

# *Aufbaukurs Bioinformatik*

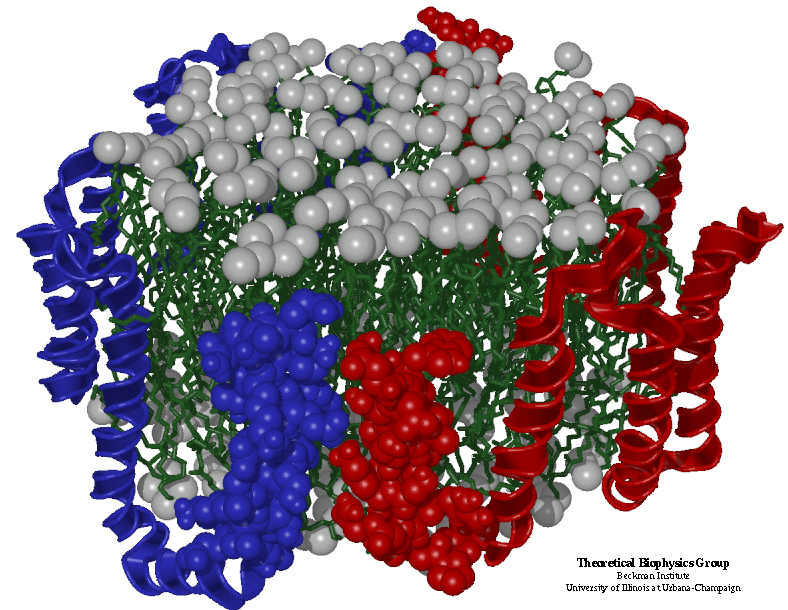
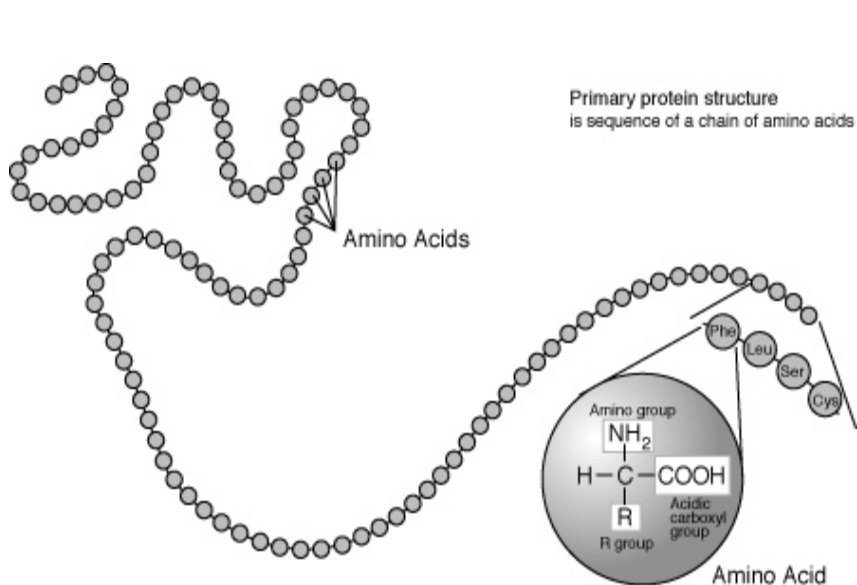


