

# Vorlesung 12205: Einführung ins Molecular Modeling

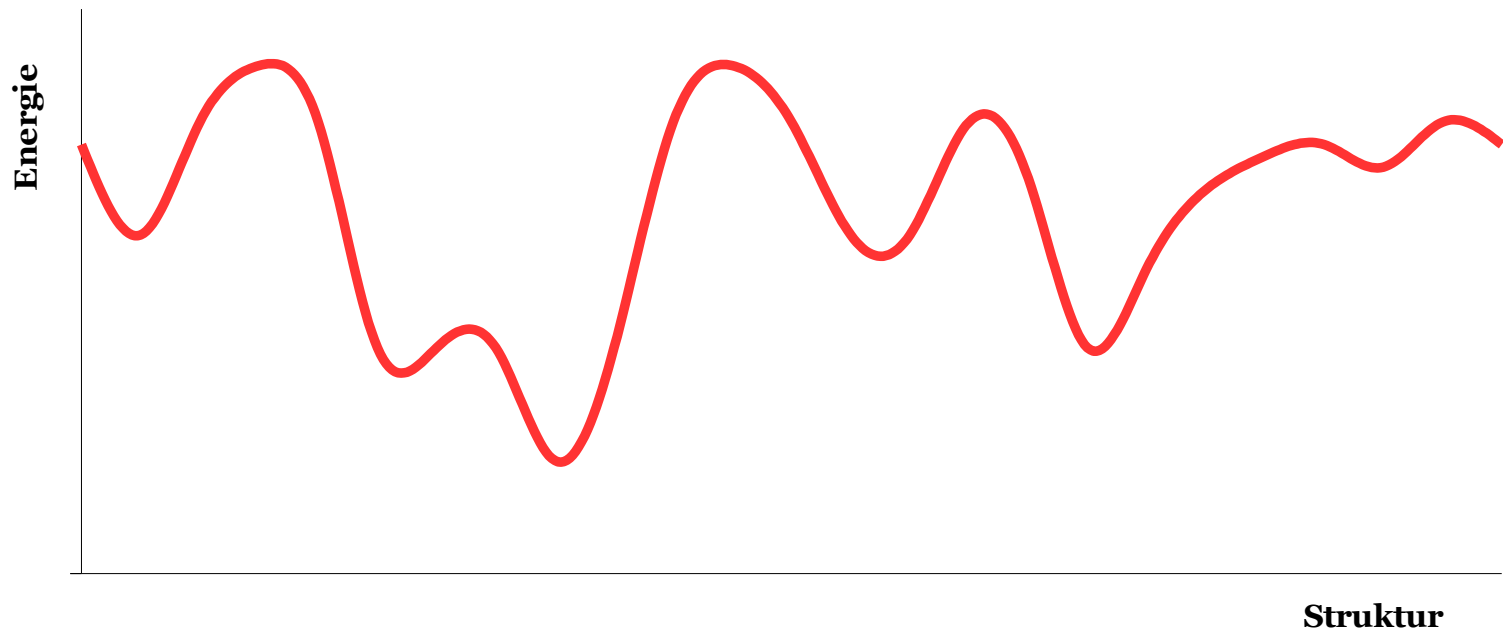
M. Smieško & A. Vedani — Departement Pharmazeutische Wissenschaften, Universität Basel, HS-2017



# Strukturoptimierung / Energieminimierung

*(Structure optimization / Energy minimization)*

- biologische Prozesse (z.B. auch Arzneistoffbindung am Target) folgen normalerweise den energetisch günstigsten Pfad → das rechnerische Modell sollte diese Strategie kopieren
- Verfeinerung von rohen strukturellen Daten (X-ray, NMR, Homologiemodellierung)



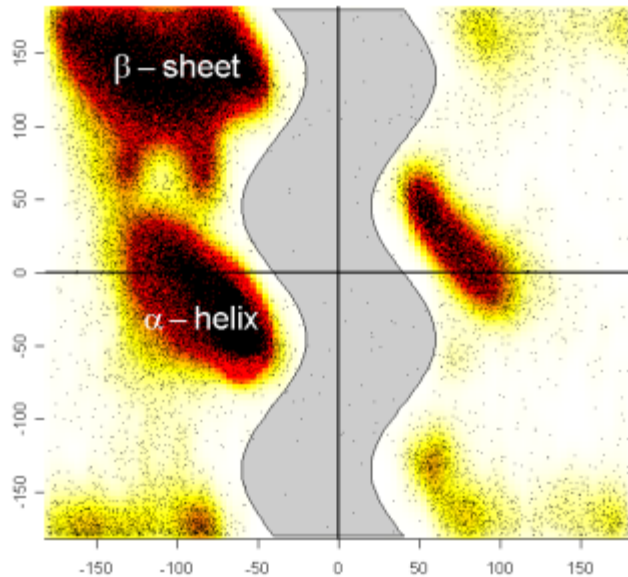


# Strukturoptimierung / Energieminimierung

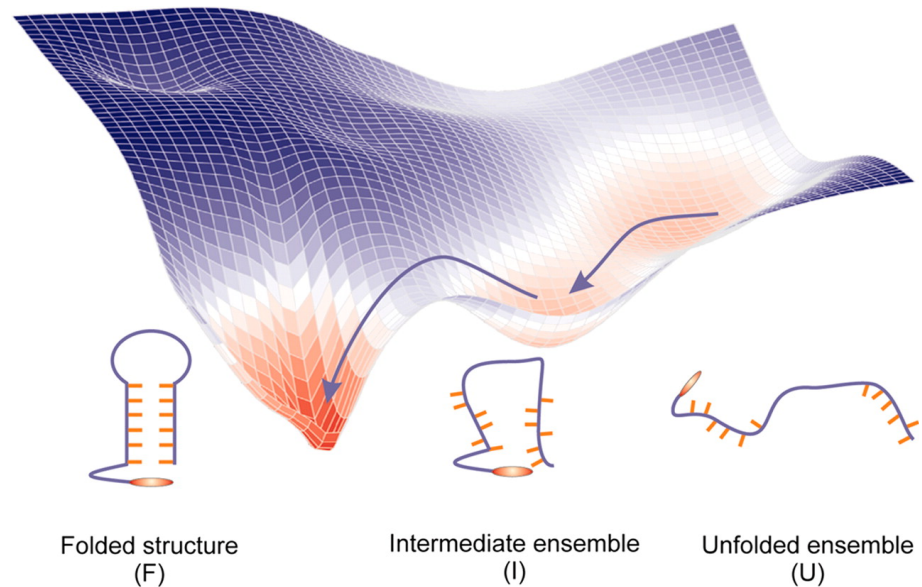
(Structure optimization / Energy minimization)

- biologische Prozesse (z.B. auch Arzneistoffbindung am Target) folgen normalerweise den energetisch günstigsten Pfad → das rechnerische Modell sollte diese Strategie kopieren
- Verfeinerung von rohen strukturellen Daten (X-ray, NMR, Homologiemodellierung)

