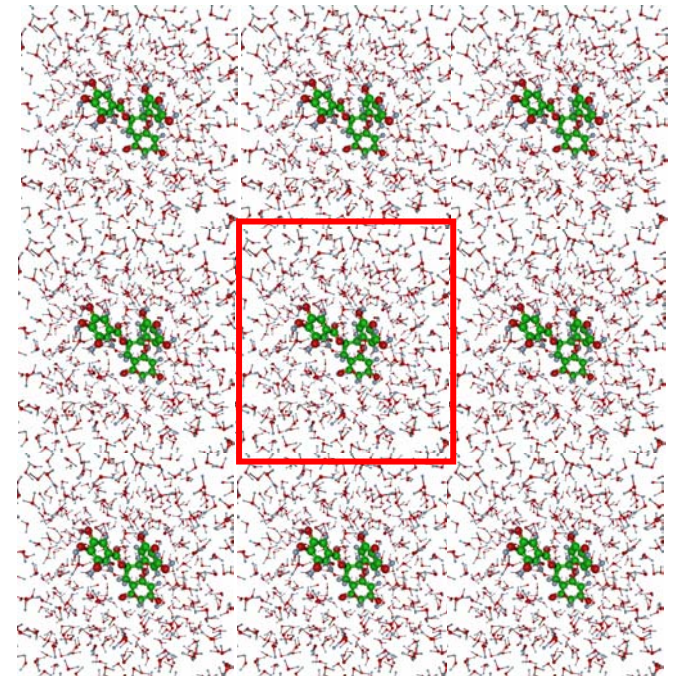
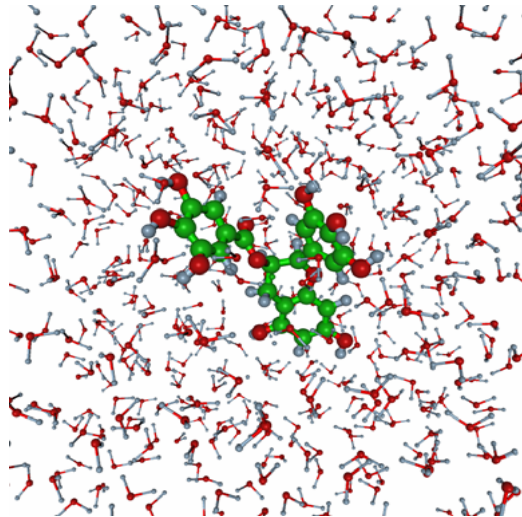
Modellierung:

1. explizit innerhalb des Kraftfeldes
→ Aufwand steigt
2. Ankopplung vereinfachter Lösungsmittelmodelle (z.B. Poisson-Boltzmann)
→ Kopplung & Genauigkeit?

Preisfrage: Wieviel Lösungsmittel simuliert man?