2. Tag: Proteindomänen, -struktur, Datenbanken  
und Werkzeuge

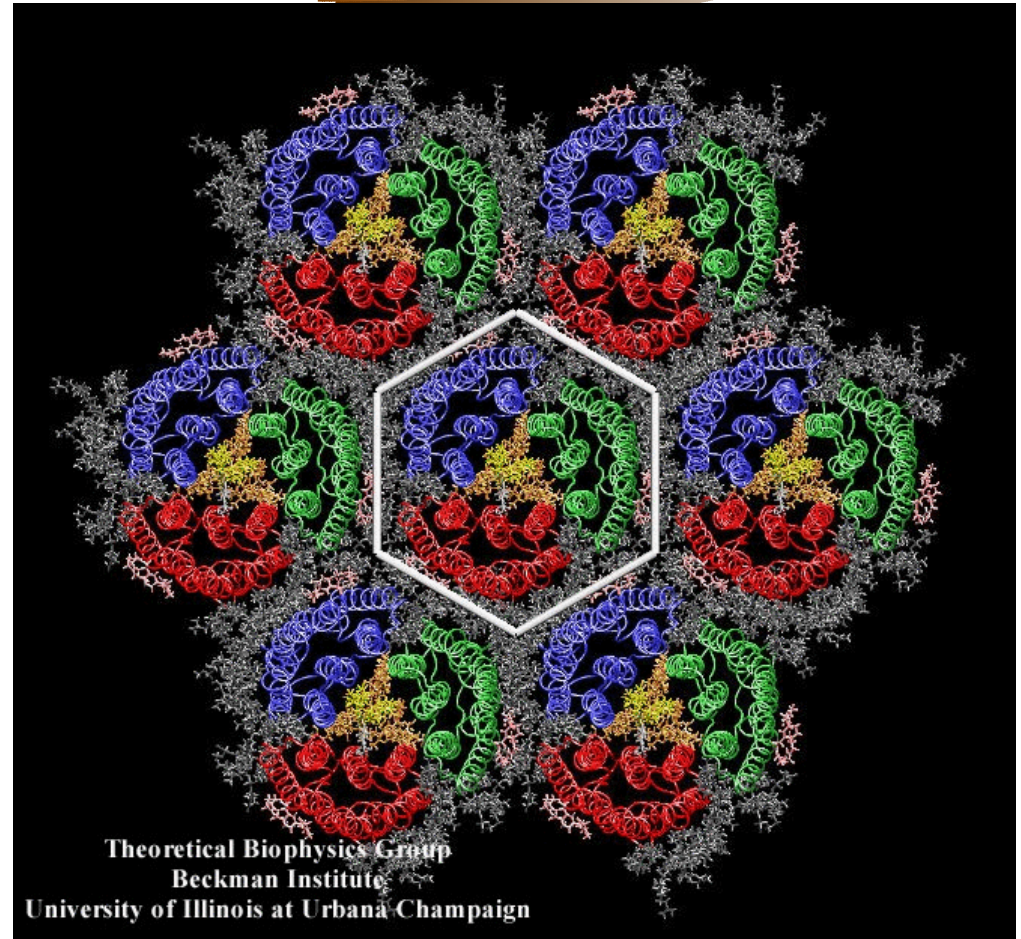
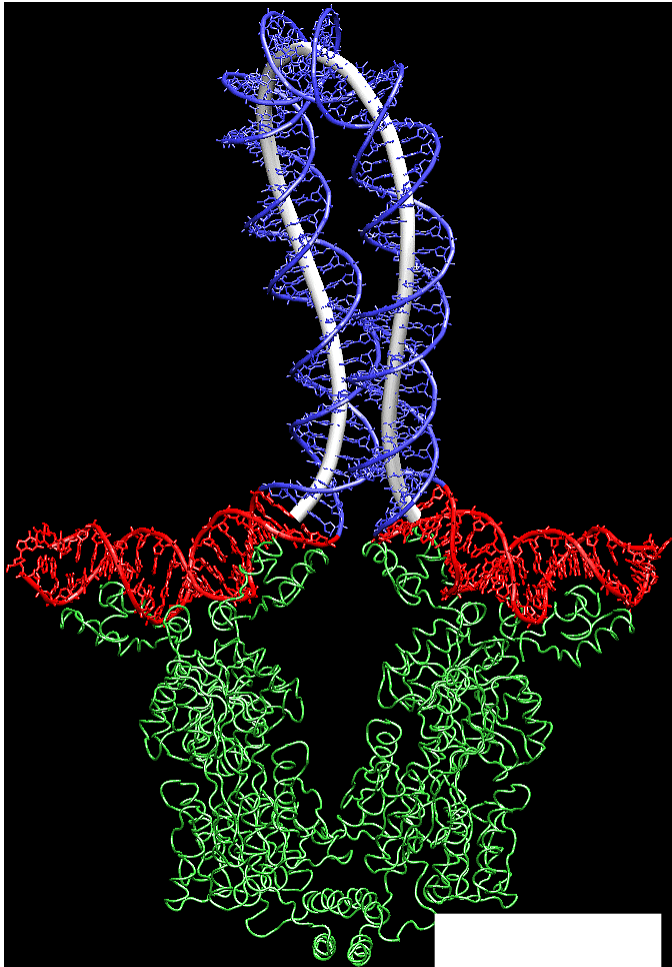
Ursula Kummer, Sven Sahle  
mit Ulla Rost, Katja Wegner und Andreas Weidemann

# Proteine

- Proteine sind Makromoleküle, die viele Abläufe in der Zelle bestimmen (Strukturbildung- und Erhaltung, Transport, Schutz und Abwehr, Steuerung und Regelung, Katalyse, Bewegung, Speicherung)
- Im menschlichen Körper werden etwa 100 000 verschiedene Proteine vermutet.