*Ramachandran Plot (2D)*



*Potentialoberfläche (3D, Energy landscape)*

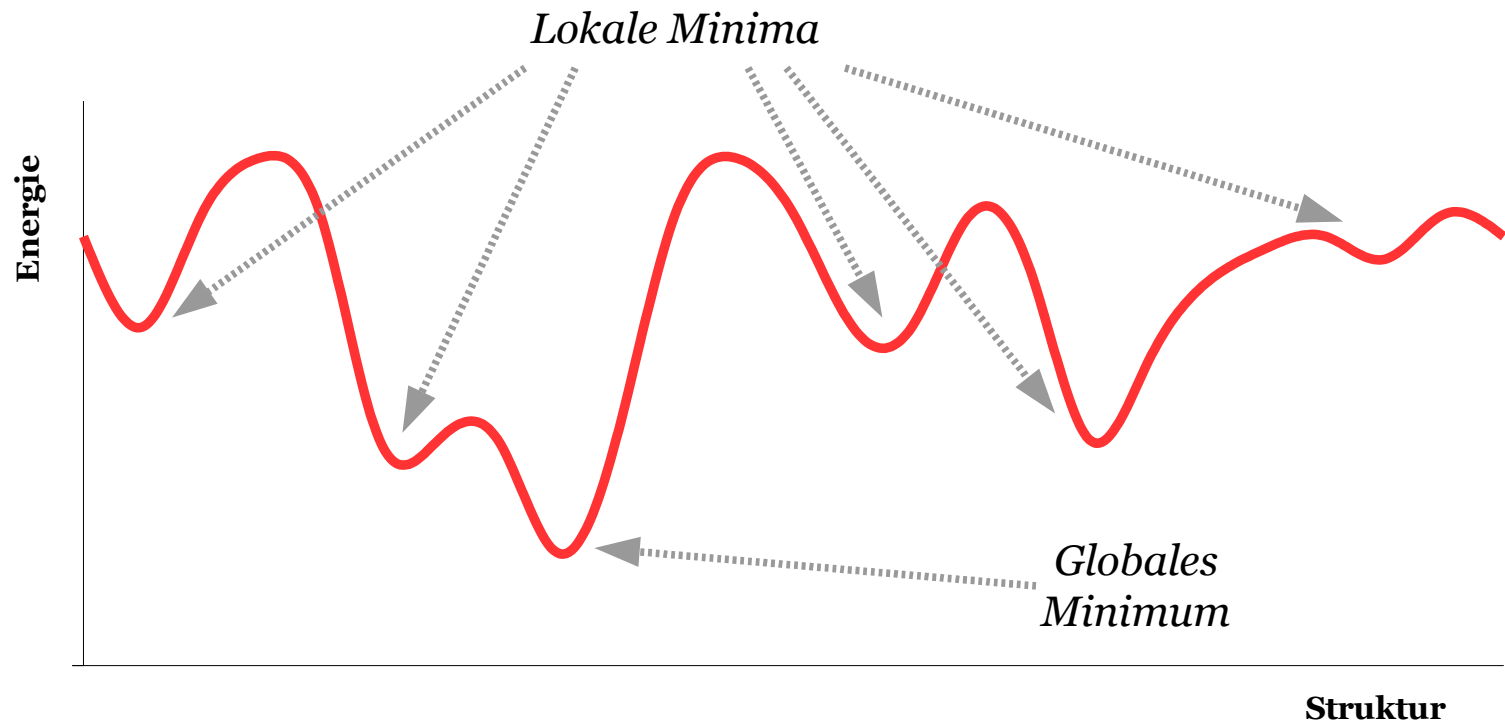




# Strukturoptimierung / Energieminimierung

*(Structure optimization / Energy minimization)*

- biologische Prozesse (z.B. auch Arzneistoffbindung am Target) folgen normalerweise den energetisch günstigsten Pfad → das rechnerische Modell sollte diese Strategie kopieren
- Verfeinerung von rohen strukturellen Daten (X-ray, NMR, Homologiemodellierung)



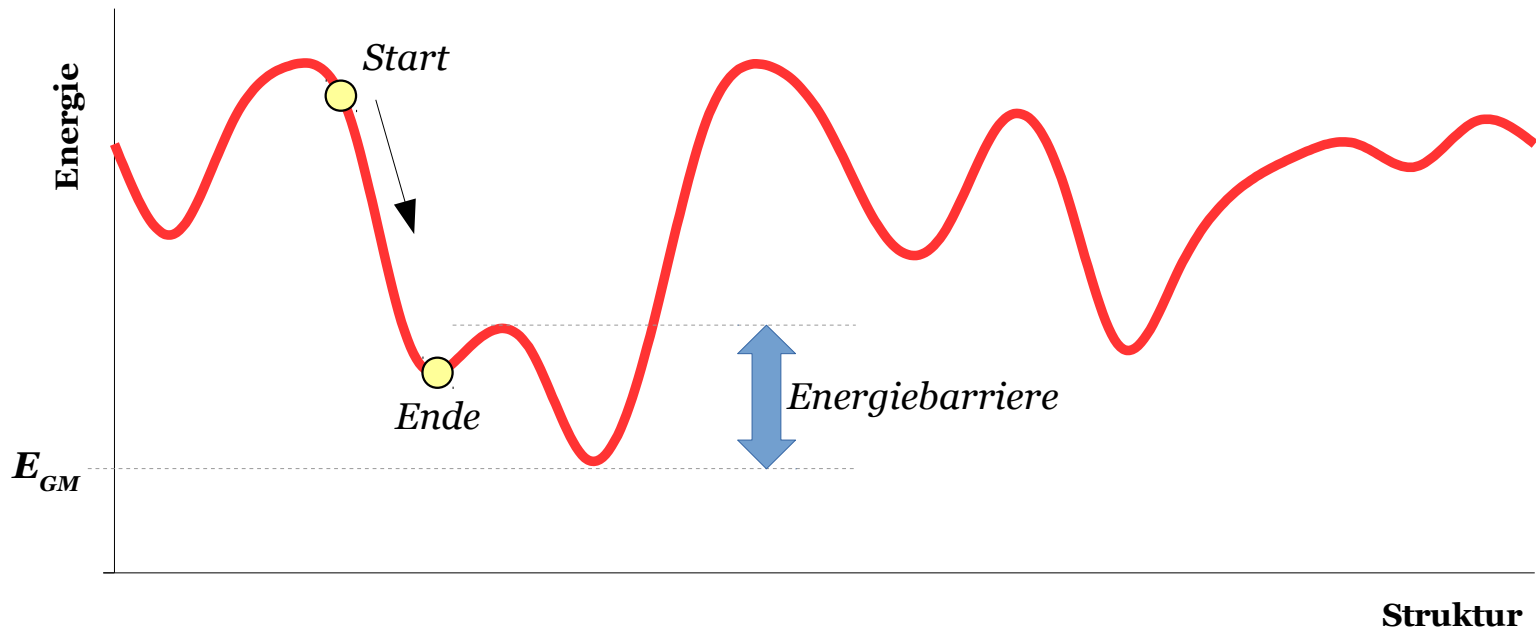




# Strukturoptimierung / Energieminimierung

(Structure optimization / Energy minimization)

Die 1. Ableitung der Kraftfeldgleichung zeigt die Richtung der Minimierung an  
(Gradient Minimierung)

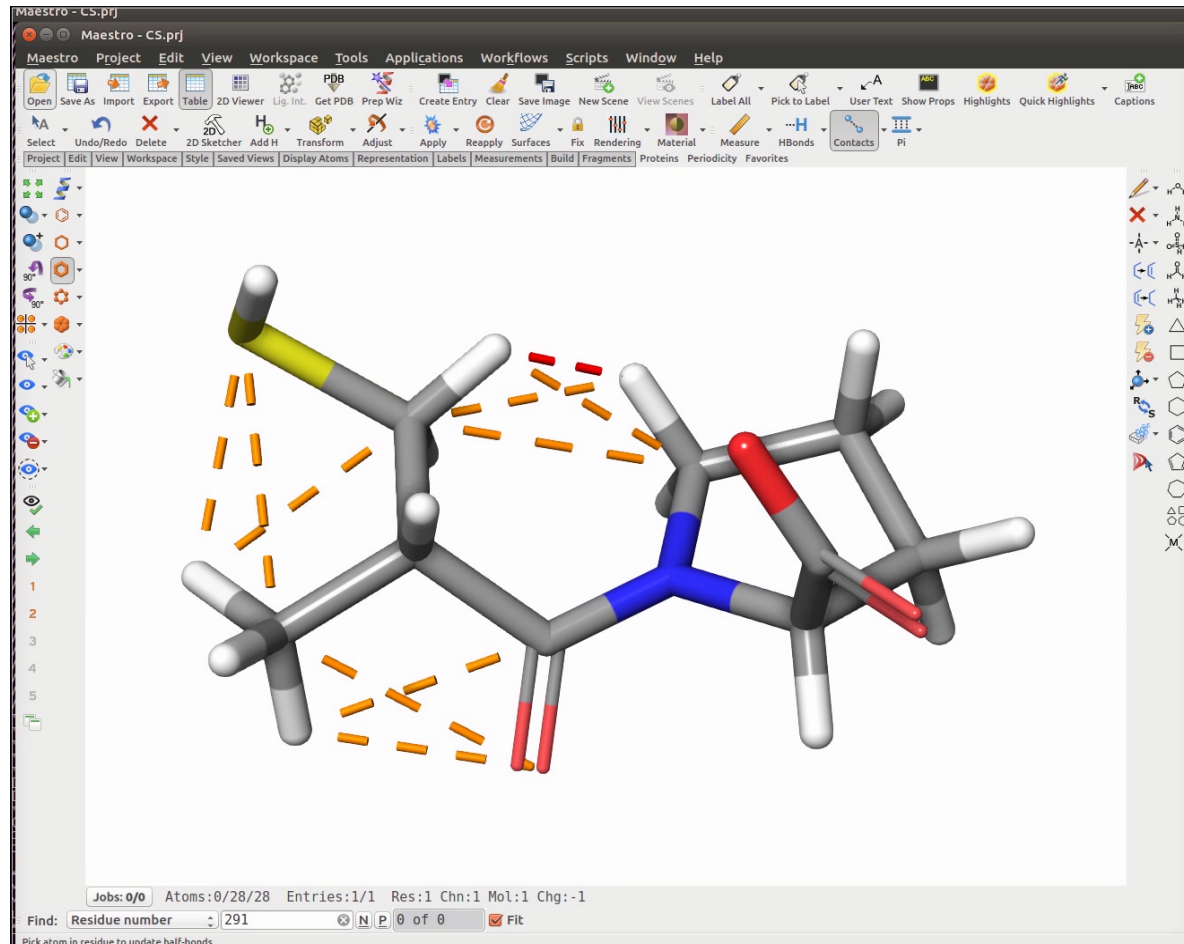


Molekülmechanik-Optimierungen enden immer im nächsten lokalen Minimum!