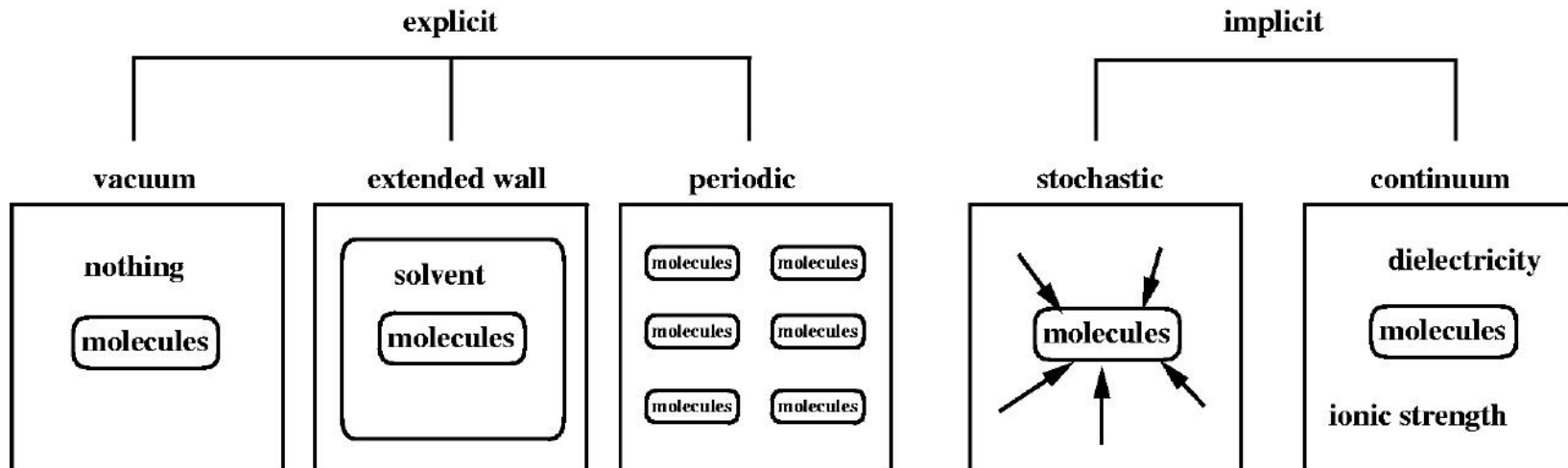
Das Lösungsmittel

Die **Struktur** und **Eigenschaften** eines Moleküls können entscheidend von dem **umgebenden Lösungsmittel** abhängen (z.B. kehren sich die hydrophilen Teile eines Proteins in Wasser nach außen).



Periodische Randbedingungen

Randbedingungen



Was braucht man alles?

1. 3d **Struktur** des Moleküls
2. Mathematisch-physikalisches Modell für die **Dynamik**
3. Numerisches Verfahren zur **Simulation**

Numerische Integratoren

Problem: Analytisch kann man die Bewegungsgleichungen
i. A. nicht lösen \rightarrow numerische Approximation

$$\begin{aligned} \dot{q} &= p \\ \dot{p} &= -DV(q) \end{aligned} \quad \leftrightarrow \quad \dot{R} = f(R)$$

Idee: um von $t=0$ (Startzeit) zu $t=T$ (Endzeit) zu kommen,
gehe in vielen kleinen Zeitschritten Δt vor.

$$r_0 \rightarrow r_{\Delta t} \rightarrow r_{2\Delta t} \rightarrow r_{3\Delta t} \cdots \rightarrow r_{m\Delta t} = r_T$$

Bsp: expliziter Euler

$$\dot{R} \approx \frac{r_{t+\Delta t} - r_t}{\Delta t} \quad \rightarrow \quad r_{t+\Delta t} = r_t + \Delta t \cdot f(r_t)$$

Wichtige Fragen: Genauigkeit, Effizienz, Erhaltungseigenschaften

Leapfrog/Verlet

DAS Verfahren in der Moleküldynamik ist der Leapfrog- oder Verlet-Integrator:

gegeben: \rightarrow q_t, p_t \rightarrow $q_{t+\Delta t}, p_{t+\Delta t}$ \leftarrow gesucht

Hilfsvariable \rightarrow

$$q_{t+0.5\Delta t} = q_t + 0.5\Delta t \cdot p_t$$
$$p_{t+\Delta t} = p_t - \Delta t \cdot DV(q_{t+0.5\Delta t})$$
$$q_{t+\Delta t} = q_{t+0.5\Delta t} + 0.5\Delta t \cdot p_{t+\Delta t}$$

Vorteile: explizit, nur eine Potentialauswertung, symplektisch (erhält Volumen und beinahe die Energie)

Was macht einem das Leben schwer?

(1) unterschiedliche Zeitskalen:

schnellste Dynamik: Bindungsschwingung (fs)

Konsequenz: Schrittweite $\Delta t \approx 1$ fs

(2) unterschiedliche Zeitskalen:

Bindungsschwingung (fs) \leftrightarrow Faltung/globale Konformationsänderungen (ms)

Verhältnis wie 1sec \leftrightarrow 30000 Jahre

Konsequenz: Langzeitintegration (große Anzahl an Simulationsschritten)

(3) „Kondition“ des Problems

Langzeitintegration sind schlecht konditioniert (kleine Anfangsfehler können große Auswirkungen auf das Endresultat haben)

Konsequenz: Reformulierung in viele Kurzzeitsimulationen

Was macht einem das Leben schwer?

