

# V6 Proteinstrukturen / Sampling

Berechnung von Eigenschaften des Systems aus MD-Simulationen

- Wassermoleküle und H-Brücken
- Zustandssumme (Wdh.)
- Ensembles (Wdh.)
- Umbrella Potential-Simulationen (Zwangspotentiale): Sampling von energetisch ungünstigen Zuständen

Proteinstrukturen (aus Kristallstrukturen und Homologiemodellen)

- Hinzufügen und Zuordnung von H-Atomen (apolare und polare)
- Rotamere
- Protonierungszustände titrierbarer Aminosäuren
- Protonierungszustände von Liganden

# Eigenschaften des Solvens

Im Gegensatz zu der Mehrzahl aller anderen Verbindungen kommt Wasser in allen drei Aggregatzuständen natürlich vor.

Schmelzpunkt	$0^{\circ} \text{ C} = 273 \text{ K}$	
Siedepunkt	$100^{\circ} \text{ C} = 373 \text{ K}$ (bei Normaldruck)	
Wärmekapazität	$75.366 \text{ J mol}^{-1} \text{ K}^{-1}$	} sehr hohe Werte!
Verdampfungsenthalpie	$40.7 \text{ kJ mol}^{-1}$	
Schmelzenthalpie	$6.01 \text{ kJ mol}^{-1}$	

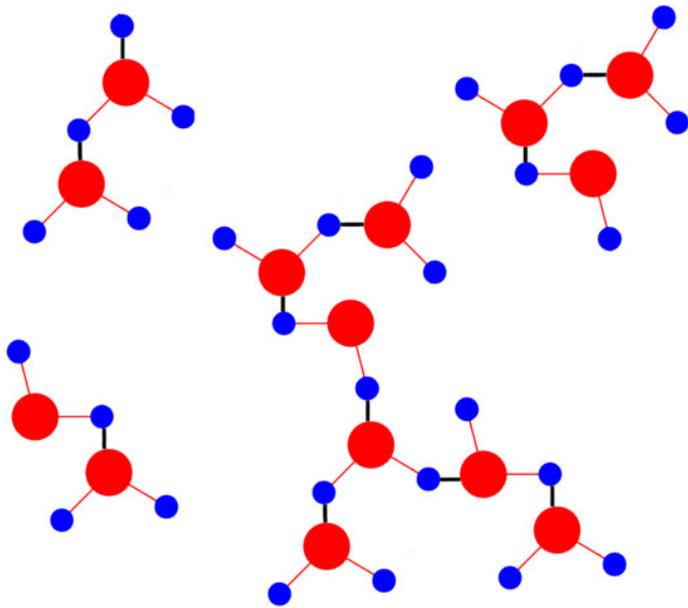
Anomalie:  
größte Dichte bei  $4^{\circ} \text{ C}$



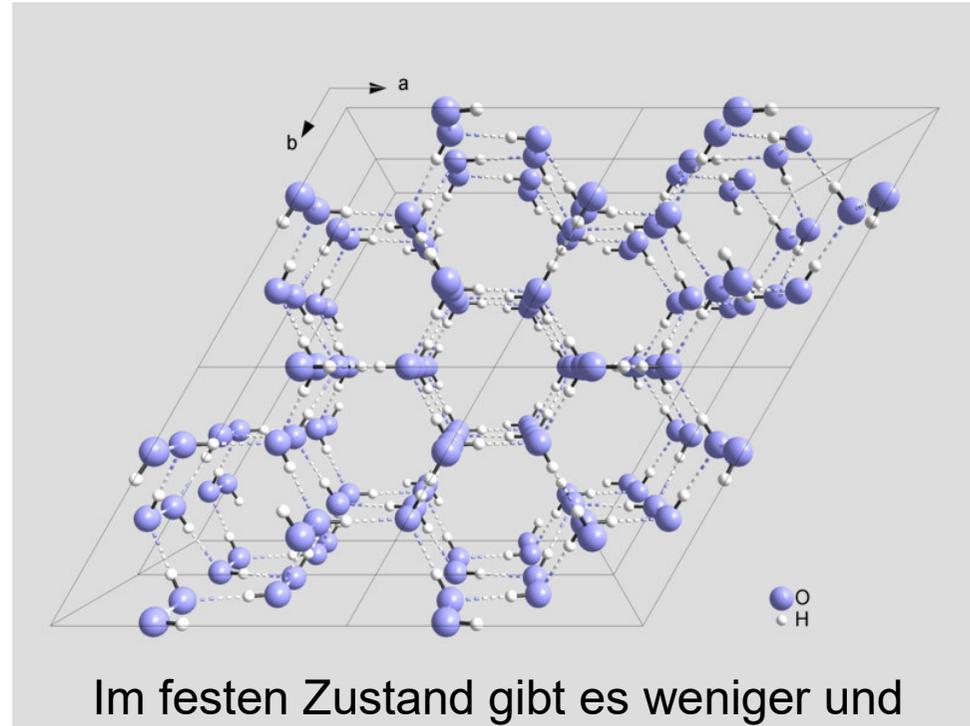
Bildquelle: Wikipedia

# Eigenschaften des Wassers (I)

Eine ganze Reihe der erwähnten Besonderheiten des Wassers lassen sich durch die Wasserstoffbrücken erklären:



Stärke einer H-Brücke  
O–H ... O ca. 21 kJ/mol  
→ hoher Siedepunkt



Im festen Zustand gibt es weniger und längere H-Brücken als im flüssigen.  
→ Die Dichte von Eis ist kleiner als die von flüssigem Wasser

# Wassermodelle (I)

Wieviele H-Brücken bildet ein Wassermolekül im Durchschnitt aus ?

→ MD-Simulation

Wie modelliert man Wasser adäquat ?

Die Modelle müssen experimentelle Größen richtig wiedergeben, z.B. Siedepunkt, Schmelzpunkt, Verdampfungsenthalpie, Dipolmoment, Struktur, ... ?

Ein einfaches Modell kann nicht allen diesen Anforderungen perfekt gerecht werden.

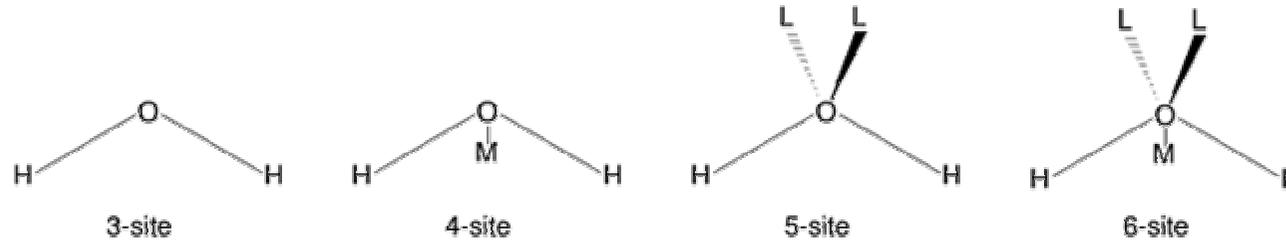
Im einfachsten Fall stellen wir die Wassermoleküle wie alle anderen Atome in einem Kraftfeld durch Bindungen, Punktladungen und van der Waals Radien dar. Die Wechselwirkungen ergeben sich dann aus elektrostatischen und van der Waals-Kräften.

$$E = \sum_{j>i}^{vdW} \left( \frac{A}{r^{12}} - \frac{B}{r^6} \right) + \sum_{j>i}^{Estat} \frac{q_i q_j}{r_{ij}}$$

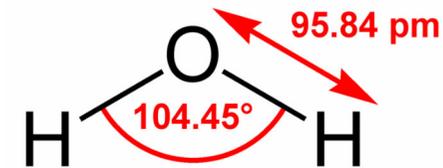
# Wassermodelle (II)

Gebräuchliche Wassermodelle (>40 verschiedene wurden publiziert)

B.Guillot, A reappraisal of what we have learnt during three decades of computer simulations on water, *J. Mol. Liq.* **101** (2002) 219-260



	SPC	TIP3P	TIP4P	TIP5P	exptl.
$r(\text{OH})$ [Å]	1.0	0.9572	0.9572	0.9572	
HOH [°]	109.47	104.52	104.52	104.52	
$q(\text{O})$ [e]	-0.82	-0.834	-	-	
$q(\text{H})$ [e]	0.41	0.417	0.52	0.241	
$q(\text{M/L})$ [e]	-	-	-1.04	-0.241	
$\epsilon$	65	82	53	81.5	78.4
$\mu$ [Debye]	2.27	2.35	2.18	2.29	1.8 (Vakuum, Dampf)
					2.3 - 2.95 (flüssig)



Bildquelle: Wikipedia

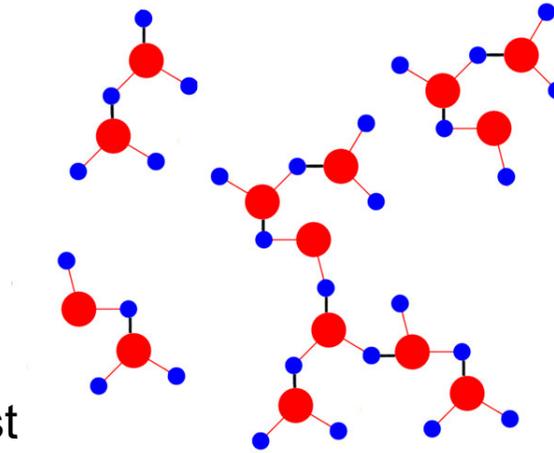
# Eigenschaften des Wassers (II)

Wieviele H-Brücken bildet ein Wassermolekül im Durchschnitt aus ?