# Proteine



# Proteinstrukturen



- **Primärstruktur:** Aminosäuresequenz
- **Sekundärstruktur:**  $\alpha$  -Helix,  $\beta$  -Faltblatt u.a.
- **Tertiärstruktur:** Kombination von  $\alpha$  -Helices und  $\beta$ -Faltblättern zu Domänen bzw. Untereinheiten
- **Quartärstruktur:** Kombinationen von verschiedenen Proteinsequenzen durch schwache nicht kovalente Bindungen zu größeren Proteinkomplexen

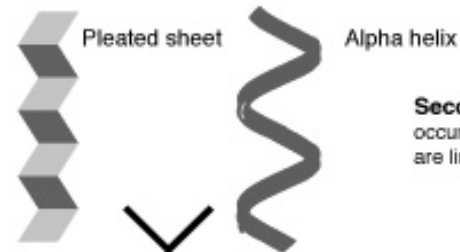
# Proteinstrukturen

## Primärstruktur



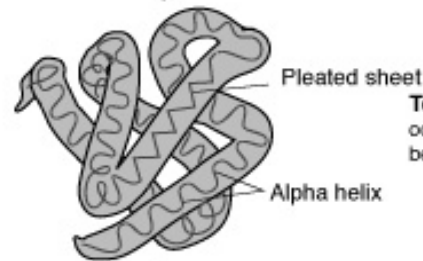
**Primary protein structure**  
is sequence of a chain of amino acids

## Sekundärstruktur



**Secondary protein structure**  
occurs when the sequence of amino acids  
are linked by hydrogen bonds

## Tertiärstruktur



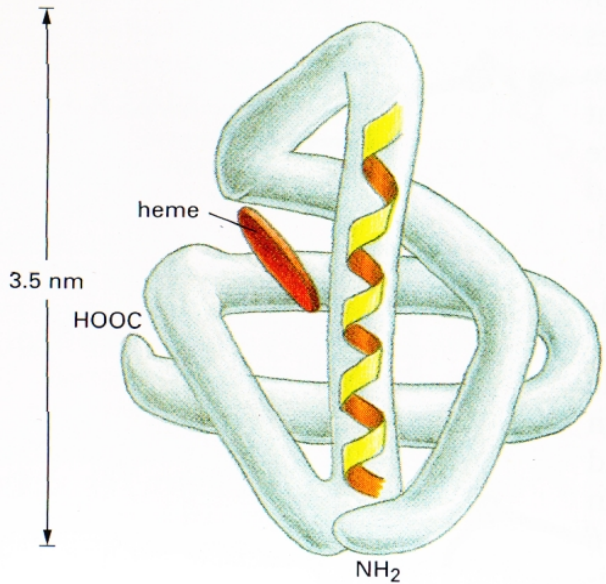
**Tertiary protein structure**  
occurs when certain attractions are present  
between alpha helices and pleated sheets.

## Quartärstruktur

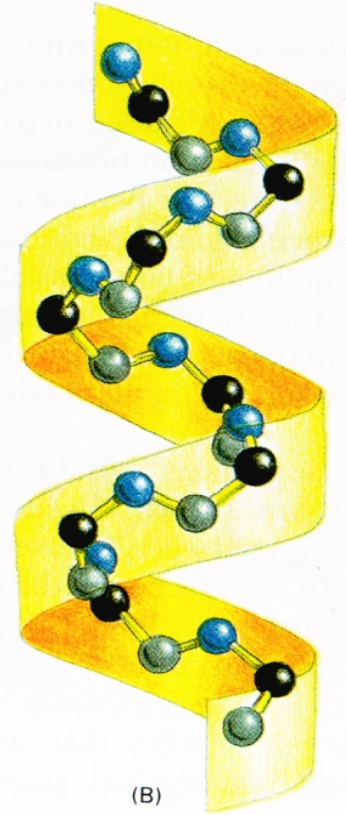


**Quaternary protein structure**  
is a protein consisting of more than one  
amino acid chain.

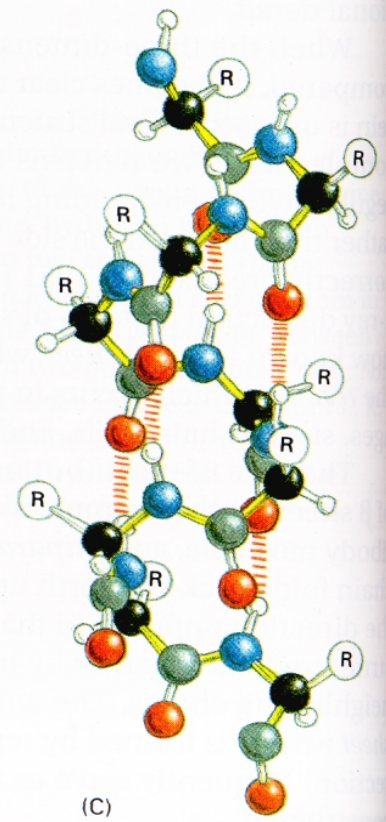
# $\alpha$ -Helix



(A)