(4) Dimension des Problems

Proteine haben hundert-zehntausend Freiheitsgrade

Konsequenz: z.B. „Einfrieren“ schneller Freiheitsgrade

(5) Aufwand der Simulation

Berechnung langreichweitiger Wechselwirkungen
(Elektrostatik) proportional $(Anzahl\ Atome)^2$

Konsequenz: Vereinfach der langreichweitigen Potential-
terme durch sog. „Cut-Off“

(6) Metastabilität/„trapping problem“

zerklüftete Potentiallandschaft mit vielen Energie-Barrieren

Konsequenz: z.B. Simulation bei höheren Temperaturen

Was macht einem das Leben schwer?

(4) Dimension des Problems

Proteine haben hundert-zehntausend Freiheitsgrade

Konsequenz: z.B. „Einfrieren“ schneller Freiheitsgrade

Vorlesung im kommenden Semester:

→ Numerik gewöhnlicher
Differentialgleichungen

(b) Metastabilität, „trapping problem“

zerklüftete Potentiallandschaft mit vielen Energie-Barrieren

Konsequenz: z.B. Simulation bei höheren Temperaturen

Worth Worrying About

„Remember: Simulations are fiction aspiring to emulate reality. Pretty pictures and even a few good numbers do not guarantee good science.“

P.J. Steinbach „Introduction to Macromolecular Simulation“

Typische Fragestellungen

- **Faltung**: beliebige Ausgangsstruktur → Faltungszustand (**Optimierungsprozess**)
- **Kinetik** des Faltungsprozesses (**Dynamik**)
- Bindungsaffinität beim **Docking** (Energiedifferenz)
Energieerwartungswerte (**Statistik**)
- **Konformationsdynamik**: Identifikation wesentlicher
3d-Strukturen, ihrer Lebensdauer sowie der
Übergangsraten (**Statistik & Dynamik**)
- ...