TIP4P: 3.59 bei 25°C, 3.24 bei 100°C (höhere Mobilität und geringere Dichte)

Abhängig von der Definition von Wasserstoffbrücken (max. Reichweite, Winkelabhängigkeit) erhält man kleinere Werte (2.357 bei 25°C)

Übliche Kriterien für H-Bindungen sind eine maximale Distanz von  $\leq 0.35$  nm zwischen Donor- und Akzeptoratom sowie ein Winkel  $D-H-A \geq 120^\circ$  (also O-H-O in möglichst linearer Anordnung)



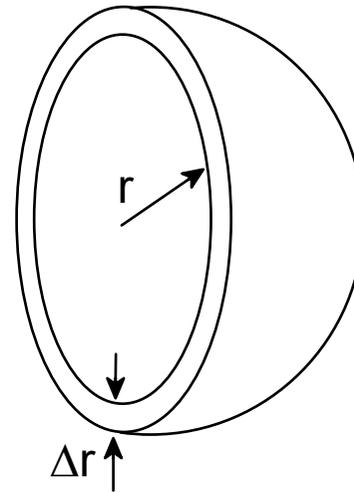
Lit: W. L. Jorgensen and J. D. Madura (1985). "Temperature and size dependence for Monte Carlo simulations of TIP4P water". *Mol. Phys.* **56** (6): 1381.

Jan Zielkiewicz (2005). "Structural properties of water: Comparison of the SPC, SPCE, TIP4P, and TIP5P models of water". *J. Chem. Phys.* **123**: 104501.

# Radiale Verteilungsfunktion (I)

Zurück zur Frage, wieviele H-Brücken bildet ein Wassermolekül im Wasser aus? Dazu sehen wir uns zunächst an, wieviele Wassermoleküle sich in der Entfernung  $r$  um dieses eine Molekül befinden und stellen eine radiale Verteilungsfunktion auf:

$$g(r, \Delta r) = \frac{V}{N^2} \frac{\langle N(r, \Delta r) \rangle_M}{2\pi r^2 \Delta r}$$



Aus der Simulation mit  $M$  Stützpunkten suchen wir also die Anzahl aus allen Teilchen  $N$  aus dem Gesamtvolumen  $V$ , die sich in der Kugelschale der Dicke  $\Delta r$  befinden.

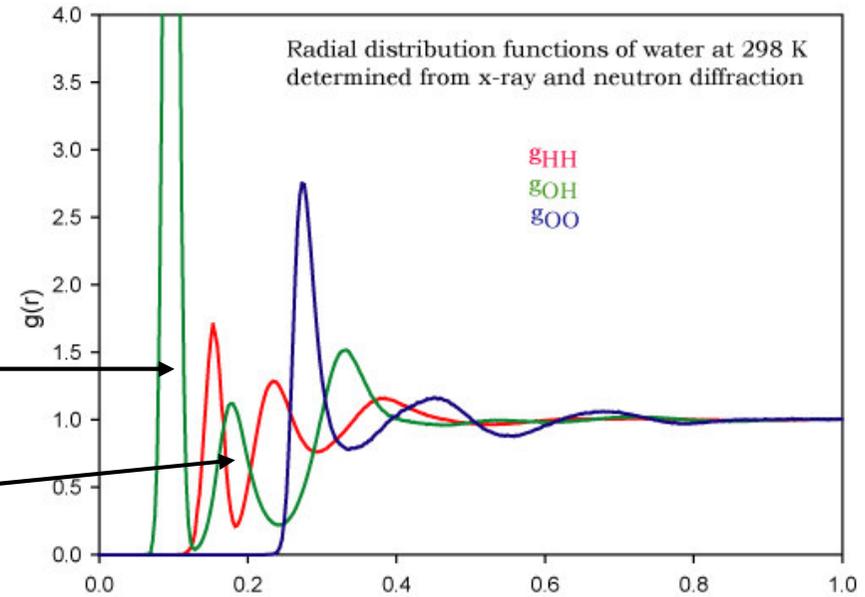
In der Praxis weisen wir die Teilchen in jeder Kugelschale einem *bin* zu, der den jeweiligen Abstand  $r$  repräsentiert und können so ein Histogramm erzeugen.

# Radiale Verteilungsfunktion von Wasser

Radiale Verteilungsfunktion  $g(r)$   
für flüssiges Wasser

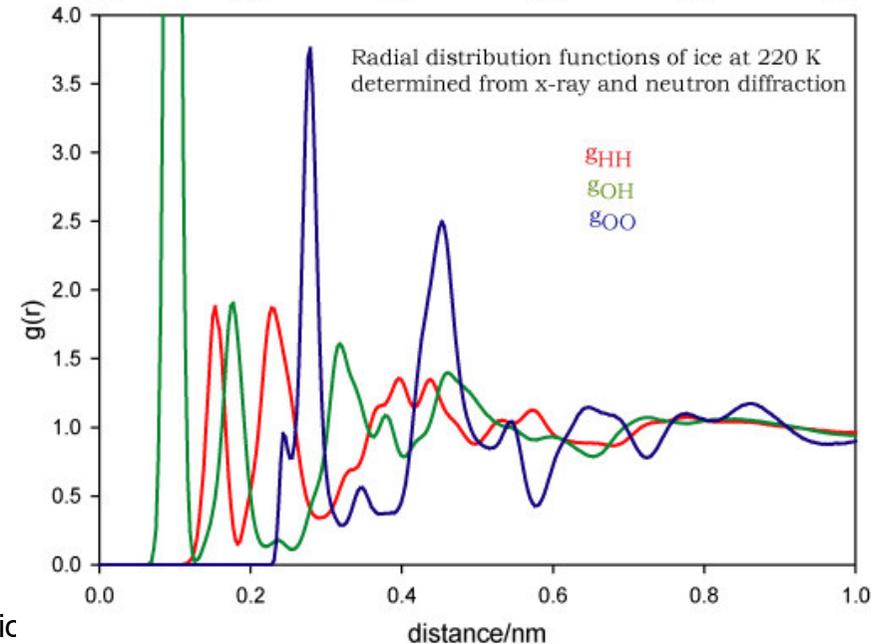
O-H Bindung

O-H H-Brücke



Aufgrund der geordneten Struktur im Eiskristall sieht hier die Verteilung anders aus als im flüssigen Wasser, z.B. sind die H-Brücken schärfer definiert und es existiert eine größere Fernordnung

Bildquelle: <http://ptcl.chem.ox.ac.uk>



## Radiale Verteilungsfunktion (II)

Die graphische Auftragung einer solchen radialen Verteilungsfunktion bietet sich natürlich für den Vergleich von Simulationsdaten mit experimentellen Daten (soweit diese verfügbar sind) an.

z.B. Kalibrierung der Wassermodelle am Experiment

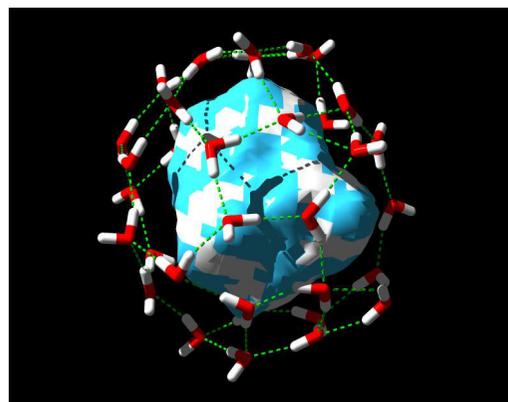
Weitere Anwendungen:

Solvatationshülle von Ionen, Aminosäuren in Bindungstaschen

Beobachtung:

Wassermoleküle über hydrophoben Proteinbereichen sind weniger mobil und bilden vorwiegend gegenseitige H-Brücken aus.

→ „mikroskopische Eiskristalle“



Problem: Zuordnung von Wassermolekülen in Kristallstrukturen

E.Nittinger et al. Evidence of Water Molecules - A Statistical Evaluation of Water Molecules Based on Electron Density, *J.Chem.Inf.Model.* **55** (2015) 771-783



## Zustandssumme (Wdh.)

Die Energie eines Moleküls ergibt sich als Summe seiner (temperatur-abhängigen) Energieterme aufgrund seiner Freiheitsgrade für die Translation, Rotation und Vibration. Folglich muss diese Summe  $Q$  der Nenner in der Gleichung für die Boltzmannverteilung sein

$$\frac{\text{Anzahl Teilchen im Zustand } i}{\text{Summe aller Teilchen}} = \frac{N_i}{N} = \frac{\exp\left(-E_i/kT\right)}{\sum \exp\left(-E_i/kT\right)} = \frac{\exp\left(-E_i/kT\right)}{Q}$$

Die Summe  $Q$  der inneren Energie wird auch als **Zustandssumme (partition function)** bezeichnet.

Je nachdem, welche Größen beim betrachteten Ensemble (an Molekülen) konstant sind (z.B. Druck, Volumen, Temperatur, Teilchenzahl) verändert sich auch die Zustandssumme  $Q$ .

# Sampling (I)

Als Simulation werden solche Methoden bezeichnet die eine statistische Komponente beinhalten. Deshalb können Fehler auch von der zeitlich begrenzten Simulationsdauer stammen, da wir nur über diesen Zeitabschnitt das Sampling durchführen.

Jede makroskopische Größe/Observable  $X$  (z.B. Enthalpie, Entropie, Wärmekapazität) läßt sich aus der Zustandssumme  $Q$  ableiten. (vgl. Vorlesung 2)

$$\text{Enthalpie } H \quad H = kT^2 \left( \frac{\partial \ln Q}{\partial T} \right)_V + kTV \left( \frac{\partial \ln Q}{\partial V} \right)_T$$

$$\text{Freie Enthalpie } G \quad G = -kT \ln Q + kTV \left( \frac{\partial \ln Q}{\partial V} \right)_T$$

$$\text{Entropie } S \quad S = k \left( \frac{\partial \ln Q}{\partial T} \right)_V + k \ln Q$$

## Sampling (II)

Für ein einzelnes Molekül können wir die Zustandssumme  $Q$  für die einzelnen translatorischen, rotatorischen und vibronischen Terme explizit formulieren (also nur kinetische Energie). Bei mehreren Molekülen müssen wir die intra-molekularen WW-Energien hinzunehmen (potentielle Energie). In einem Ensemble mit vielen Teilchen werden die einzelnen Energieniveaus jedoch so dicht beieinander sein, daß wir von einer kontinuierlichen Verteilung ausgehen können. Deshalb ersetzen wir die Summation diskreter Energieterme durch die Integration über alle Koordinaten  $\mathbf{r}$  und Impulse  $\mathbf{p}$ , den sog. Phasenraum.

$$Q = \sum_i^{\text{Zustände}} e^{-E_i/kT} = \int e^{-E(\mathbf{r},\mathbf{p})/kT} d\mathbf{r} d\mathbf{p}$$

Weiterhin nehmen wir an, daß die  $N$  Teilchen unseres Ensembles miteinander wechselwirken.