(B)

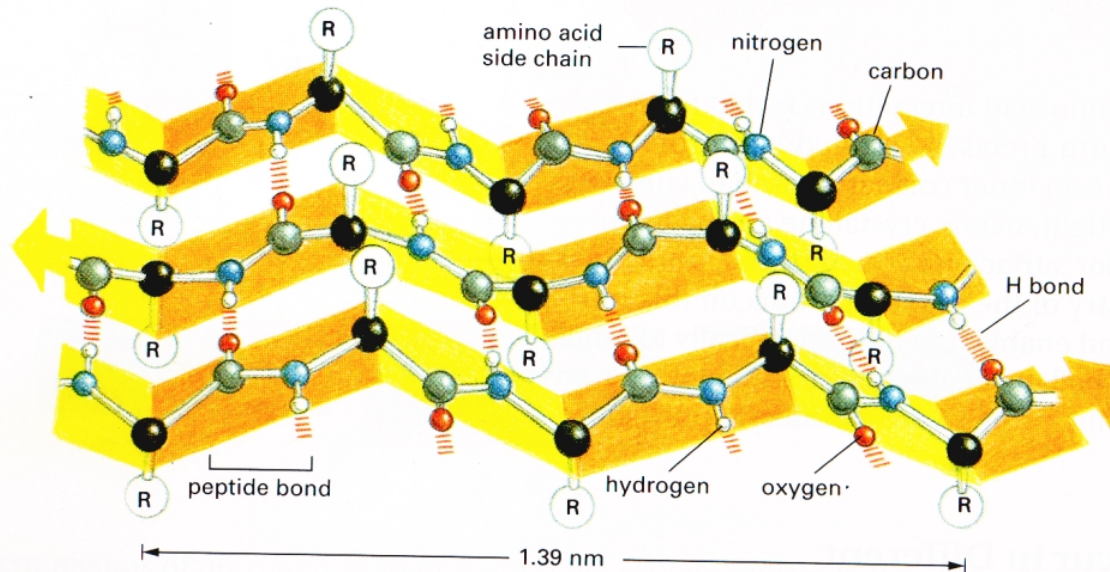
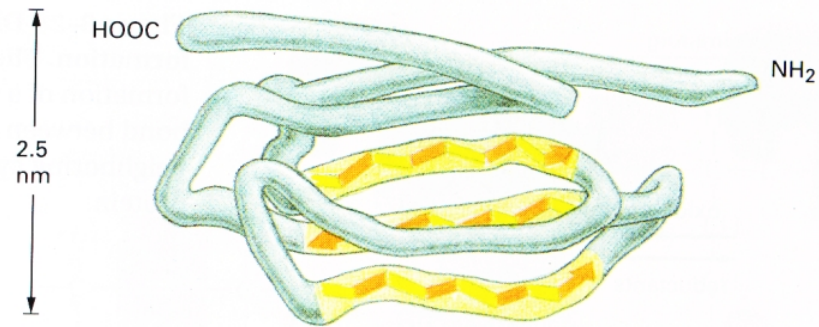


(C)



- 32-38% aller AS globulärer Proteine
- H-Brücken zwischen C=O und H-N zwischen den Aminosäuren  $i$  und  $i+4$  -> regelmäßige, stabile Anordnung
- Pitch: 5.4 Å entlang der Achse, 3.6 AS pro Umdrehung
- Anstieg der Helix pro AS: 1.5 Å