# Strukturoptimierung / Energieminimierung

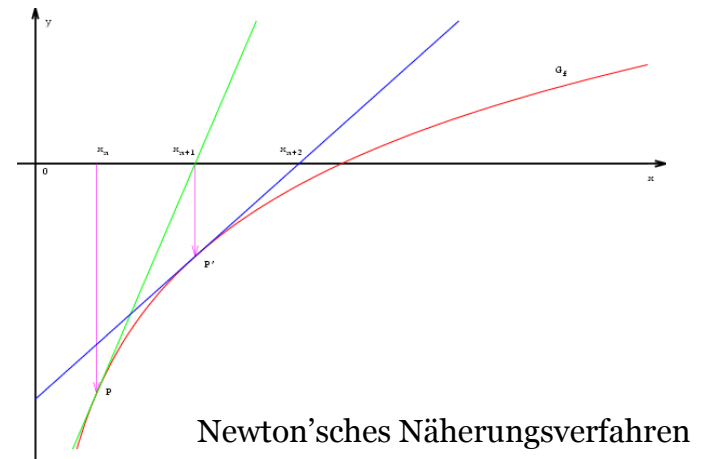
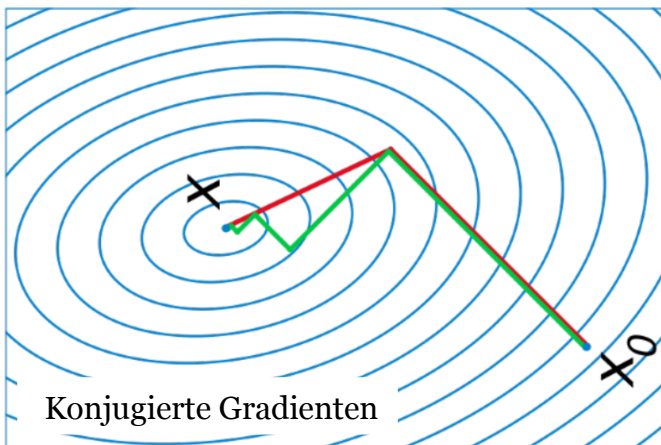
(Structure optimization / Energy minimization)





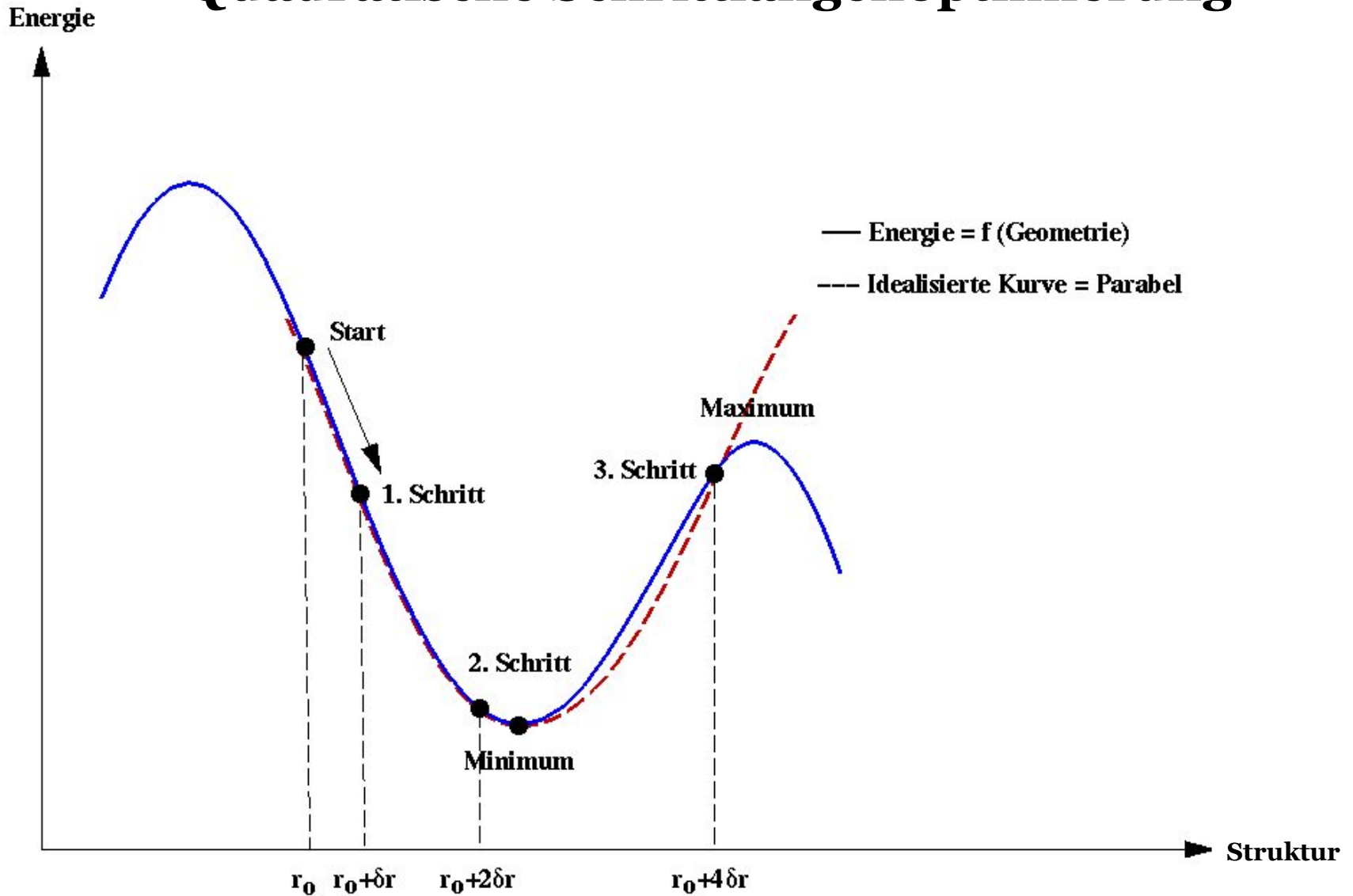
# Gradienten

- Die Methode des “steilsten Abstiegs” (*steepest descent*) führt immer dem aktuellen Gradienten (der 1. Ableitung) entlang und wird dem Minimum **nur infinitesimal nahe kommen**, es aber nie exakt erreichen.
- Konjugierte Gradienten (*conjugate gradient*) Verfahren haben ein viel besseres Konvergenzverhalten und erreichen das Minimum (Geometrie niedrigster Energie) bei  $n$  Variablen spätestens in  $n$  Schritten (Ligand–Protein Komplex im expliziten Wasser: ca. 25,000 Atome  $\rightarrow$  75,000 Variablen/Schritte!)
- Newton-Verfahren werden auch die 2. Ableitungen berechnet, wo insbesondere die Kreuzterme – wie reagiert eine Variable, wenn gleichzeitig noch eine andere verändert wird – von Bedeutung sind. Solche Verfahren werden aus Gründen der Rechenzeit nur bei kleinen und mittleren Molekülen (bis 250 Atome) angewendet (eine  $n \times n$  Matrix invertiert werden muss  $\rightarrow$  zu aufwändig bei Proteinen); Newton-Verfahren geben auch Auskunft über Maxima und Sattel- und Wendepunkte





# Quadratische Schrittlängenoptimierung





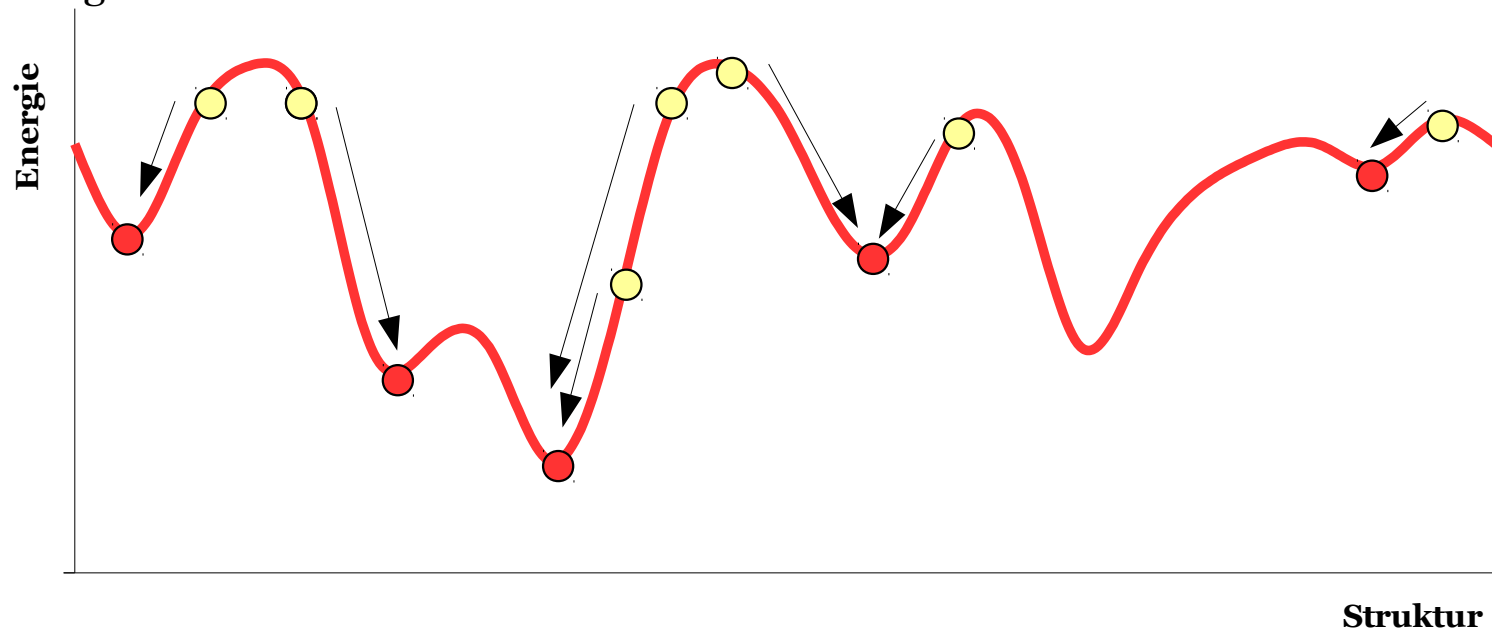
# Konformationsuche

**Zweck:** Auffinden von energetisch günstigen Zuständen (→ globales Minimum und lokale Minima) eines Moleküls oder eines kleineren Systems