## Sampling (III)

Im Falle eines idealen Gases müssen wir nur die kinetische Energie betrachten, für alle anderen Systeme auch die potentielle Energie (Anziehung/Abstoßung) wie wir sie aus einem Kraftfeld erhalten.

$$E_{\text{potentiell}} = \sum^{\text{Bindungen}} \frac{k}{2} (r - r_0)^2 + \sum^{\text{Winkel}} \frac{k}{2} (\theta - \theta_0)^2 + \sum^{\text{Torsionen}} \sum C_n \cos(1 + n\omega) \\ + \sum_{j>i}^{\text{vdW}} \left( \frac{A}{r^{12}} - \frac{B}{r^6} \right) + \sum_{j>i}^{\text{Estat}} \frac{q_i q_j}{r_{ij}} + \dots$$

Zusammengenommen ergeben diese Energie die innere Energie  $U$

$$U = kT^2 \left( \frac{\partial \ln Q}{\partial T} \right)_V$$

Da  $\frac{\partial \ln Q}{\partial T} = \frac{\partial Q}{Q \partial T}$  erhält man

$$U = \frac{kT^2 \partial Q}{Q \partial T} = \sum_i^{\text{Zustände}} \frac{E_i}{Q} e^{-E_i/kT}$$

# Sampling (IV)

Da alle Energiezustände  $E_i$  der Boltzmannverteilung unterliegen, erhalten wir dafür eine sog. Boltzmannsche Verteilungsfunktion  $P$

$$P(E_i) = Q^{-1} e^{-E_i/kT} \quad \text{im diskreten bzw.}$$

$$P(r, p) = Q^{-1} e^{-E(r,p)/kT} \quad \text{Im kontinuierlichen Fall}$$

Setzen wir  $P$  in die Gleichungen für die innere Energie  $U$  ein, so erhalten wir

$$U = \sum_i^{\text{Zustände}} E_i P(E_i)$$

$$U = \int E(r, p) P(r, p) dr dp$$

Damit ist die innere Energie  $U$  also die Summe aller Energien, gewichtet mit der Wahrscheinlichkeit in diesem Zustand vorzukommen, d.h.  $U$  ist die durchschnittliche Energie des Systems.

# Sampling (V)

Generell kann jede makroskopische Observable  $X$  als Erwartungswert (Durchschnitt)  $\langle X \rangle$  der entsprechenden mikroskopischen Größe  $X(r,p)$  gewichtet mit der Boltzmannverteilung berechnet werden.

Dazu brauchen wir eine ausreichende (repräsentative) Menge an *samples* (Stützpunkten)  $M$ .

$$\langle X \rangle_M = \frac{1}{M} \sum_i^M X(r_i, p_i)$$

Eine Ansammlung entsprechender Konfigurationen (die bei der entsprechenden Temperatur vorwiegend populiert sind) wird in der Regel aus MD-Simulationen erhalten.

Bsp. Innere Energie  $U$

$$\langle U \rangle_M = \frac{1}{M} \sum_i^M E_i = \langle E \rangle_M$$

# Sampling (VI)

Der so berechnete Erwartungswert einer Größe  $X$  weist einen statistischen Fehler  $\sigma(X)$  auf, der umgekehrt proportional zur Anzahl der Stützpunkte  $M$  ist

$$\sigma(X) \propto \frac{1}{\sqrt{M}}$$

Neben einer ausreichenden Anzahl von Stützpunkten (= lange Simulationszeiten) muß darüberhinaus auch noch das betrachtete Ensemble repräsentativ sein (hinreichend groß), ansonsten wird der systematische Fehler groß.

Da Größen wie die Freie Energie  $G$  und die Entropie  $S$  direkt von der Zustandssumme  $Q$  abhängen (und deshalb von der Boltzmann-gewichteten Energie), werden Sie v.a. durch energiereiche Zustände bestimmt die entsprechend selten vorkommen. D.h. die genaue Bestimmung von  $G$  oder  $S$  wird sehr (Rechenzeit-)aufwendig, da wir über lange Zeitspannen simulieren müßten.

# Umbrella sampling

Problem: Wie kann ich energetisch ungünstige Zustände adäquat sampeln, zumal diese selten vorkommen?

Dazu muß die Energiefunktion so modifiziert werden, daß diese Zustände „wahrscheinlicher“ werden.

$$E'(r) = E(r) + W(r)$$

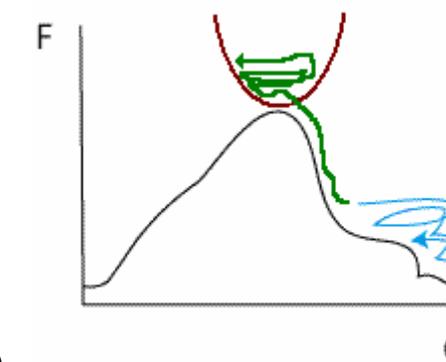
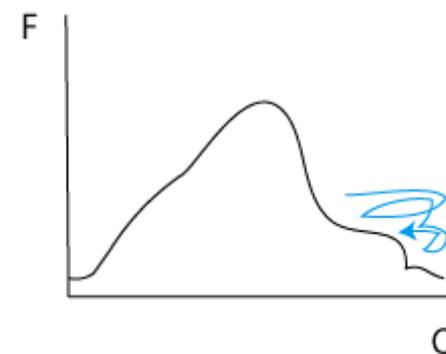
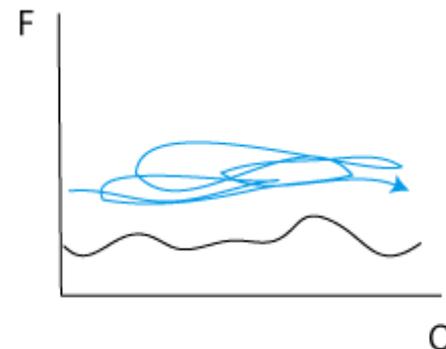
$$W(r) = k(r - r_o)^2$$

$r_o$  und  $k$  in der Gewichtungsfunktion  $W(r)$  werden so gewählt, daß die Energie an der gewünschten Position günstiger wird.

Da nun aber keine echte Boltzmannverteilung mehr vorliegt, muß auch der Erwartungswert korrigiert werden

$$\langle A \rangle = \frac{\langle A(r) e^{+W(r)/kT} \rangle}{\langle e^{+W(r)/kT} \rangle}$$

Bildquelle: wikipedia (modifiziert)



# Monte Carlo (MC)

Wie generiert man repräsentative Ensembles?

In sog. Monte Carlo Verfahren fügt man dem System eine zufällige Änderung zu, und akzeptiert diese falls die neue Energie günstiger ist, oder eine höhere Energie entsprechend der Boltzmann-Verteilung der Energie zufällig angenommen werden kann.

→ Metropolis Algorithmus

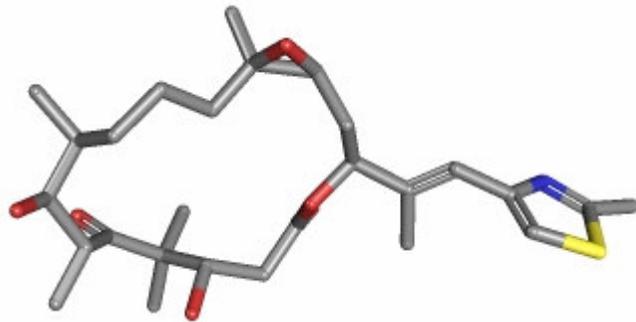
Im Gegensatz zu MD bleibt die Energie deshalb nicht erhalten, sondern fluktuiert (stark).

In der Praxis braucht man hier einige Millionen Schritte um die Energiehyperfläche um die Startkonfiguration herum angemessen zu sampeln.

→ Ideal wenn die Erzeugung neuer Konfigurationen rechentechnisch einfach ist (z.B. durch stochastische Geometrieänderungen anstelle der Verwendung von Ableitungen wie bei MD)

# Konformationssampling von Macrozyklen (I)

Viele natürliche vorkommende Wirkstoffe enthalten große, flexible Ringsysteme, sog. Macrozyklen. Um deren mögliche Konformationen (in Lösung oder bei der Bindung an Enzyme) adäquat zu sampeln, werden auch stochastische Verfahren eingesetzt, da zwingende Randbedingungen berücksichtigt werden müssen, etwa die Stereochemie.



Epothilone B

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

Ansätze:

a) Systematische Variation der Diederwinkel gefolgt von Energieminierung und Clustering.

Wang et al. *J.Chem.Inf.Model.* **56** (2016) 2194.