# $\beta$ -Faltblatt





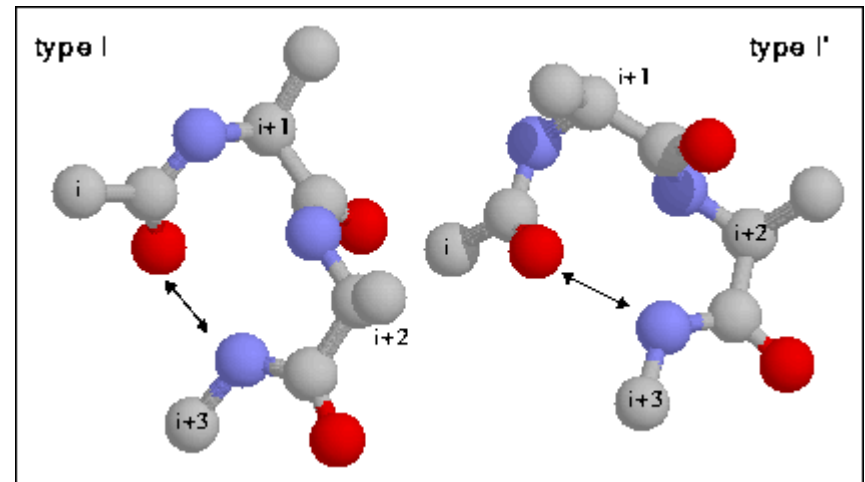
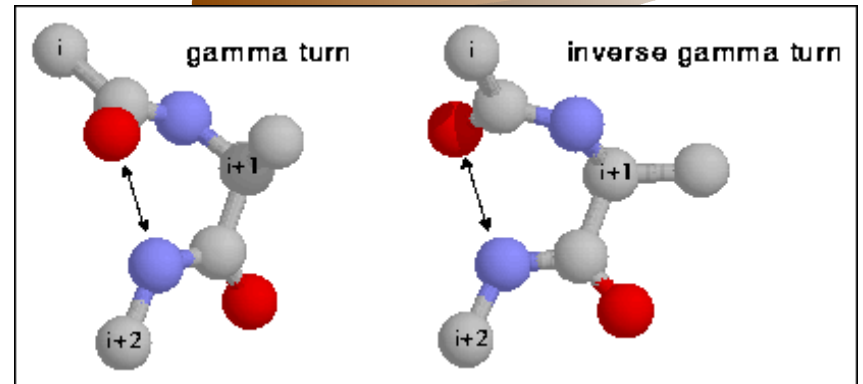
# *$\beta$ -Faltblatt*



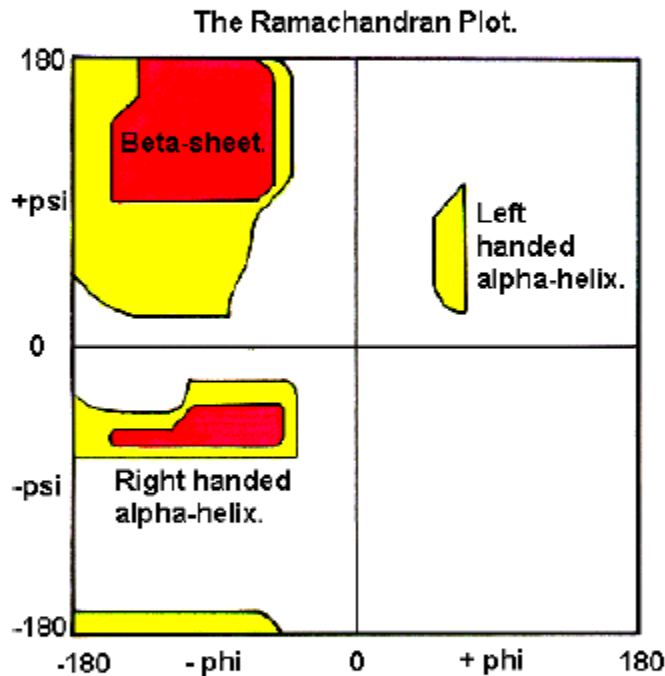
- Zickzack-artige Anordnung und wesentlich gestreckter als eine  $\alpha$ -Helix (Pitch 6.8 Å)
- Peptidbindungen benachbarter Aminosäuren zeigen in entgegengesetzte Richtung, ebenso die Seitenketten (senkrecht zur Ausrichtung der CO-NH-Gruppen)
- Mehrere  $\beta$ -Stränge können sich durch Ausbildung von Wasserstoff-Brücken zu Faltblättern anordnen
- Länge des  $\beta$ -Strangs  $\sim$  #AS  $\times$  3.4 Å

# Turns

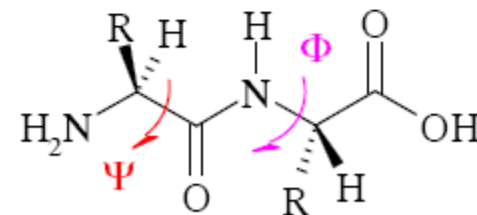
Um die Richtung einer Proteinkette umzukehren und um ein  $\beta$ -Faltblatt zu bilden, sind *turns* notwendig