**Ablauf:**

- 1) Generieren von der Startkonformation (gelb; zufällig, systematisch, rationell...)
- 2) Strukturoptimierung zum nächsten Minimum (rot; Gradient Methoden)

Die Zeitfolge ist nicht wichtig, nur ob das globale Minimum und alle relevante lokale Minima gefunden wurden

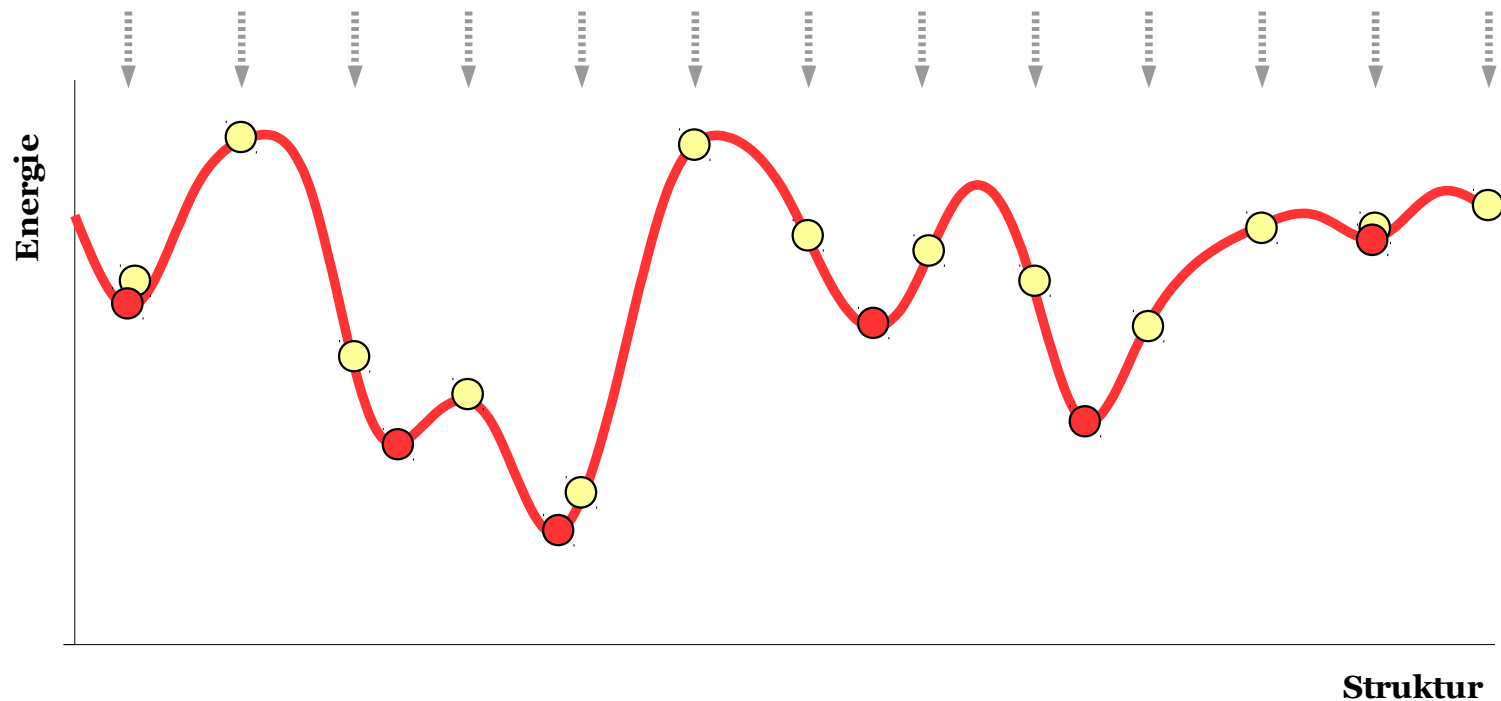




## Systematische Konformationsuche

- fähig alle Minima (globales + alle lokale) zu finden
- alle Freiheitsgrade  $n$  (Torsionswinkeln  $\rightarrow$  frei-drehbare Bindungen) werden systematisch variiert (Kleinmoleküle:  $n < 10$ ; Proteine  $n > 500$  !!!  $\rightarrow n/a$ )
- nur für kleine Systeme: z.B. Molekül mit 5 frei-drehbaren Bindungen, Schrittweite  $30^\circ$ ,  $N = (360^\circ/30^\circ)^5 = 12^5 = 248\ 832$  Konformere

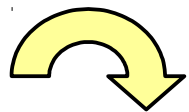
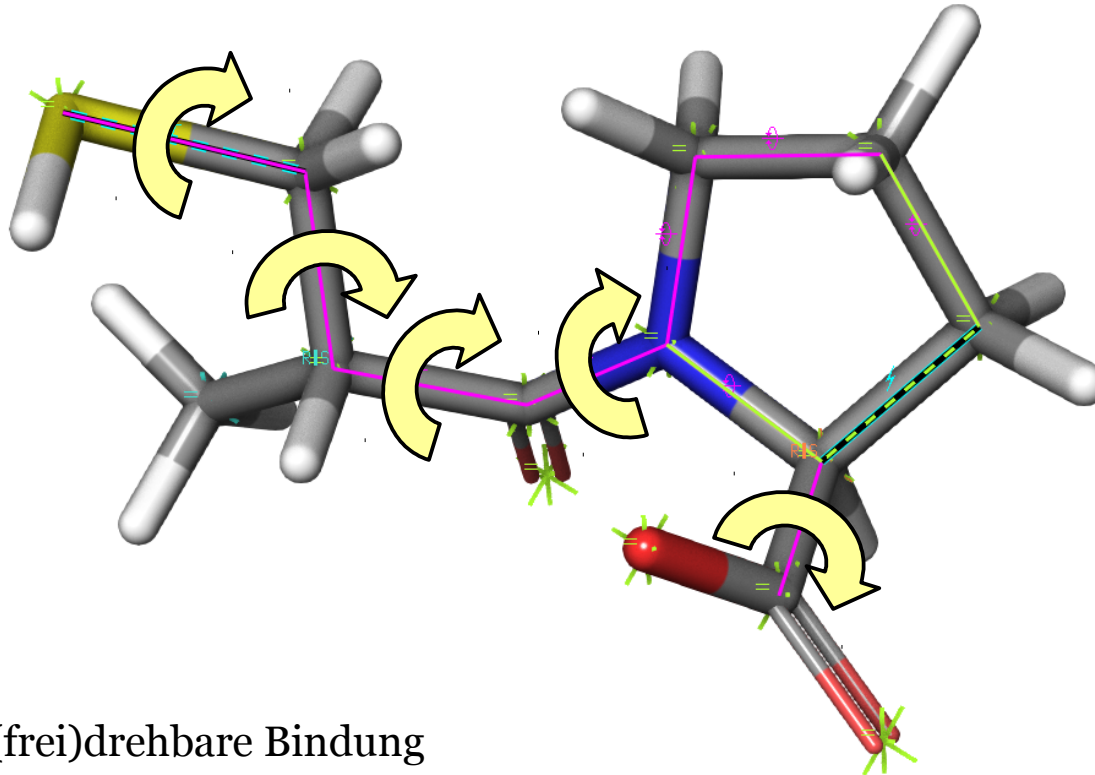
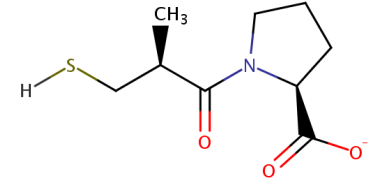
$$N_{Konf.} = (360^\circ/\text{Schrittweite})^n$$