# Konformationssampling von Macrozyklen (II)

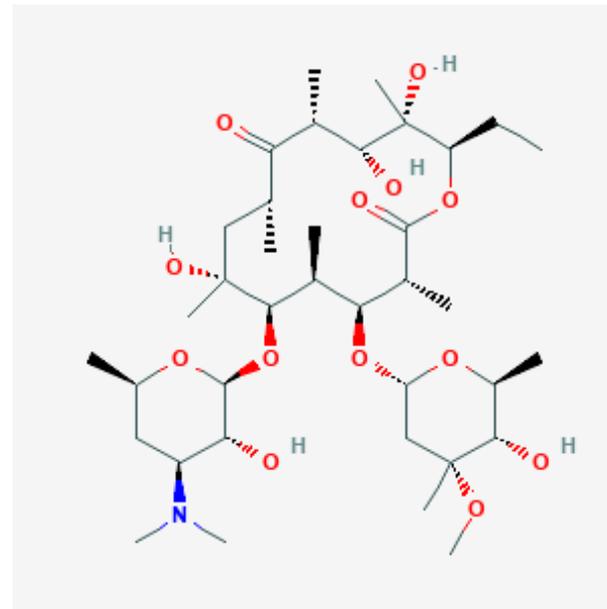
Stochastische Ansätze:

- b) Zufällige Generierung von neuen Atomkoordinaten bzw. Platzierung von Fragmenten mit fester Stereochemie. Danach Energieminimierung (unterschiedliche  $\epsilon_r$  je nach Umgebung) und Clustering (z.B. RMSD-Abweichung  $< 2.0 \text{ \AA}$  innerhalb eines Clusters)
- c) Kurze MD-Simulation bei sehr hoher Temperatur (5000 K) um neue Konformation zu erhalten. Danach Energieminimierung, Überprüfung korrekter Stereochemie und Clustering

Nebeneffekt: Anhand einer Boltzmann-gewichtung der Energien der Konformere läßt sich die Entropie abschätzen

Erythromycin

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



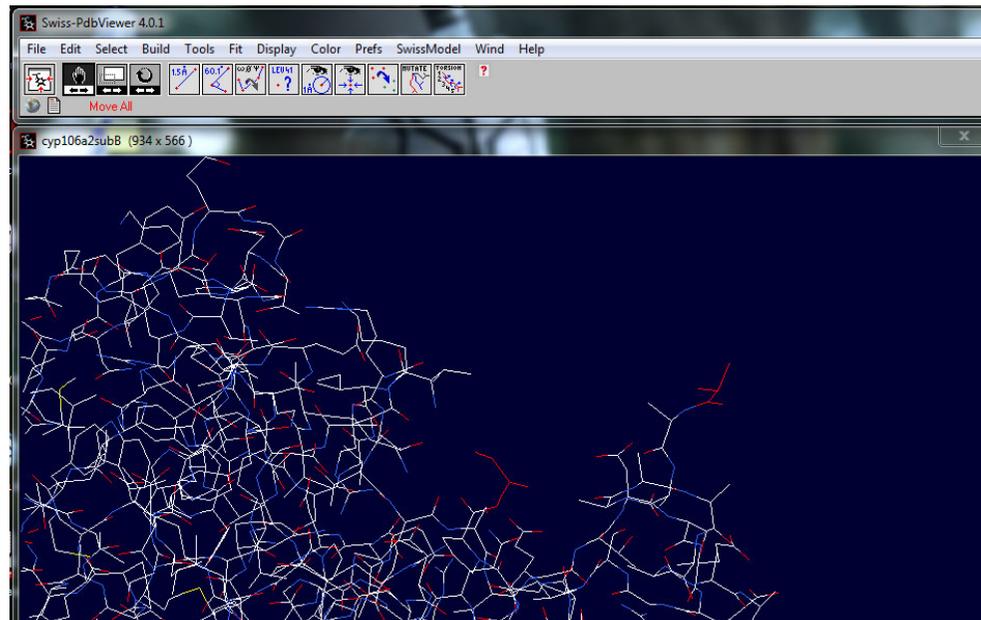
# Loopgenerierung bei Proteinen (I)

Da Loops flexibel sind, sind diese Bereiche in Kristallstrukturen oft nicht enthalten (*wieso?*) (*missing residues*).

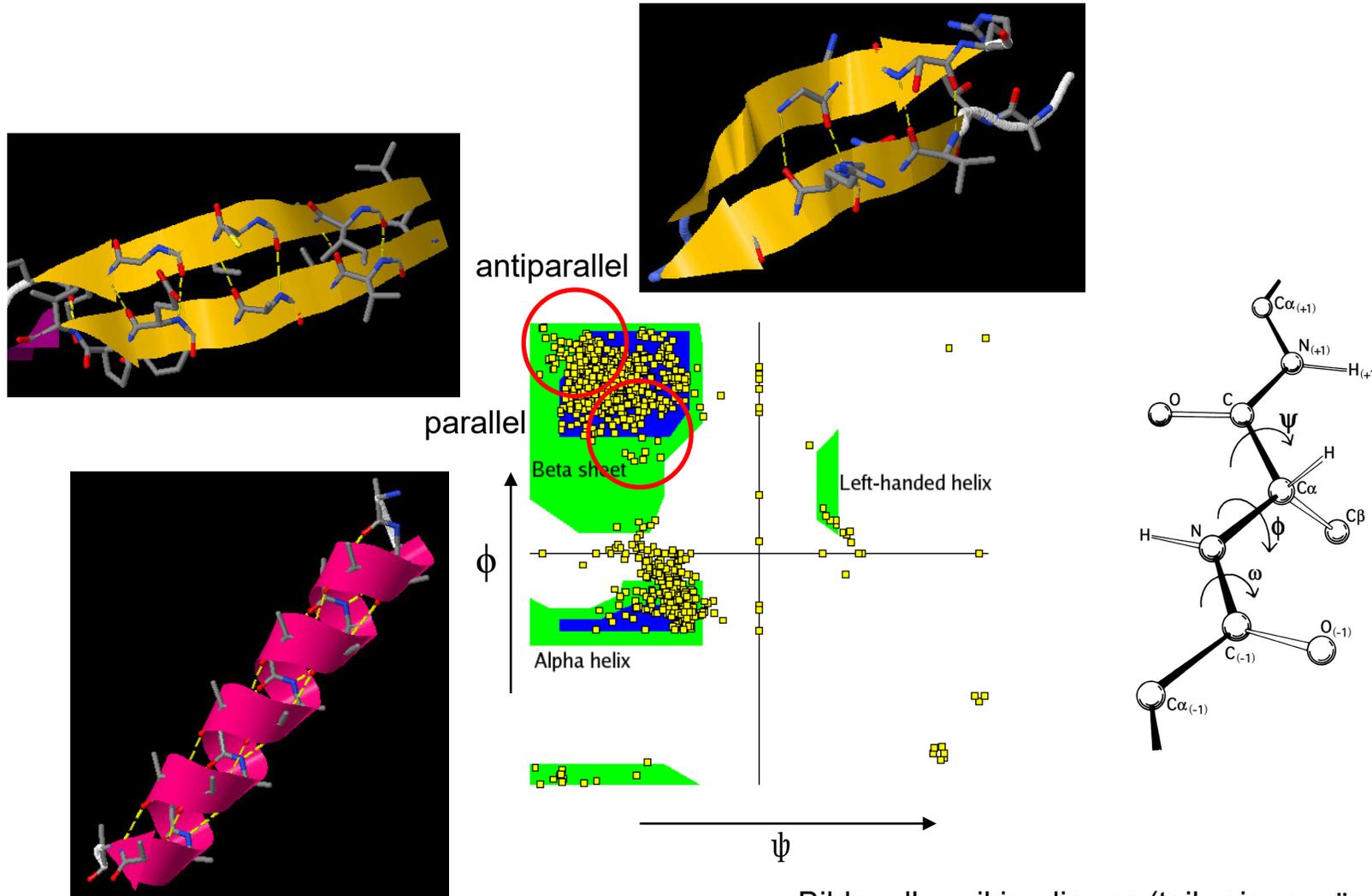
Ebenso müssen bei der Homologiemodellierung die nicht in der Vorlage (*template*) enthaltenen Bereiche konstruiert werden.

Ansätze im Programm spdbv (SWISS-Pdb-Viewer)

1. Loopdatenbank aus bekannten Loops entsprechender Länge in Kristallstrukturen
2. de novo Konstruktion (systematische Variation erlaubter Diederwinkel. Vgl. Ramachandranplot)