# Ramachandranplot



- | Blick vom N-Terminus  
Drehung im Uhrzeigersinn  $\Rightarrow$  positive Winkel
- | Weisser Bereich: Atome naher als die Summe ihrer van der Waals-Radien  
 $\Rightarrow$  sterisch verboten (Ausnahme: Gly)
- | **Rot**: keine sterischen Hinderungen ( $C_{\beta}$ - $CH_2$ )  
 $\Rightarrow$  ideale Sekundarstruktur
- | **gelb**: erlaubte Region bei leicht verkurzten van der Waals-Radien  
 $\Rightarrow$  zusatzliche linkshandige Helix



# AS bestimmen die Struktur

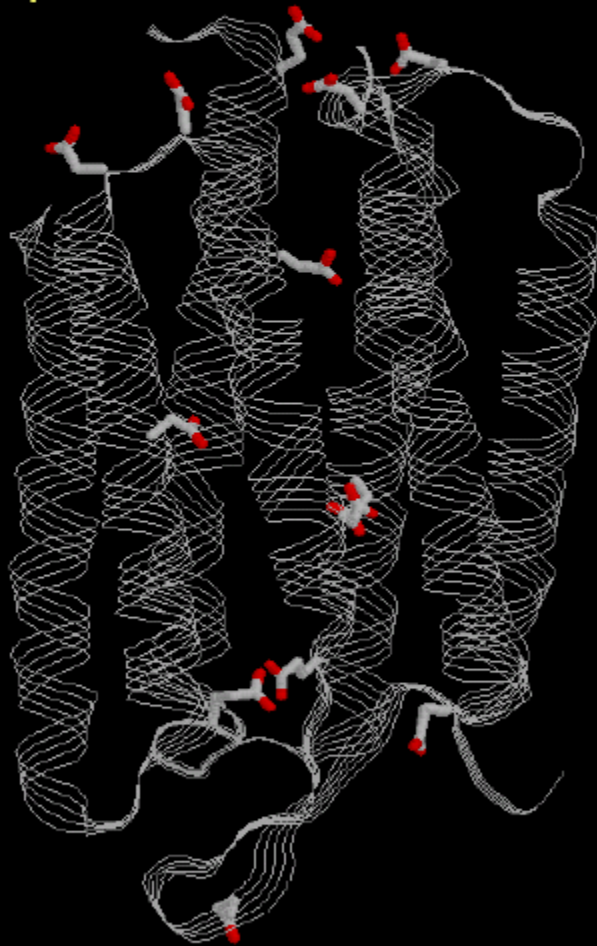
## Sekundärstruktur-Statistik

**Table 9.1**  $\alpha$ -helix and  $\beta$ -chain character of individual residues (Ptitsyn and Finkelstein 1983)

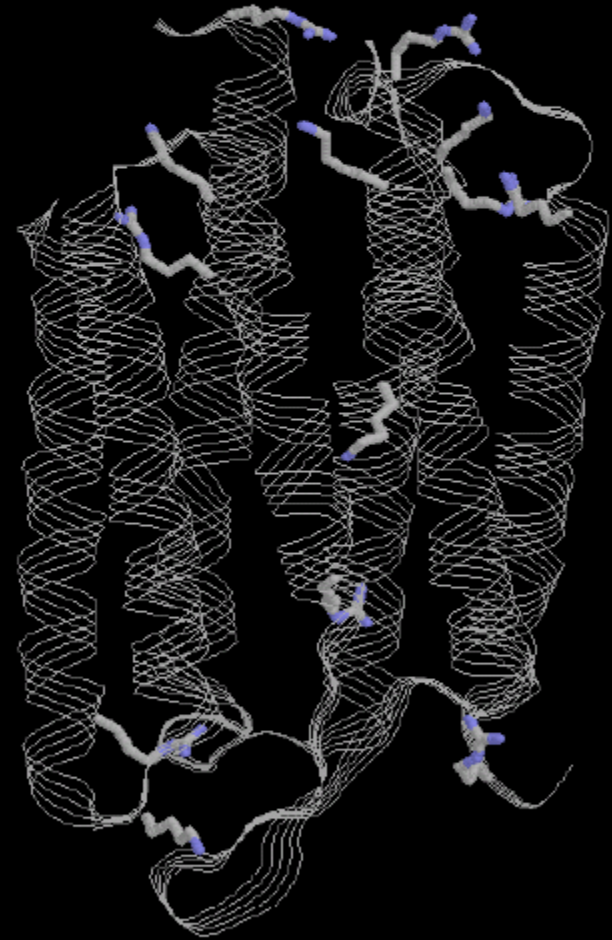
|             | $\alpha$ -helix   | $\beta$ -chain   |
|-------------|---|--|
| former      | leu, met, lys, glu, ala,<br>arg, ile, phe, trp                                      | val, ile, phe, tyr, trp, leu,<br>cys, met, thr, his, lys   |
| indifferent | tyr, gln, lys <sup>+</sup> , glu <sup>-</sup> , his,<br>arg <sup>+</sup> , val, cys | glu, arg <sup>+</sup> , ala, gln, ser  |
| breaker     | his <sup>+</sup> , asn, asp, thr, ser,<br>asp <sup>-</sup> , gly, pro               | his <sup>+</sup> , lys <sup>+</sup> , arg <sup>+</sup> , glu <sup>-</sup> ,<br>asp, asn, asp <sup>-</sup> , glu, pro |