## Einfache Konformationsuche

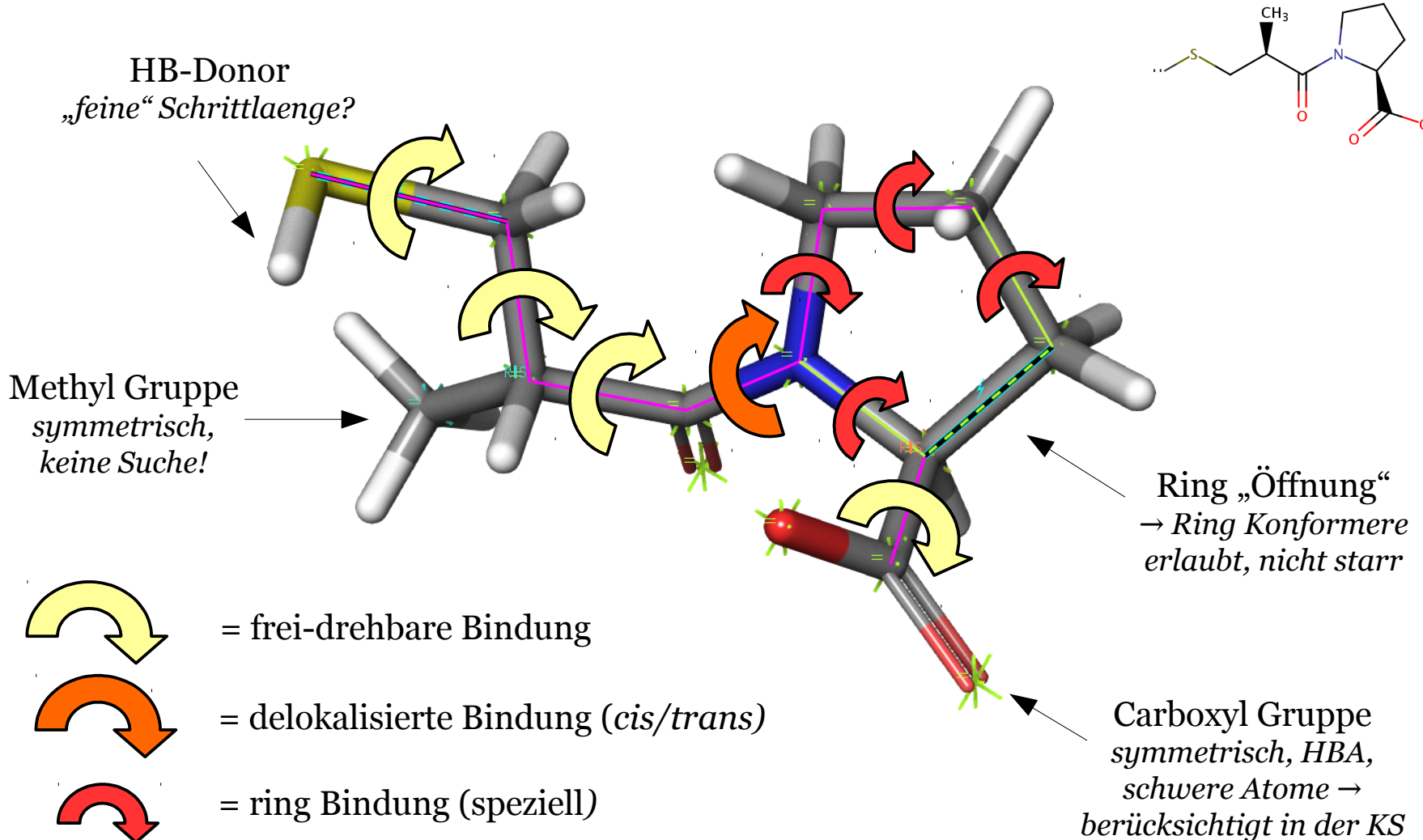


= (frei)drehbare Bindung

$$N_{Konf.} = (360^\circ/30^\circ)^5 = 12 \cdot 12 \cdot 12 \cdot 12 \cdot 12 = 248\ 832$$

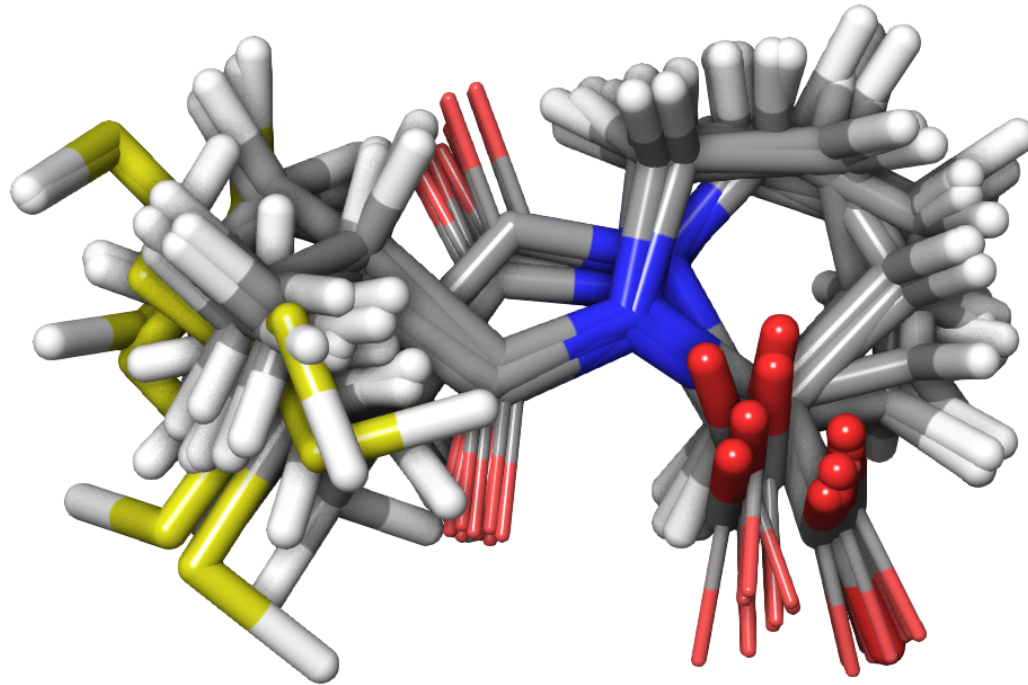
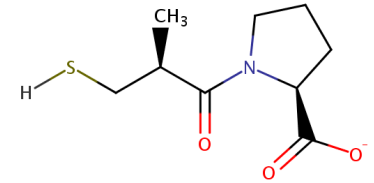


# Fortgeschrittene Konformationsuche





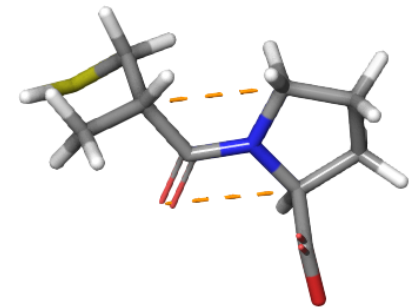
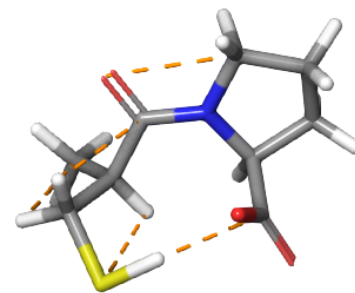
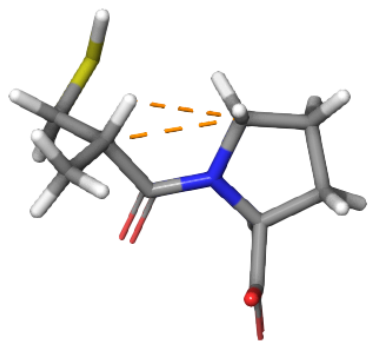
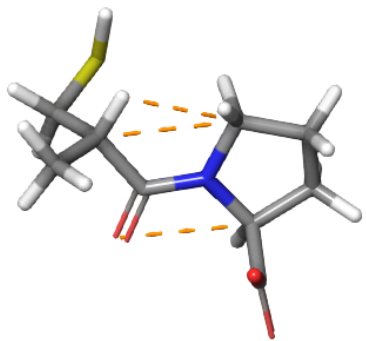
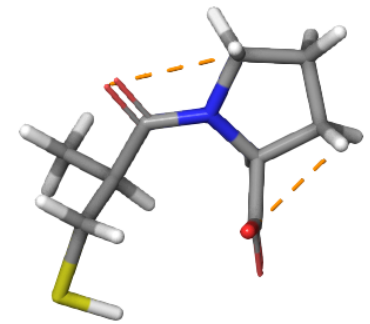
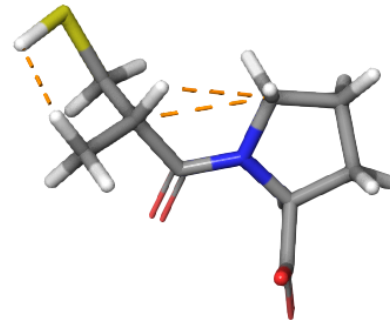
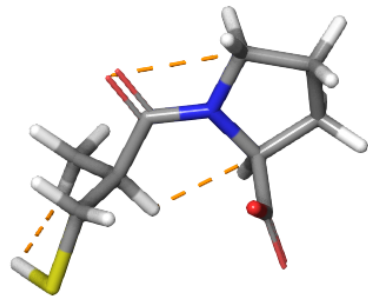
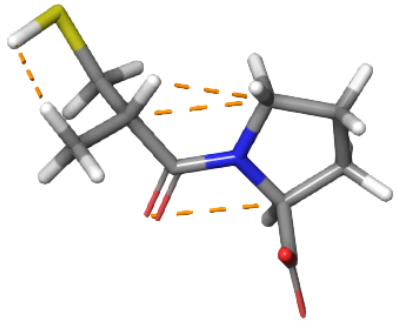
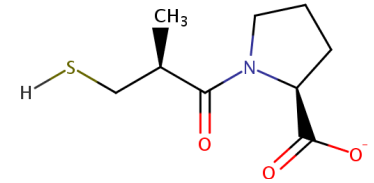
# Konformationsuche



Überlagerung (RMSD, schwere Atome) von energetisch günstigen Konformeren des *Captoprils*