Virtual Screening

Targets

X-ray, NMR, Modell

energy function + 3D data

conformation-dynamics

pharma site

Ligands

2-D data

energy function + 3-D data

conformation-dynamics

pharma-filter

Input

preparation

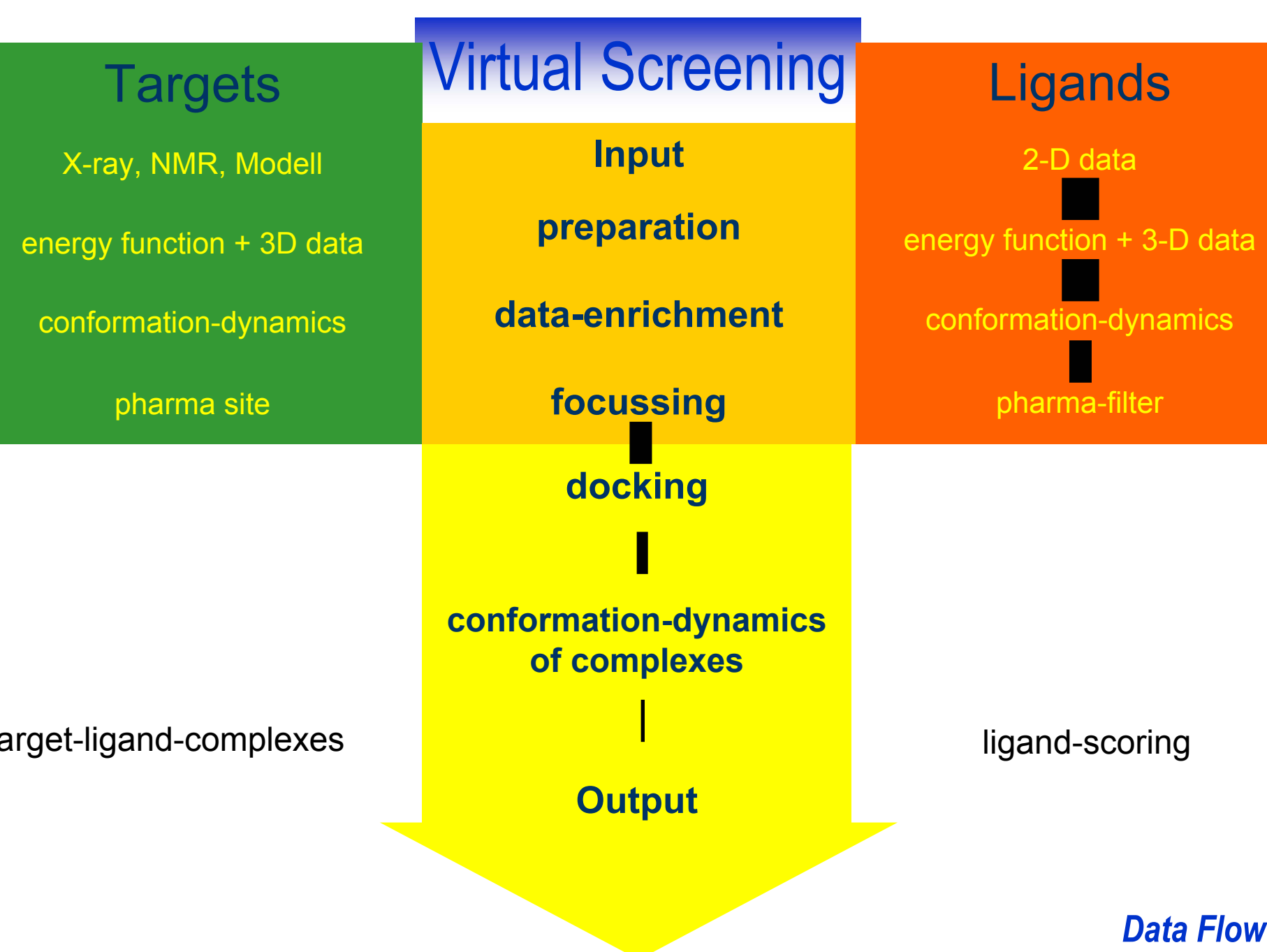
data-enrichment

focussing

docking

**conformation-dynamics
of complexes**

Output



Targets

X-ray, NMR, Modell

Virtual Screening

Input ■

Ligands

2-D data

Targets

X-ray, NMR, Modell

energy function + 3D data

Virtual Screening

Input

preparation

Ligands

2-D data

energy function + 3-D data

$$\begin{aligned} V(q) = & \sum_{k,l} V_{\text{bond}}(q_k, q_l) + \\ & \sum_{k,l,j} V_{\text{angle}}(q_k, q_l, q_j) + \\ & \sum_{k,l,j,m} V_{\text{out-of-plane}}(q_k, q_l, q_j, q_m) + \\ & \sum_{k,l,j,m} V_{\text{dihedral}}(q_k, q_l, q_j, q_m) + \\ & \sum_{k,l} V_{\text{Lennard-Jones}}(q_k, q_l) + \\ & \sum_{k,l} V_{\text{Coulomb}}(q_k, q_l) \end{aligned}$$

MMFF:

quantum chemically based Force Field

automatic parameterization of „any“

biomolecule

Virtual Screening

Targets

X-ray, NMR, Modell

energy function + 3D data

conformation-dynamics

Input

preparation

data-enrichment

Ligands

2-D data

energy function + 3-D data

conformation-dynamics

$$\dot{q} = M^{-1}p, \quad \dot{p} = -\text{grad } V$$

$$f_0(x) = \underbrace{\frac{1}{Z_p} \exp\left(-\frac{\beta}{2} p^T M^{-1} p\right)}_{P(p)} \underbrace{\frac{1}{Z_q} \exp(-\beta V(q))}_{Q(q)}$$