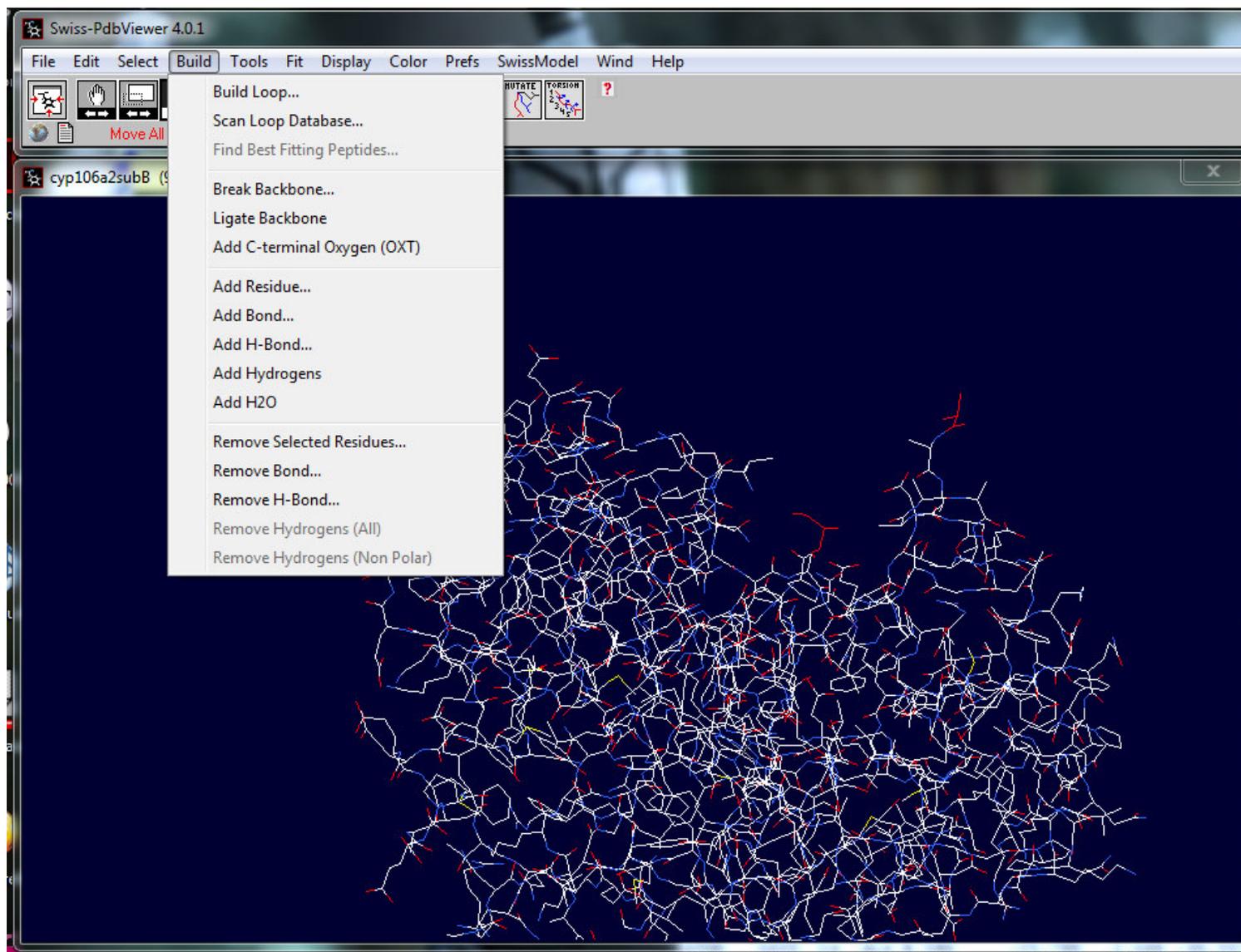


# Ramachandran Plot

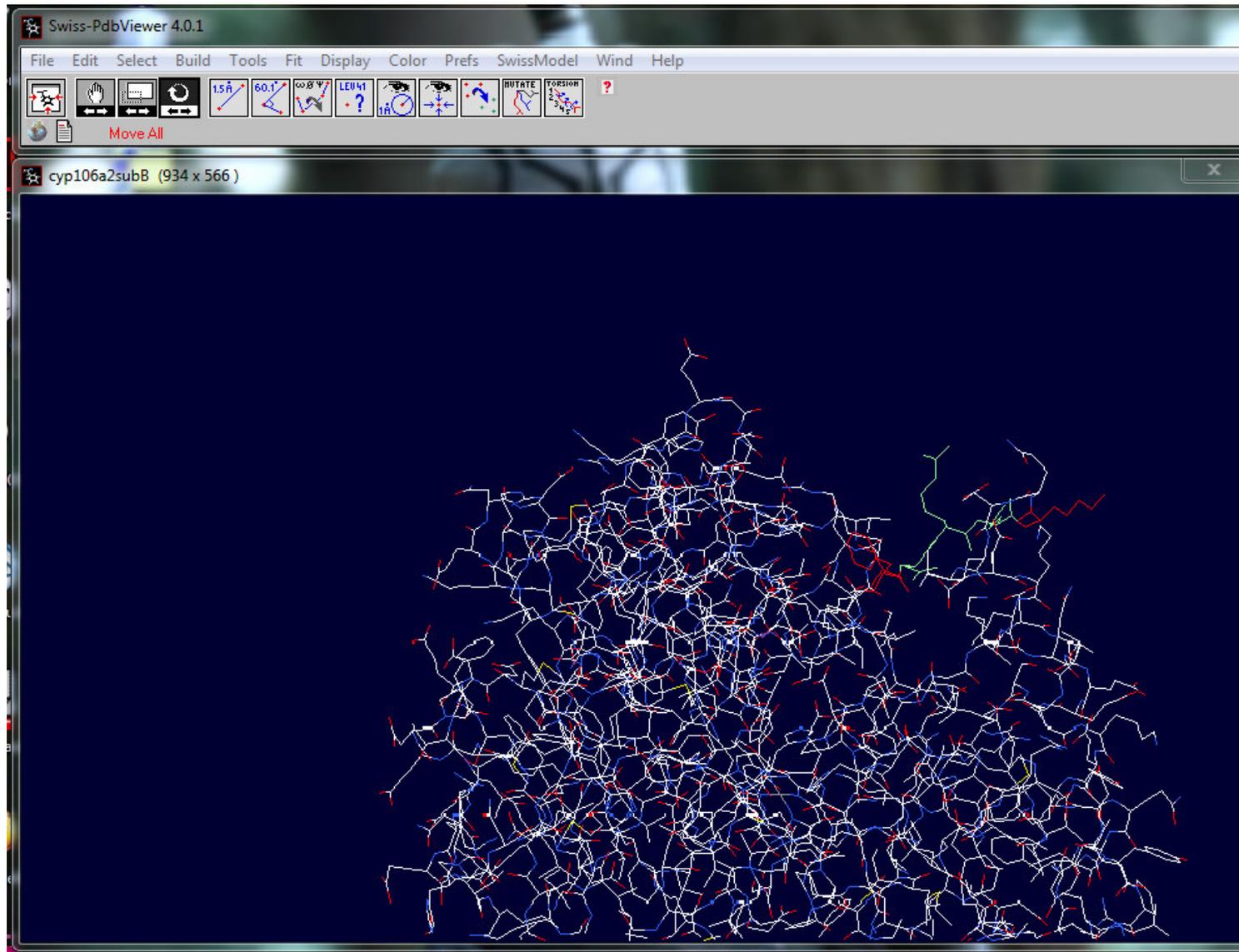


Bildquelle: wikipedia.org (teilweise verändert)

# Loopgenerierung bei Proteinen (II)



# Loopgenerierung bei Proteinen (III)



## Wdh: mikroskopische Reversibilität (“detailed balance”)

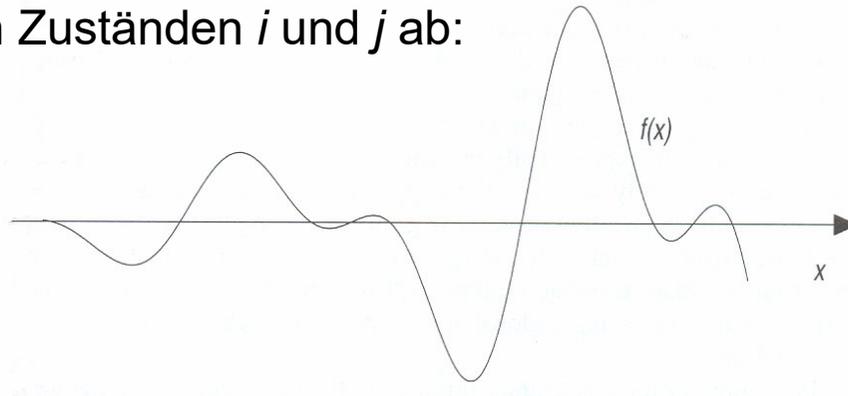
Lege Übergangsmatrix fest durch Definition einer Übergangswahrscheinlichkeit  $k_{ij}$  für jeden Übergang von  $X_i$  nach  $X_j$ , so dass die mikroskopische Umkehrbarkeit erfüllt ist:

$$\rho_i k_{ij} = \rho_j k_{ji}$$

$\rho_i$  und  $\rho_j$  sind die Besetzungswahrscheinlichkeiten der beiden Zustände  $i$  und  $j$ .

Im kanonischen Ensemble sind dies die Wahrscheinlichkeiten aus der Boltzmann-Verteilung. Damit hängt das Verhältnis der Übergangswahrscheinlichkeiten nur vom Energieunterschied zwischen den Zuständen  $i$  und  $j$  ab:

$$\frac{\rho_i}{\rho_j} = \frac{k_{ji}}{k_{ij}} \propto \exp(-\beta \Delta E_{ij})$$
$$\Delta E_{ij} = E(X_i) - E(X_j)$$



Durch die Erfüllung dieser Bedingung ist das Gleichgewicht stationär.

Es gilt die **detailed balance**.

# Ensembles (Wdh.)

Bei MD-Simulationen sorgt man zunächst dafür, daß sich das System in einem Gleichgewichtszustand befindet, bevor man mit dem Sampling beginnt:

$10^4 - 10^5$  Schritte Equilibrierung, gefolgt vom Produktionslauf mit  $10^3 - 10^5$  Stützpunkten.

Ensembles werden charakterisiert durch die Parameter die konstant sind und diejenigen die durch die Simulation berechnet werden sollen.

Monte Carlo hat üblicherweise eine konstante Anzahl Teilchen  $N$ , Volumen  $V$  und Temperatur  $T$  = kanonisches Ensemble  
Die Energie fluktuiert aufgrund des Metropolis-Algorithmus.

MD hat üblicherweise eine konstante Anzahl Teilchen  $N$ , Volumen  $V$  und Energie  $E$  = Mikrokanonisches Ensemble, wobei die Entropie  $S$  ein Maximum anstrebt. Die Energie bleibt konstant aufgrund der Newtonschen Bewegungsgleichungen.

Bei Bedarf lassen sich auch andere Szenarien generieren.

## Beziehung zwischen den Ensembles

Die Zustandssumme kann als Summe über die Zustände fester Energie dargestellt werden: 
$$Q(N, V, T) = \sum_E \exp[-\beta E] \Omega(N, V, E)$$

Die Anzahl der Zustände  $\Omega(N, V, E)$  nimmt sehr rasch mit  $E$  zu, wogegen die Besetzung dieser Zustände stark abnimmt (Boltzmann-Verteilung).

Das Produkt der beiden Funktionen hat daher ein scharfes Maximum bei einem Wert  $\bar{E}$  und das System wird meist eine Energie sehr eng bei diesem Wert haben.

→ Daher besteht in der Praxis meist kein grosser Unterschied zwischen dem kanonischen (Monte Carlo) und dem mikrokanonischen System (MD).