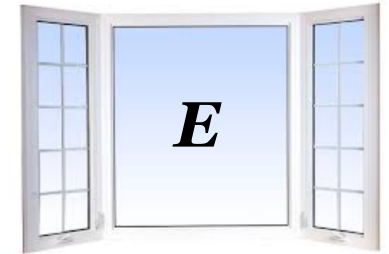
# Konformationsuche



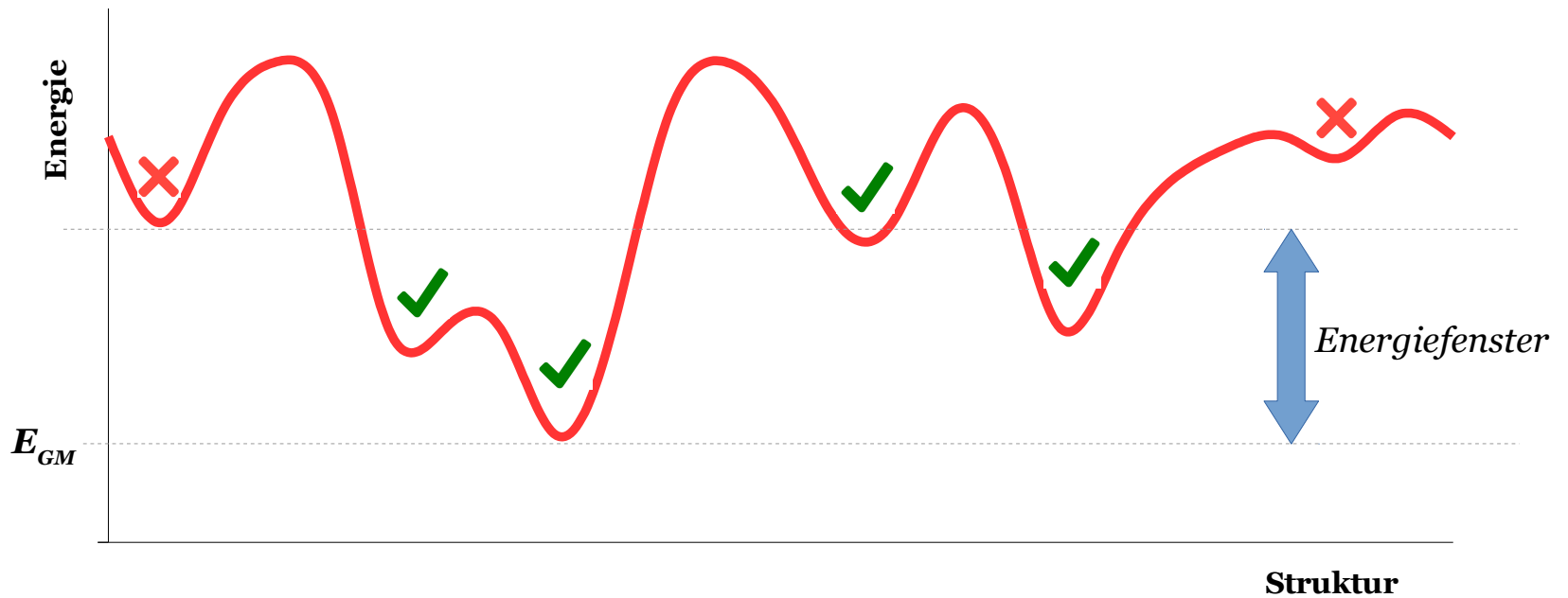
Energetisch günstige Konformere des *Captoprils*



# Energiefenster (energy window)



- relativ zum globalen Minimum
- Typische Grösse (Kleinmoleküle): 5-10 kcal/mol
- filtert Konformere mit hohen E vom finalen Konformersatz





## Was ist Molekulares Docking?

- **Auffinden** von der bevorzugten Orientierung und Konformation – der **bioaktiven Pose** – eines (typischerweise kleinen) **Moleküls innerhalb von der Bindungsstelle** eines anderen (Makro)Moleküls mittels einer **Computer Simulation**
- folgt das **Schloss-Schlüssel** Konzept (Komplementarität der wechselwirkenden Moleküle); heute beinhaltet auch *Induced Fit* → **Hand-und-Handschuhe Konzept**
- ist **von fundamentaler Bedeutung für den rationalen Arzneistoffdesign** (Virtual Screening) und Voraussage von *Off-Target* Bindung (Nebenwirkungen, Protein-vermittelte Toxizität)

### Terminologie

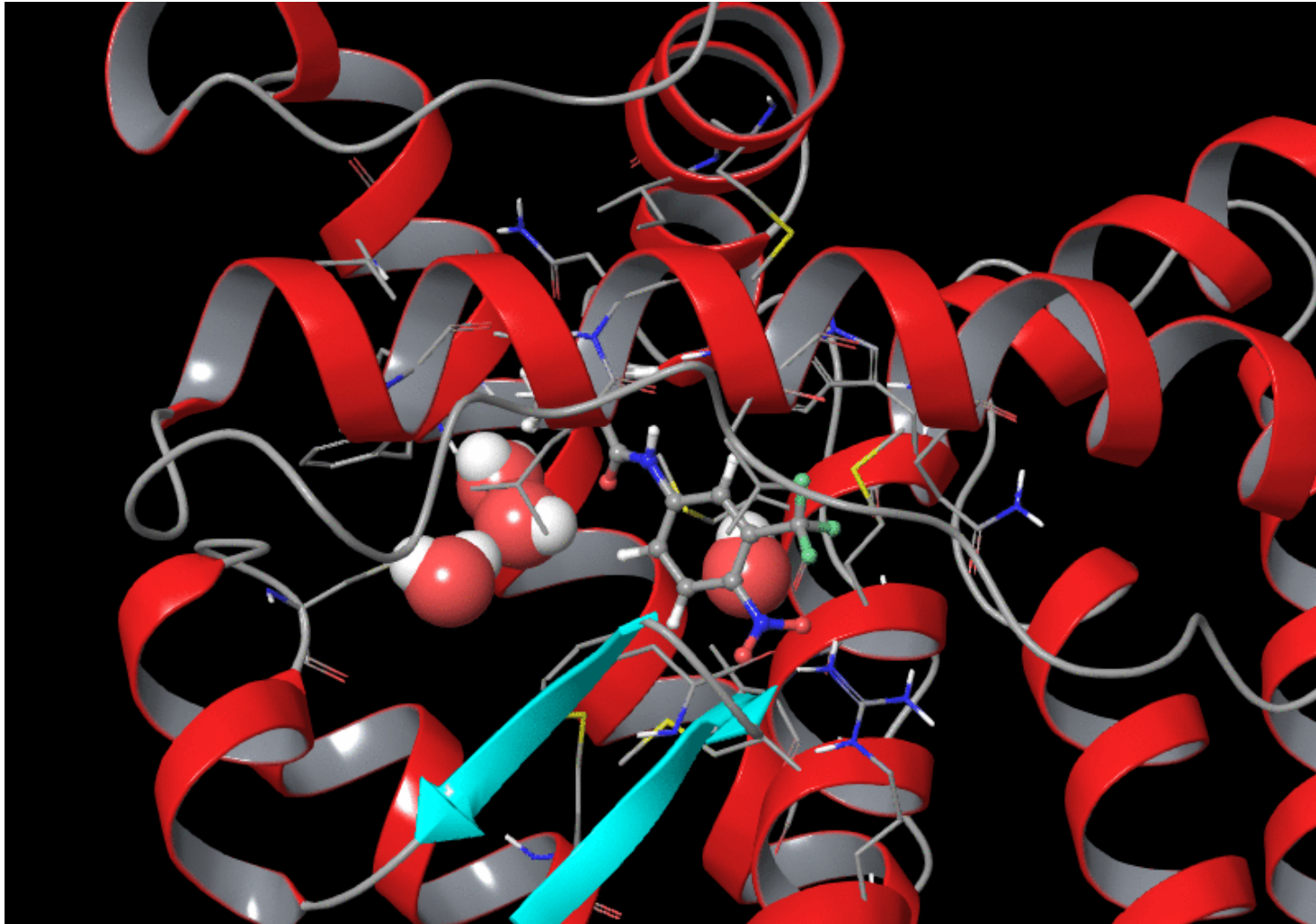
- **Rezeptor** (Gastgeber, Schloss, Handschuh) = empfangendes Molekül, meistens Protein oder NS
- **Ligand** (Gast, Schlüssel, Hand) = das komplementäre Partnermolekül, das am Rezeptor bindet
- **Bindungsmodus** = die Orientierung des Ligandes relativ zum Rezeptor und auch die Konformation des Ligandes und des Rezeptors, wenn zusammengebunden
- **Scoring** = Auswertung der Pose durch Quantifizieren der intermolekularen Wechselwirkungen (z.B. Wasserstoffbrücken, hydrophobe Kontakte, Elektrostatik...)
- **Ranking** = Anordnen der Liganden basierend auf das Score (typischerweise  $\Delta G_{\text{binding}}$ )





## Molecular Docking

extrem komplexes, aber gleichzeitig extrem attraktives rechnerisches Konzept





## Die bioaktive Pose mittels Docking zu finden, aber wie?

**Sampling** = ausforschen aller (?) möglichen Bindungsmodi eines kleinen Moleküls in der Bindungsstelle eines Proteins

**Manuelles Docking** = “fundierte Vermutung”, subjektiv, nicht reproduzierbar, aufwändig, limitierter Datendurchlauf

**Random Sampling** = ineffektiv, könnte theoretisch zur besten Lösung führen, aber Wahrscheinlichkeit ist sehr gering selbst für kleine flexible Moleküle