# *Verteilung einiger AS in Bakteriorhodopsin*

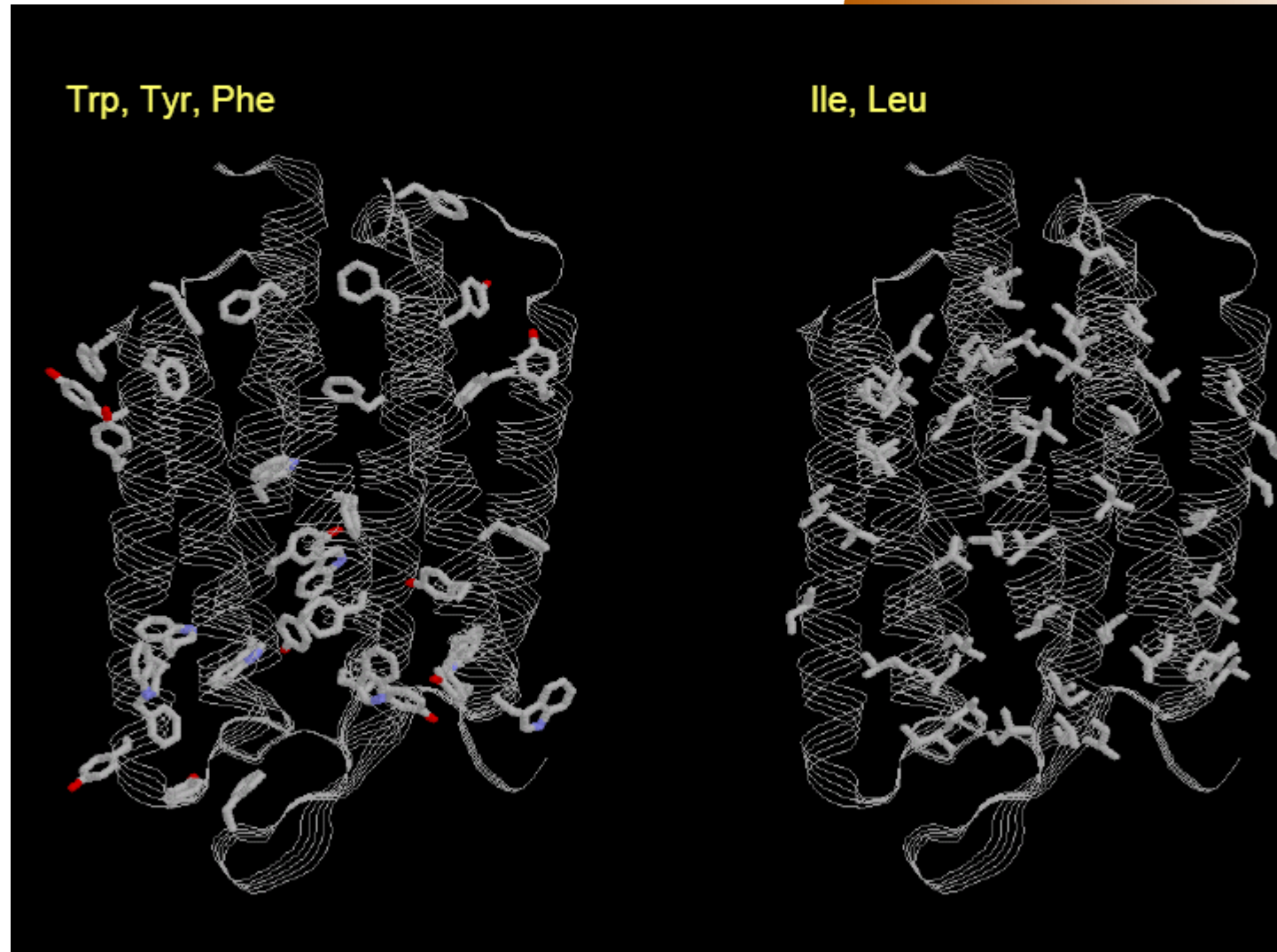
Asp, Glu



Arg, Lys



# Verteilung einiger AS in Bakteriorhodopsin



- Funktionale Einheiten eines Proteins
  - Bedeutung für die Wechselwirkung mit anderen Molekülen in der Zelle haben
  - Sequenz-Domäne: zusammenhängender String innerhalb der der Sequenz, der stark konserviert ist und somit ein charakteristisches Motiv darstellt. Es gibt eine (wenn auch nicht immer klar erkennbare) funktionale Bedeutung
  - Struktur-Domäne: strukturell unabhängiger Bereich im Protein