Aufgrund des Gesetzes der grossen Zahlen, haben messbare physikalische Grössen sehr kleine Fluktuationen.

Die Abweichungen zwischen den Ensembles werden jedoch grösser, je kleiner die Systeme werden → Vorsicht also bei kleinen Simulationssystemen!

## gebräuchliche statistische Ensembles

Das mikrokanonische Ensemble  $NVE = \text{konstant}$  (MD,  $p$  schwankt)

Das kanonische Ensemble  $NVT = \text{konstant}$  (Monte Carlo,  $E$  schwankt)

Das isotherme-isobare Ensemble  $NPT = \text{konstant}$  (MD,  $V$  schwankt)

Das grosskanonische Ensemble  $\mu VT = \text{konstant}$  (MD,  $N$  schwankt)

Das chemische Potential  $\mu$  geht in die Freie Enthalpie ein. Dies führt dazu, daß ein spontaner Übergang von einer Phase in die andere nur dann erfolgt, wenn die chemischen Potentiale dieser beiden Phasen unterschiedlich sind. Im Gleichgewichtsfall sind dann auch die chemischen Potentiale der beiden Phasen gleich.

# Anwendungen

Das mikrokanonische Ensemble  $NVE = \text{konstant}$

→ überprüfe Stabilität von Integrationsalgorithmen,  
Erzeugung von „wahrer“ Dynamik (typische MD-Simulationen)

Das kanonische Ensemble  $NVT = \text{konstant}$

→ simuliere Vorgänge unter Druckschwankungen (Monte Carlo Simulationen)

Das isotherme-isobare Ensemble  $NPT = \text{konstant}$

→ übliches Ensemble für Simulationen von Biomolekülen (MD-Simulationen)

Das grosskanonische Ensemble  $\mu VT = \text{konstant}$

→ Stuchebrukhov-Paper über Zahl der Wassermoleküle in Cytochrome c Oxidase  
(siehe folgende Seite)

# Anwendung: identifiziere Wassermoleküle in Proteinen (I)

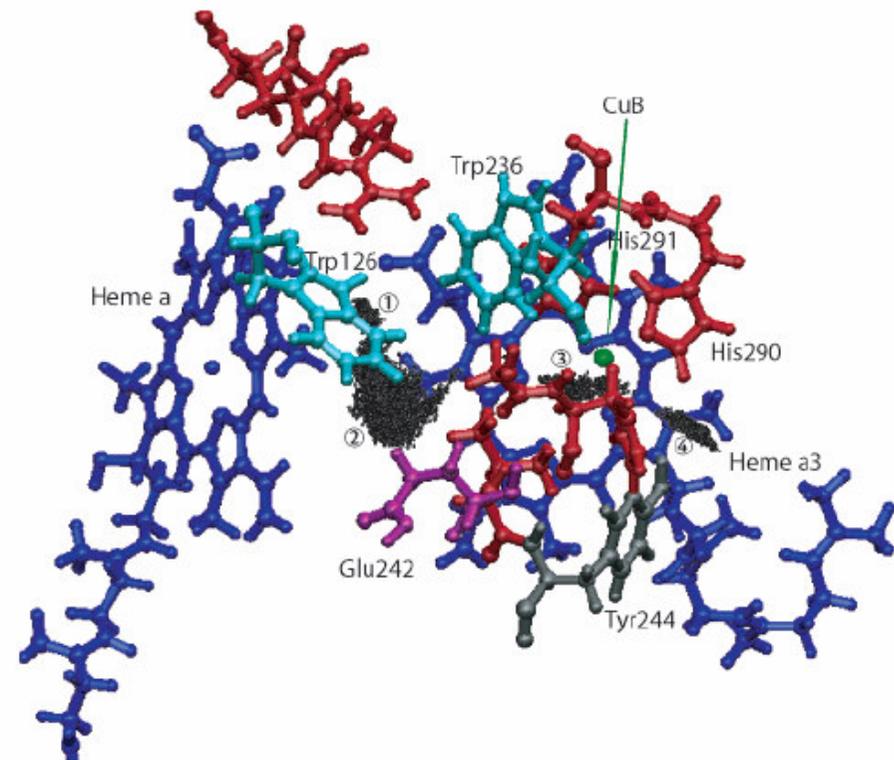
Kristallstrukturen von Proteinen zeigen oft nicht die Position interner Wassermoleküle, da diese zu mobil sind (B-Faktor zu hoch, obwohl Elektronendichte vorhanden ist).

→ Verwende Computersimulationen um die „wahre“ Hydratation zu finden.

Welches ist das beste Ensemble um Wassermoleküle in ein Protein zu positionieren?

Das gross-kanonische Ensemble, da hier das chemische Potential  $\mu$  konstant ist.

Ref. Tashiro, Stuchebrukhov, *J. Phys. Chem. B* **109** (2005) 1015.



**Figure 1.** Structure of CcO around the enzyme catalytic center. Only the important residues are shown. The key elements are heme  $a$ , heme  $a_3$ , the  $\text{Cu}_B$  center, Glu242, His291, and the propionates of heme  $a_3$ . The hydrophobic cavity discussed in the text extends from Glu242 to the  $\Delta$ -propionate of heme  $a_3$  and from Glu242 to the Fe–Cu BNC. The black points show the favorable position of a water molecule in the region.

## Anwendung: identifiziere Wassermoleküle in Proteinen (II)

Betrachte Transfer aus dem Lösungsmittel Wasser ins Proteininnere.

In Lösung besitzt ein Wassermolekül eine freie Lösungsenthalpie von -12.3 kcal/mol.

Wenn es im Proteininneren eine günstigere Position findet, dann wird diese im zeitlichen Mittel mit einem Wasser besetzt sein.

Führe Simulation im semi-grosskanonischen Ensemble durch, wobei während der Simulation Wassermoleküle in das Protein hinzugefügt bzw. daraus entfernt werden.

In diesem Fall ist es gerade so günstig, die Kavität mit 2 Wassermolekülen zu füllen.

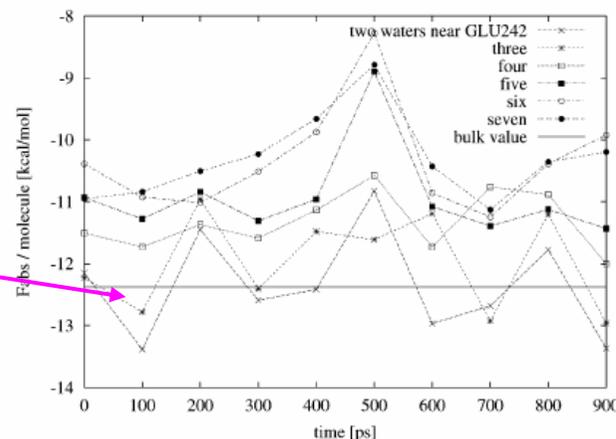


Figure 6. Absolute free energies of water clusters in the hydrophobic cavity around the BNC. The free energies are shown per one molecule in the cluster. For comparison, the corresponding free energy in bulk water (TIP3P water, ref 44) is also shown in the figure.

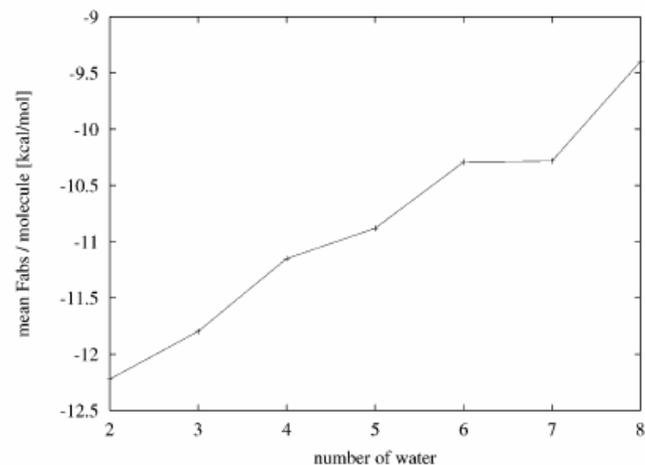


Figure 7. Mean free energy per one water molecule of water clusters of different sizes in the BNC cavity.

## Wdh: Moleküldynamik-Simulation

Basiert auf **Newtonscher Bewegungsgleichung** für ein Atom  $i$  eines Moleküls:

$$\mathbf{F}_i = m_i \frac{\partial^2 \mathbf{r}}{\partial t^2}$$

wobei  $\mathbf{F}_i$  die Kraft,  $m_i$  seine Masse und  $\frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{a}_i$  die Beschleunigung ist, die auf Atom  $i$  wirkt.

Die Kräfte lassen sich aus den Ableitungen der Energie nach den kartesischen Koordinaten ausrechnen.

- Die **Trajektorie** eines Systems ist die Aneinanderreihung der einzelnen Koordinaten und Geschwindigkeiten, also ein Film, der die Dynamik des Systems zeigt.

# Zusammenfassung Sampling

Molekülmechanik-Kraftfelder liefern eine sehr effiziente (aber grobe) Darstellung für die relativen Energien von Molekülen im Konformationsraum.

Das Absuchen des Konformationsraums kann systematisch oder stochastisch (Monte Carlo) geschehen.

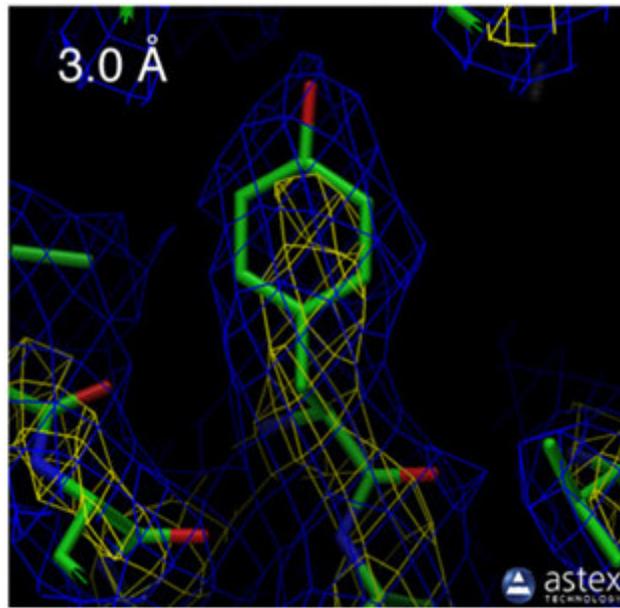
Die Erfüllung der *detailed balance*-Bedingung ist hinreichend und notwendig für MC- und MD-Simulationen im Gleichgewichtszustand.