**Rationell (Regel) basiert** = man versucht die rechnerische Komplexität vom Docking zu reduzieren  
→ ähnliche Moleküle binden mit einem ähnlichen Bindungsmodus:

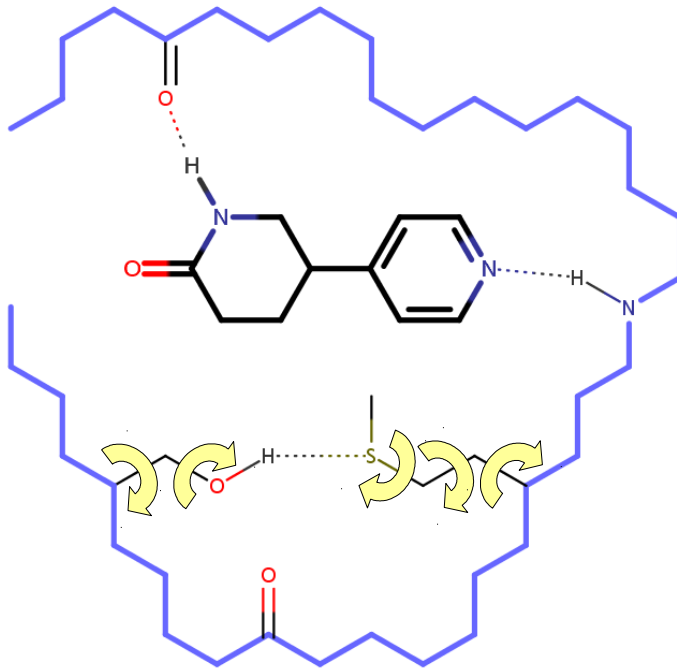
- Überlagerung des gedockten Moleküls mit dem Templatmolekül (co-kristallisiert)
- Überlagerung des gedockten Moleküls mit dem Pharmacophore (Bindungsstellenabdruck)
- Fragmentieren des gedockten Moleküls, platzieren der Fragmente in die idealen Lagen und Wiederaufbau des Moleküls innerhalb der Bindungsstelle
- viele verschiedene Algorithmen wurden entwickelt (verschiedene Vorteile/Nachteile)

**Kombinatorische Explosion** = 3 Translationen + 3 Rotationen + drehbare Bindungen (Ligand) + Rotamere der Seitenketten (Protein) + explizite Wassermoleküle (+ Kofaktoren, Hauptkette...)

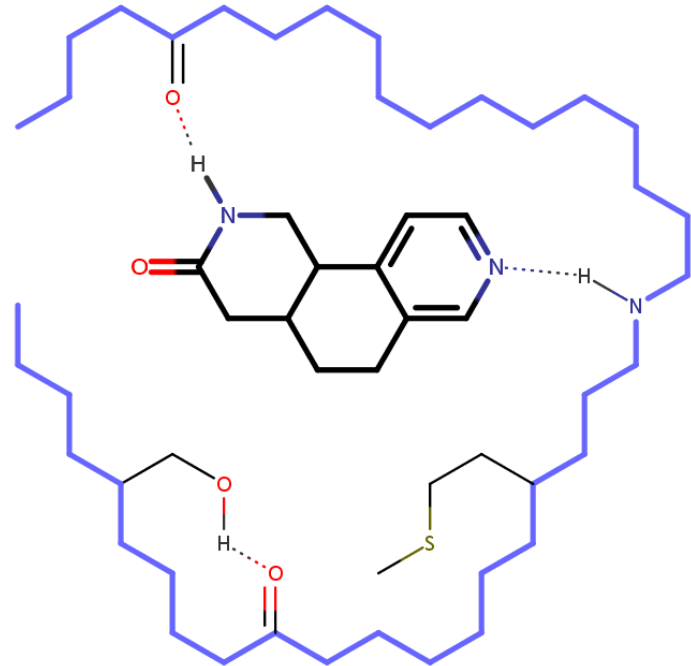


## Kombinatorische Explosion an der Proteinflexibilität

Protein-Ligand Komplex



Neues Ligand gedockt  
mit dem *induced fit*



**Serine:** 2 drehbare Bindungen, Schrittlänge  $30^\circ \rightarrow 144$  Rotamere

**Methionine:** 3 drehbare Bindungen, Schrittlänge  $30^\circ \rightarrow 1728$  Rotamere

Beide müssen gleichzeitig durchgesucht werden  $\rightarrow 248\,832$  möglich Kombinationen!



## **Rigides vs. flexibles Docking**

### **Rigides Docking**

- Molekül(e) werden rigid behandelt, das Docking ist vereinfacht auf Translationen/Rotationen
- extrem schnell, aber ziemlich ungenau → Screening von grossen Bibliotheken der kleinen einfachen Fragmente, die keine Anpassung des Proteins brauchen oder Docking von der Analoga

### **Flexibles Docking**

- Molekül(e) dürfen ihre Geometrie ändern (Torsionswinkeln bei den frei-drehbaren Bindungen)
- komplex, verbesserte Genauigkeit (?), Screening von kleineren Bibliotheken von grösseren /strukturell diversen Fragmenten/Moleküle

### **Verschiedene Kombinationen**

- Ligand flexibel & Protein rigid = Flexibilität des Ligandes in der Echtzeit / vorberechnete Konformere / fragmentierter Ligand + Wiederaufbau
- Ligand rigid & Protein flexibel = Echtzeit / vorberechnete Proteinzustände (*ensemble docking*)
- Ligand flexibel & Protein flexibel = erhöhte Komplexität

### **Weitere Faktoren die Dockingresultate stark beeinflussen können**

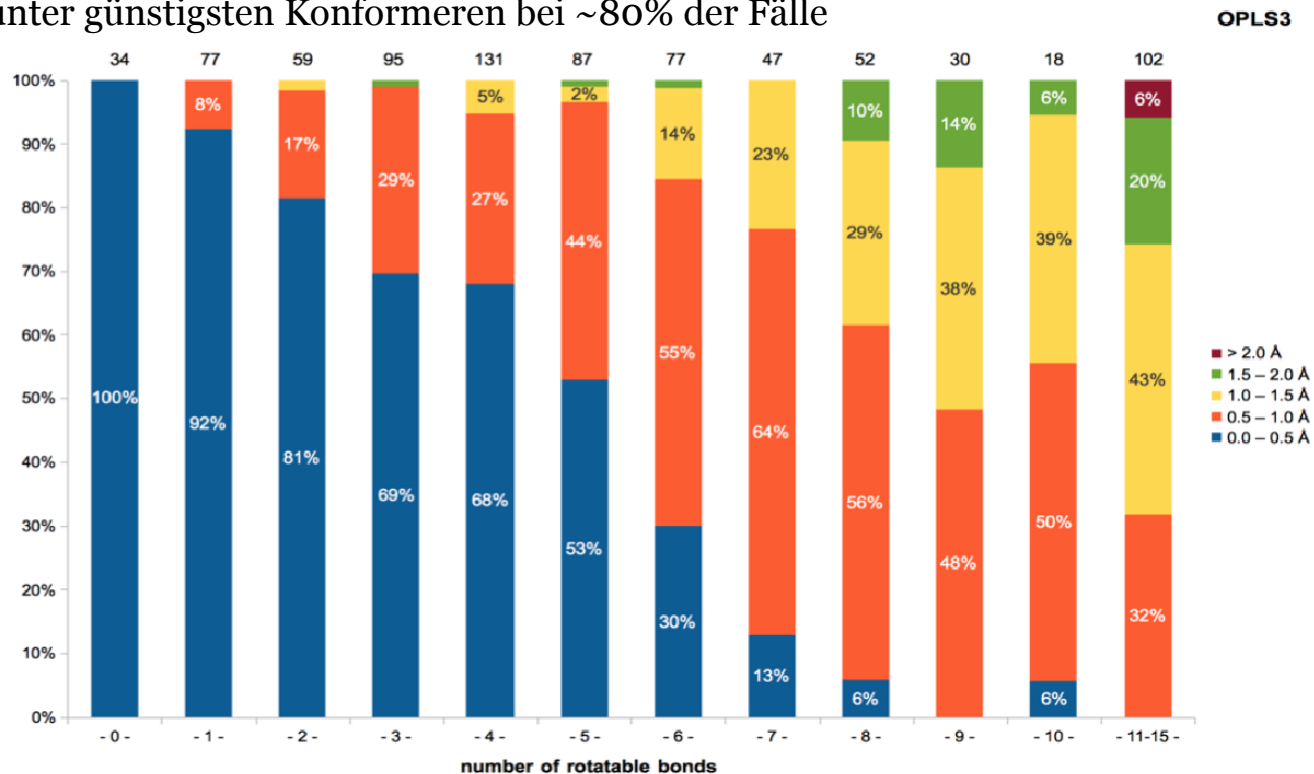
- Wasser, Kofaktoren, multiple Moleküle die simultan Binden (CYPs!)