Virtual Screening

Targets

X-ray, NMR, Modell

energy function + 3D data

conformation-dynamics

pharma site

Input

preparation

data-enrichment

focussing

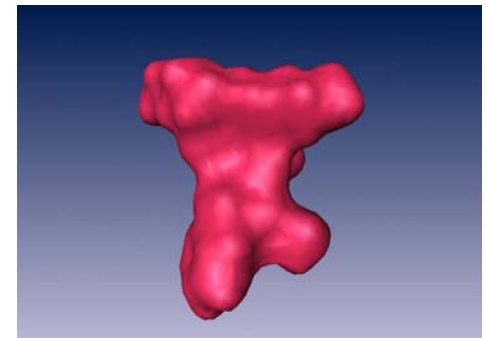
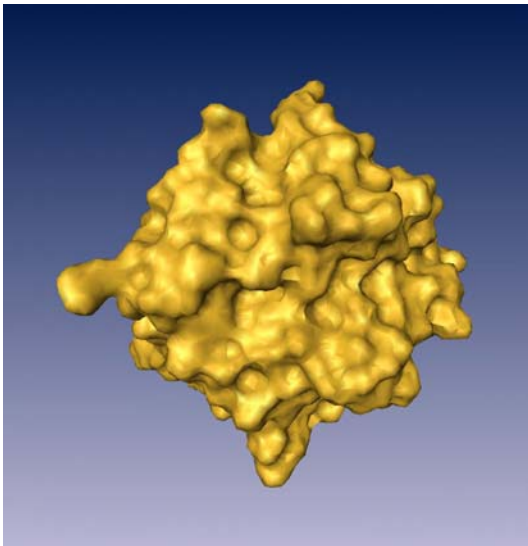
Ligands

2-D data

energy function + 3-D data

conformation-dynamics

pharma-filter



Virtual Screening

Targets

X-ray, NMR, Modell

energy function + 3D data

conformation-dynamics

pharma site

Input

preparation

data-enrichment

focussing

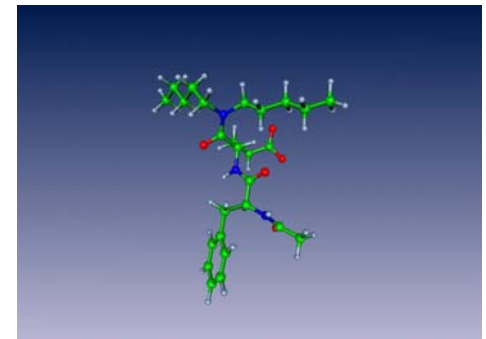
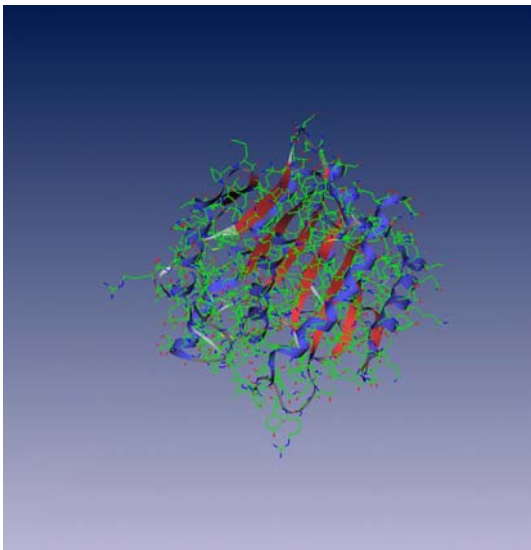
Ligands

2-D data

energy function + 3-D data

conformation-dynamics

pharma-filter



Virtual Screening

Targets

X-ray, NMR, Modell

energy function + 3D data

conformation-dynamics

pharma site

Input

preparation

data-enrichment

focussing

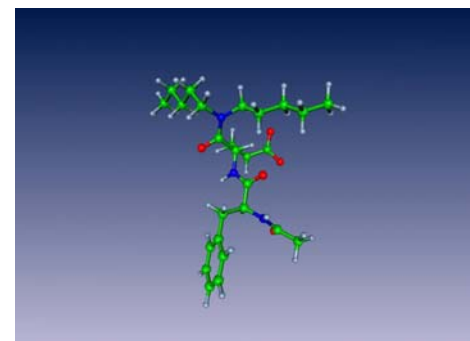
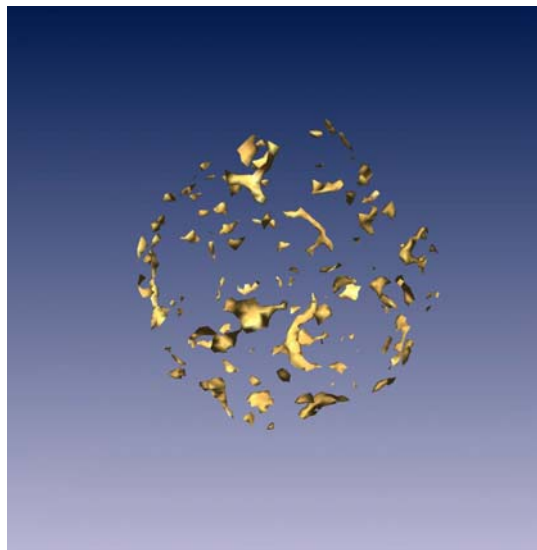
Ligands