In Simulationen können Zwangsbedingungen angewendet werden um das Sampling an bestimmten Stellen zu verbessern bzw. in die gewünschte Richtung zu lenken (umbrella potential).

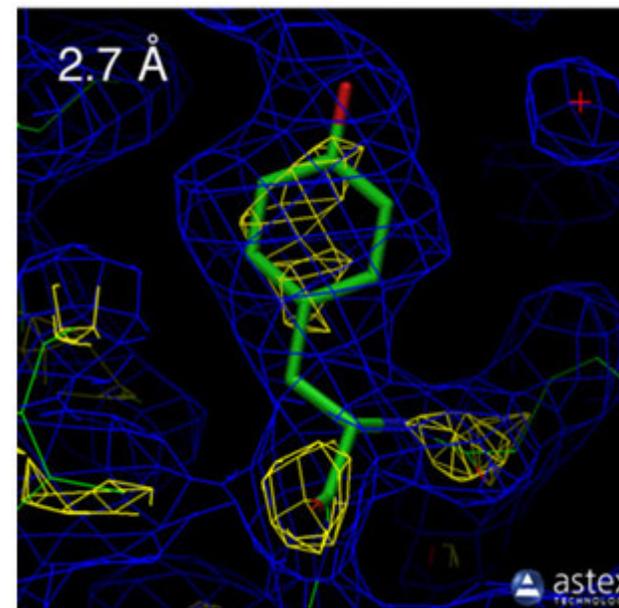
# Proteinstrukturen (1)

Da Kristallstrukturen von Proteinen (typischerweise) keine H-Atome enthalten (**warum?**) müssen diese vor der Verwendung in Kraftfeldern oder beim Docking hinzugefügt werden.

Die Auflösung (resolution) gibt die kürzeste Wellenlänge (in Å) an, die der Kristall noch streute. → Man kann also Objekte bis zu dieser Größe erkennen.



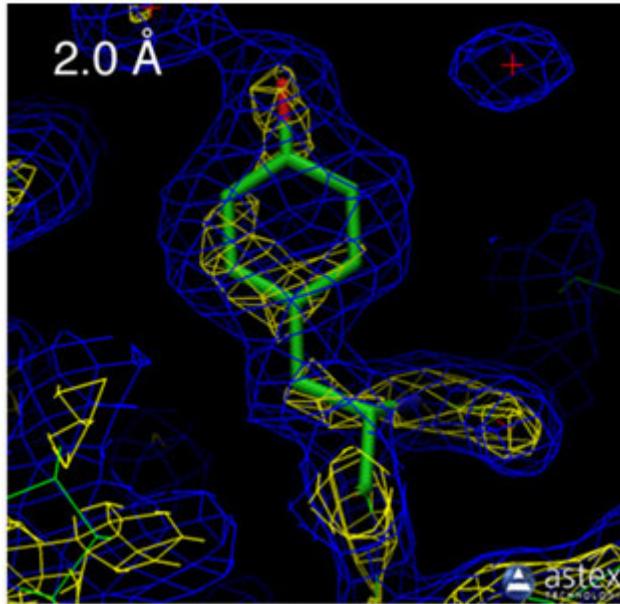
Backbone und mutmaßliche Positionen der Seitenketten



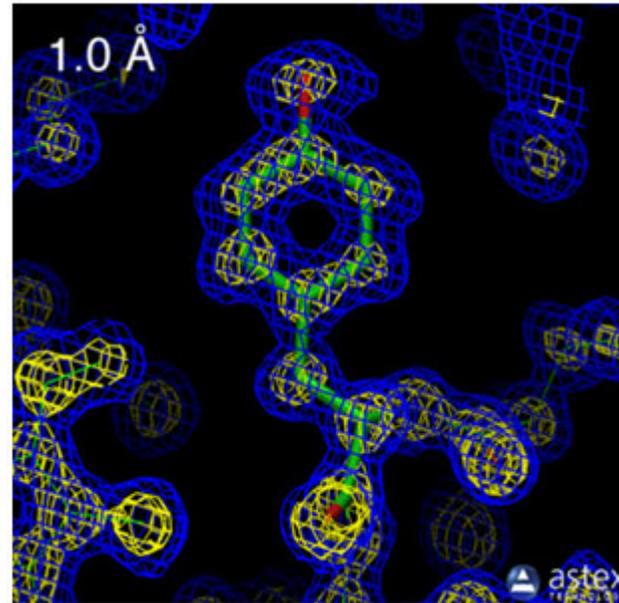
Unterscheidung zwischen grossen und kleinen Seitenketten wird möglich

## Proteinstrukturen (2)

Die in einer pdb-Datei angegebenen Atomkoordinaten sind das Ergebnis einer Kraftfeldoptimierung mit der Elektronendichte als Randbedingung.



Unterschiedliche Rotamere der Seitenketten können zugeordnet werden



Atomkoordinaten können zweifelsfrei zugeordnet werden.  
Ausbuchtungen der Elektronendichte lassen H-Atome erahnen.

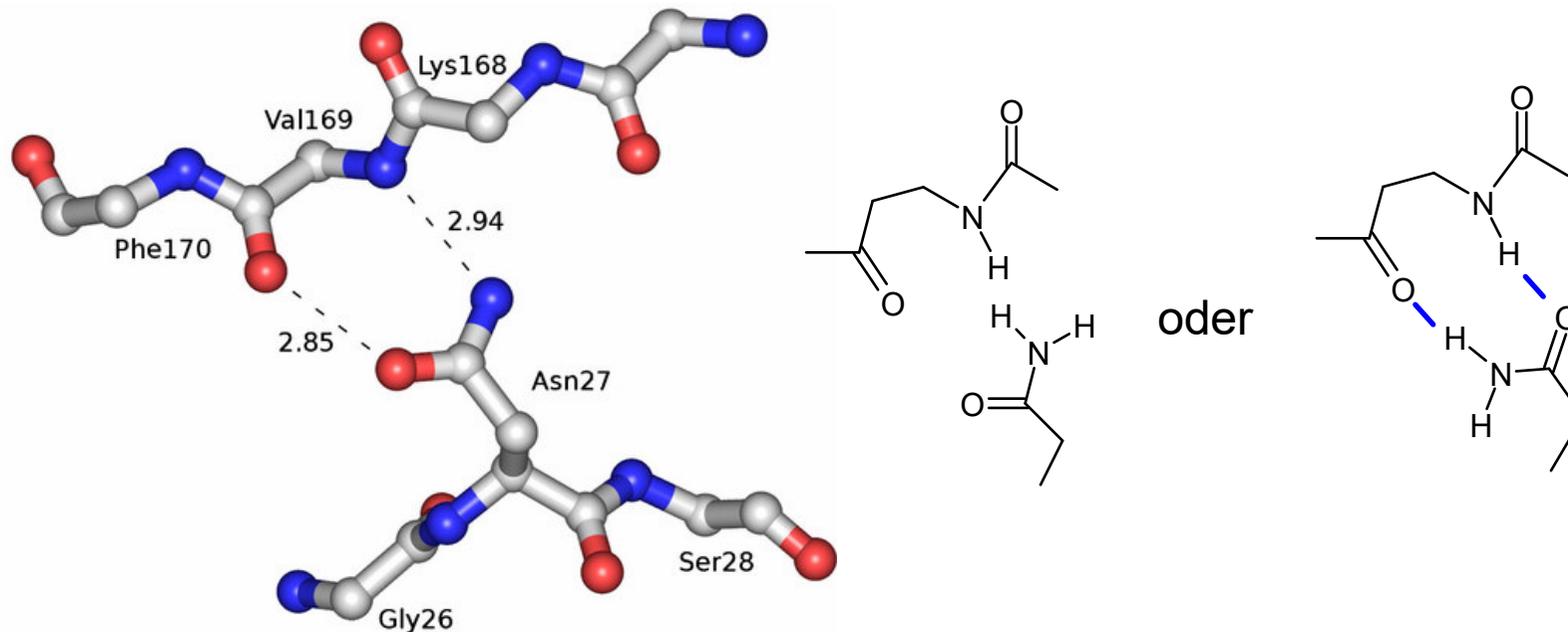
Elektronendichte:

gelb:  $1 \sigma$

blau:  $3 \sigma$

# Rotamerzuordnung von Asn und Gln

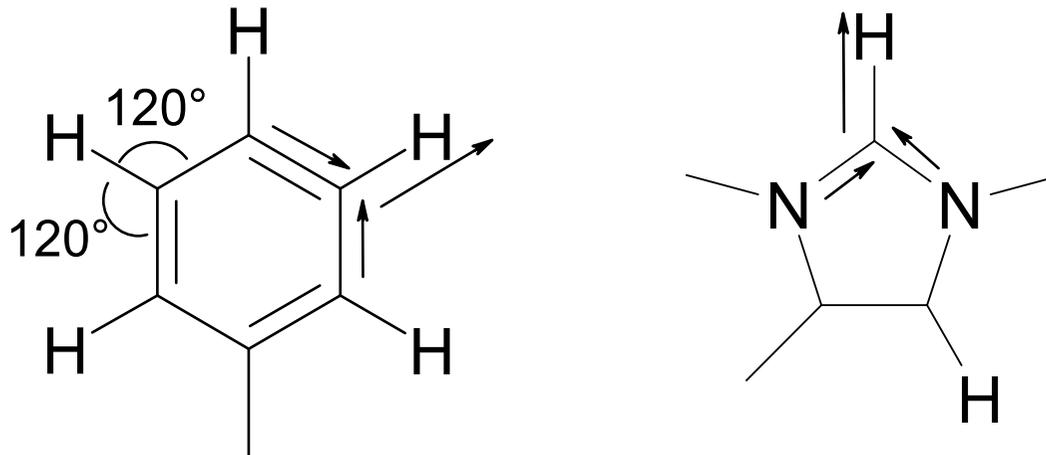
Da Sauerstoff und Stickstoff sehr ähnliche Streuquerschnitte aufweisen, ist die Zuordnung der passenden Rotamere von Asn (N) und Gln (Q) nicht ohne weiteres möglich:



Erst durch Überprüfung der H-Brücken zeigt sich welches Rotamer energetisch besser ist (NQ-flip um  $180^\circ$ ).