## Ligandflexibilität

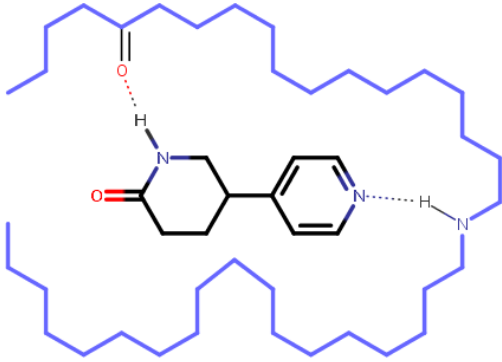
- Anpassen der Ligandengeometrie (Torsionswinkeln) „*on the fly*“ ist rechnerisch aufwändig
- Fragmentierung & Wiederaufbau des gedockten Moleküls in der Bindungsstelle
- Docking von energetisch günstigen Konformeren (aus der Konformationsuche) → Torsionswinkeln müssen nicht angepasst werden, gedockt als rigide Körper → schnell

Analyse von ~800 *drug-like* Moleküle in PDB: ein Konformer ähnlich zur Kristallpose (RMSD < 1.0 Å) ist präsent unter günstigsten Konformeren bei ~80% der Fälle

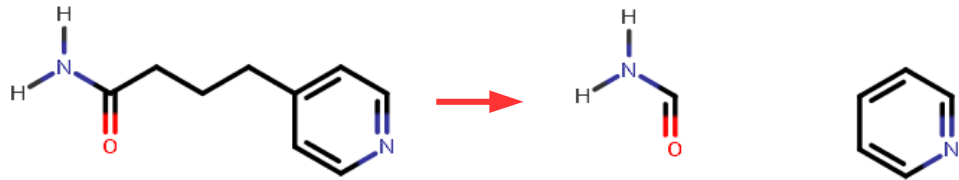




## Flexible docking – Wiederaufbau des Ligandes in der Bindungsstelle

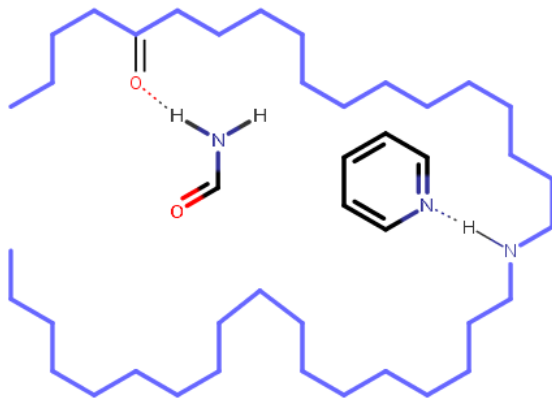


Protein-Ligand Komplex

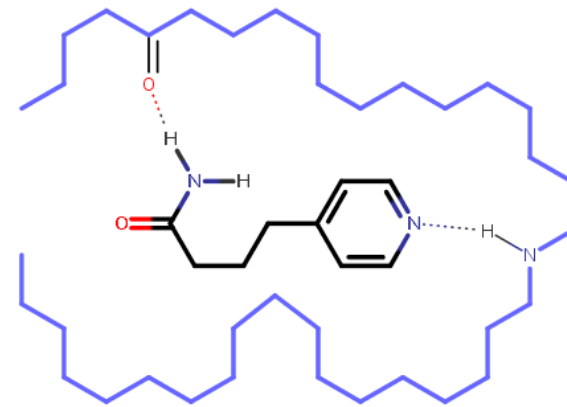


Gedocktes Molekül

fragmentiert



Fragmente platziert in die idealen Positionen

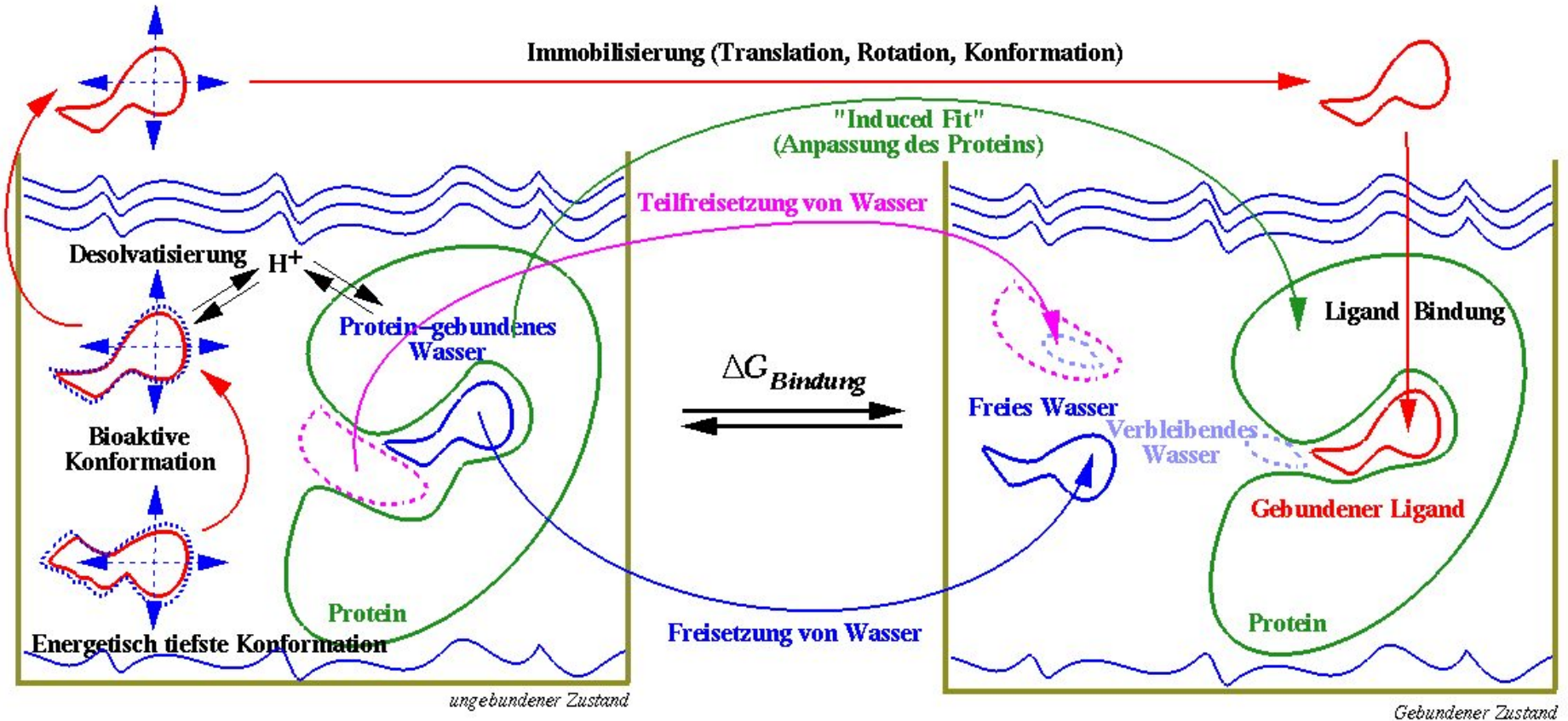


Fragmente wieder gebunden (Bewegung erlaubt)





# Abschätzung der Bindungsaffinität von Wirkstoffen



$$E_{\text{Bindung}} = E_{\text{Ligand-Protein}} + E_{\text{Ligand-Solvens}} - E_{\text{Innere Spannung}} - E_{\text{Ligand-Desolvatation}} - T\Delta S$$



# Pose scoring – Quantifizieren der Passform (Fit)

## Relative (Welche ist die beste Pose?)

- Scoring Funktion muss schnell sein, 1k bis 1M mal berechnet pro Substanz/Docking
- Methode der Wahl beim Struktur-basierten Design (Verbindung XYZ bindet besser als Templat/Lead)

**Absolute (IC<sub>50</sub>, K<sub>d</sub>, ΔG)**       $\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/t} + \Delta G_{vib}$

- extrem sensitiv
- anfällig für falsch negative (Sampling schwach) und falsch positive Resultate (Scoring Problem)
- komplex und rechnerisch intensiv – @docking: schnell, @refinement: präzise
- versagt wenn *induced fit* Effekte nicht richtig betrachtet
- viele Vereinfachungen um Schnelligkeit/Durchlauf zu gewinnen
- Entropie- Gewinne/Verluste nicht (akkurat) berechnet (abgeschätzt oder vereinfacht; MDs nötig)
- verschieden Korrektionssterme (PSA, SASA, Volumen) oder empirische Parametern (LogP, Dipol)
- ultimatives Ziel: eine Scoring Funktion basiert auf den ersten Prinzipien ohne zusätzlichen Korrekturen  
→ generelle Aussagekraft für alle Targets und Liganden (alle chemische Klassen und Bindungsstellen)

## Weitere Faktoren

Wasserverdrängung, Torsionswinkeln, Kooperativität der Wasserstoffbrücken, hydrophobe Kontakte



## Empfindlichkeit – Scoring von imperfekten Geometrien

Um den Datendurchlauf zu erhöhen sind die gedockte Bindungsmodi typischerweise nicht (komplett) minimiert und könnten ungenügend akkurat sein:

- *Intermolekulare enge Kontakte* : Van der Waals Term erlaubt keine Überlappungen → weiche Potentiale eingebaut (kleinere atomare Radii)
- *Rotationen/Translationen schrittweise!*
- *Torsionswinkeln modifiziert schrittweise!*
- *Elektrostatische WW* – überschätzt bei kleinen Distanzen