2-D data

energy function + 3-D data

conformation-dynamics

pharma-filter



Virtual Screening

Targets

X-ray, NMR, Modell

energy function + 3D data

conformation-dynamics

pharma site

Input

preparation

data-enrichment

focussing

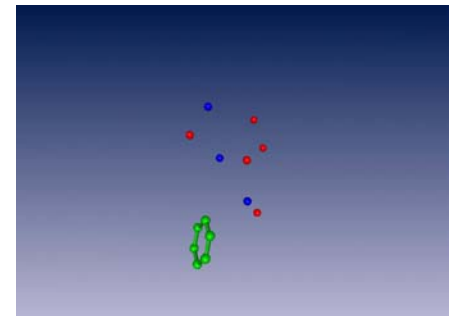
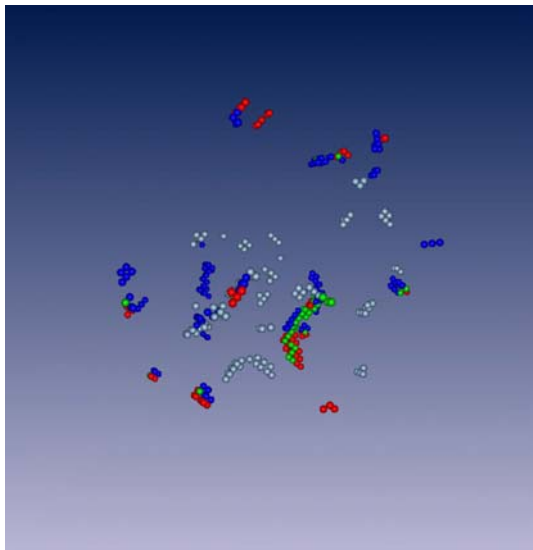
Ligands

2-D data

energy function + 3-D data

conformation-dynamics

pharma-filter



Targets

X-ray, NMR, Modell

energy function + 3D data

conformation-dynamics

pharma site

Virtual Screening

Input

preparation

data-enrichment

focussing

docking ■ ■

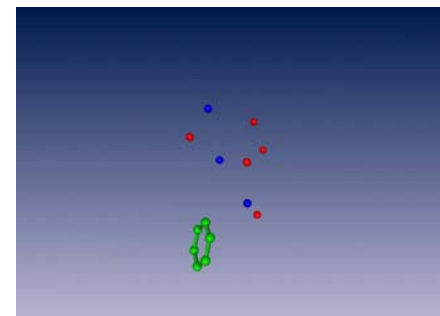
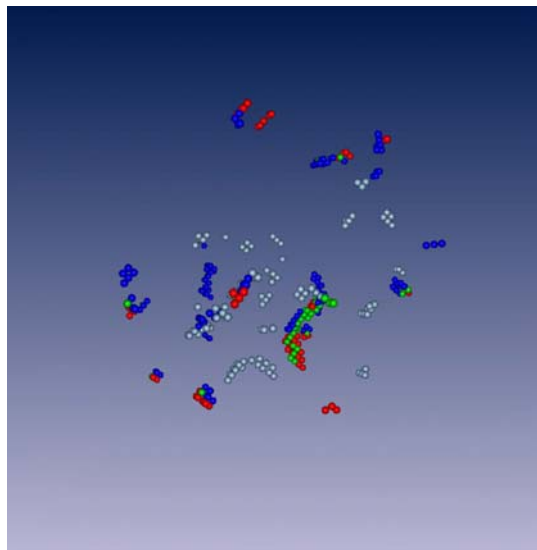
Ligands

2-D data

energy function + 3-D data

conformation-dynamics

pharma-filter



Point Matching
+
Fine Tuning

Virtual Screening

Targets

X-ray, NMR, Modell

energy function + 3D data

conformation-dynamics

pharma site

Ligands

2-D data

energy function + 3-D data

conformation-dynamics

pharma-filter

Input

preparation

data-enrichment

focussing

docking

**conformation-dynamics
of complexes**

Output

target-ligand-complexes

ligand-scoring