## apolare H-Atome (1)

Die Positionen von H-Atomen die Kohlenstoff gebunden sind lassen sich für die üblichen Aminosäuren geometrisch relativ einfach erzeugen.

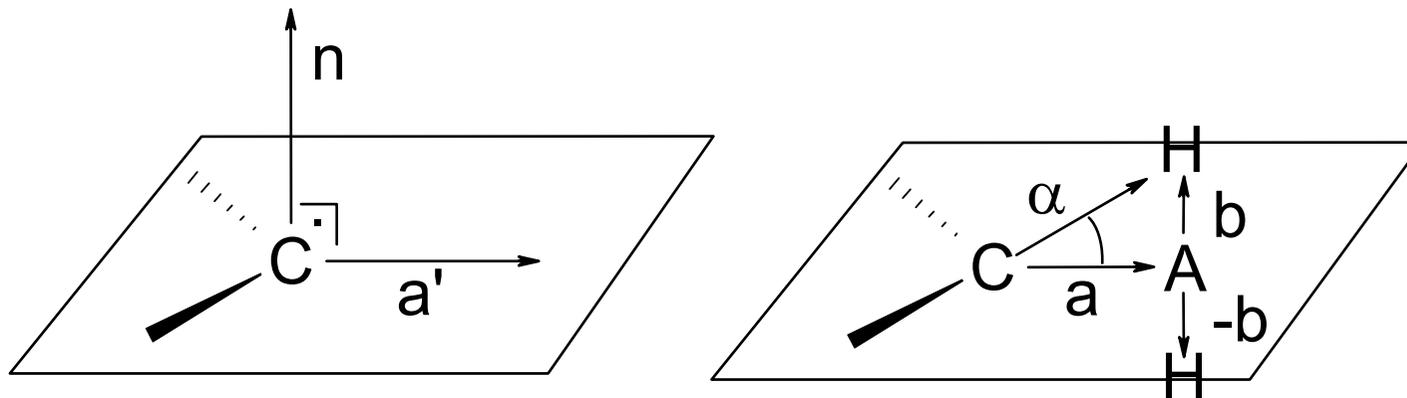


Aufgrund der Ungenauigkeiten der Atomkoordinaten hat man meistens nicht die idealen Winkel vorliegen. Stattdessen kann man die Position der H-Atome durch Addition der Vektoren und skalieren auf die C-H Bindungslänge konstruieren.

Genauso erhält man die polaren H-Atome an den Stickstoffen der Peptidbindung und in der Seitenkette von Arginin.

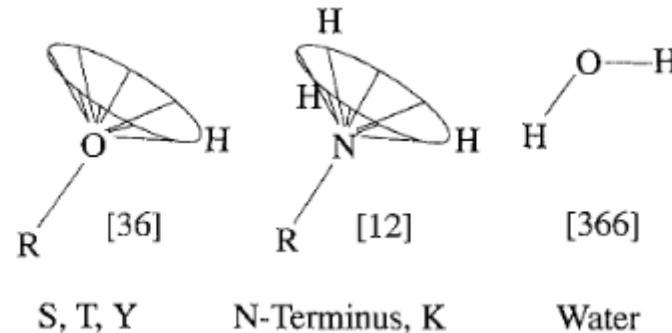
## apolare H-Atome (2)

Bei nicht-planaren Geometrien ist die Konstruktion etwas aufwendiger.  
Zunächst wird der Hilfsvektor  $a'$  durch Addition der beiden Bindungsvektoren erhalten. Dann erzeugt man den Normalenvektor  $n$  auf der Ebene der Atome.  
Länge C-A =  $1.09 \cos \alpha$  mit  $\alpha = 109.5/2$  (1.09 ist die C-H Bindungslänge)  
Länge A-H =  $1.09 \sin \alpha$   
Der Vektor  $a$  wird durch skalieren von  $a'$  auf die Länge C-A erhalten.  
Translatieren von  $n$  nach A und skalieren auf die Länge A-H erzeugt die Position der beiden gesuchten H-Atomen aus den Vektoren  $b$  und  $-b$ .



## Polare H-Atome (1)

Bei Serin, Threonin und Tyrosin beschreibt die Rotation der R–OH Bindung einen Kegel. Ähnlich bei  $\text{NH}_3^+$  von Lysin. Im Falle von Wasser liegen die beiden H-Atome auf einer Kugeloberfläche.

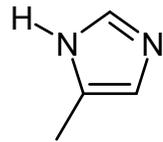


Das Programm WHATIF diskretisiert diese Positionen in  $10^\circ$  Schritten. Gleichverteilte Positionen auf einer Kugeloberfläche werden aus einer Fibonacci-Kugel erzeugt. (was allerdings nicht zu geometrisch korrekten Wassergeometrien führt)

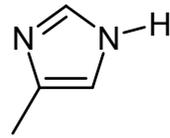
Lit. WHATIF Hooft, Sander & Vriend, *Proteins* **26** (1996) 363-376.

## Polare H-Atome (2)

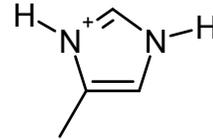
Für Histidin gibt es drei verschiedene Protonierungszustände: HID, HIE und HIS die wiederum jeweils um  $180^\circ$  rotiert werden können (= 6 Positionen)



HID



HIE



HIS

Asp und Glu sind in aller Regel deprotoniert; Arg und Lys typischerweise protoniert. Ausnahme: in apolaren Medien (membranständige Bereiche von Proteinen)

Da die Vielzahl der dadurch möglichen Zustände zu einer kombinatorischen Explosion führt, setzt WHATIF auf eine *divide and conquer* Strategie: Aufteilen auf überschaubar lösbare Subnetzwerke von H-Brücken die mittels simulated annealing optimiert werden.

WHATIF Webserver: <https://swift.cmbi.umcn.nl/servers/html/index.html>

## Polare H-Atome (3)

Weitere Programme die polare H-Atome zuordnen:

ProToss Lippert & Rarey *J. Cheminform.* **1** (2009) 13.

Die denkbaren Protonierungszustände werden als Graph erzeugt der in einen Baum transformiert wird. Ausgehend von dessen Wurzel werden die Unterbäume iterativ gelöst. Webserver:

<https://www.zbh.uni-hamburg.de/forschung/amd/software/protoss.html>

Propka3 bzw. pdb2pqr

Søndergaard et al. *J. Chem. Theo. Comput.* **7** (2011) 2284-2295.

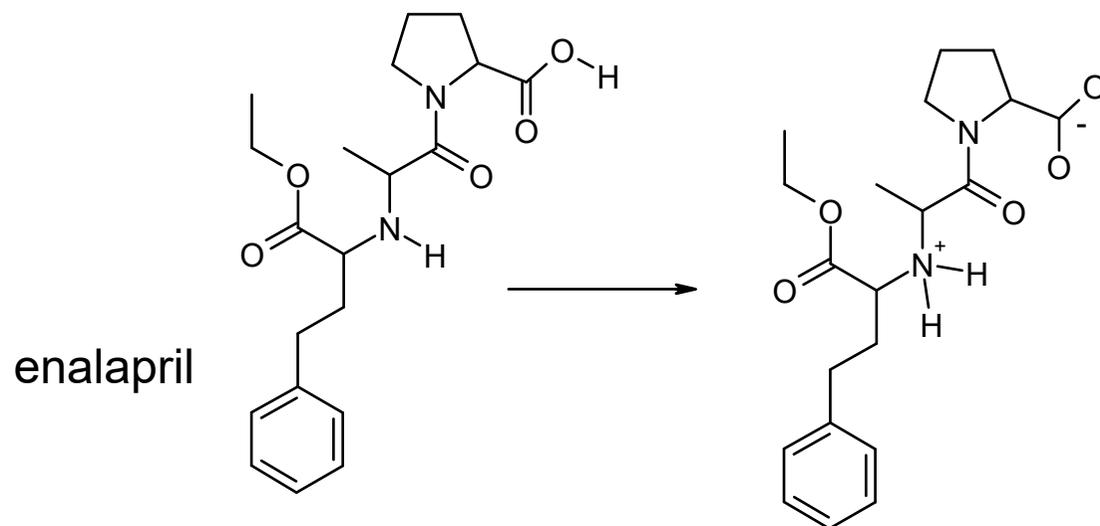
Berücksichtigt Änderungen des Protonierungszustandes von Aminosäuren durch H-Brücken mit gebundenen Liganden (falls vorhanden), sowie den pH-Wert des Lösungsmittels.

Webserver:

<https://server.poissonboltzmann.org/pdb2pqr>

## Polare H-Atome in Liganden

Auch bestimmte chemische Gruppen in Liganden verändern ihren Protonierungszustand in Abhängigkeit von umgebendem pH-Wert und der Möglichkeit H-Brücken (mit der Rezeptor) auszubilden. Allgemein gilt: Säuren werden deprotoniert, Basen protoniert



Mit dem Programm open babel lässt sich der Protonierungszustand für entsprechende pH-Werte zuordnen:

```
babel -ipdb ligand.pdb -p 7.4 -opdb ligand_with_Hs.pdb
```