# Datenbanken im Netz

- Datenbanken für Domänen (funktionale Einheiten in Proteinen)
  - Pfam <http://www.sanger.ac.uk/Software/Pfam/>  
semi-automatisch erzeugt
  - ProDom <http://protein.toulouse.inra.fr/prodom.html>  
automatisch erzeugt
  - Prosite <http://www.expasy.org/prosite>  
auch Muster, z.B. proteinfamiliäntypische
  - InterPro: <http://www.ebi.ac.uk/interpro> Integration von Informationen über Proteine und deren funktionelle Einheiten (Pfam+ProDom+Prosite+PRINTS+SWISS-PROT + Trembl), halbautomatisch erzeugt



# *Transmembrandomänen*



Algorithmus:

- > schaut nach dem wahrscheinlichen Auftreten von alpha-helices
- > wenn diese mehr als 18 AS enthält -> Transmembrandomäne
- > Achtung: Am Anfang eines Proteins: Signalsequenz
- > Falls keine Transmembrandomänen, kann die Lokalisation nicht vorhergesagt werden

# Transmembrandomänen

<http://www.cbs.dtu.dk/services/TMHMM-2.0/>

|   |        |         |                    |                              |                                |                |  |
|---|--------|---------|--------------------|------------------------------|--------------------------------|----------------|--|
| CENTER FOR<br>BIOLOGICAL<br>SEQUENCE<br>ANALYSIS<br>CBS | EVENTS | NEWS    | RESEARCH<br>GROUPS | CBS<br>PREDICTION<br>SERVERS | CBS<br>DATA<br>SETS            | PUBLICATIONS   | BIOINFORMATICS<br>EDUCATION<br>PROGRAM |
|   | STAFF  | CONTACT | ABOUT<br>CBS       | INTERNAL                     | CBS<br>BIOINFORMATICS<br>TOOLS | CBS<br>COURSES | OTHER<br>BIOINFORMATICS<br>LINKS       |

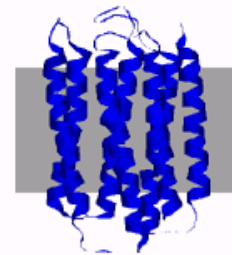
[CBS](#) >> [CBS Prediction Servers](#) >> TMHMM

## TMHMM Server v. 2.0

### Prediction of transmembrane helices in proteins

Please try the new server [Phobius](#)

NOTE: You can submit many proteins at once in one fasta file. Please limit each submission to at most 4000 proteins. Please tick the 'One line per protein' option. Please leave time between each large submission.



### [Instructions](#)

#### SUBMISSION

Submission of a local file in **FASTA** format (HTML 3.0 or higher)

OR by pasting sequence(s) in **FASTA** format:

#### Output format:

- Extensive, with graphics
- Extensive, no graphics
- One line per protein



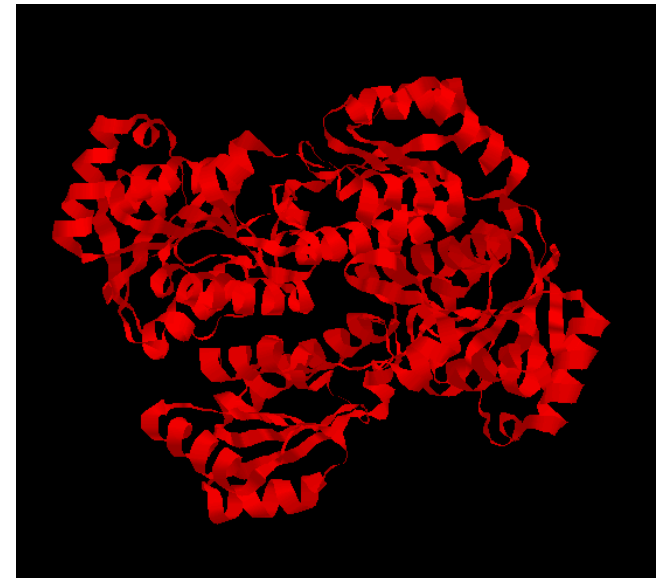
# 3D-Struktur

Kenntnisse der 3D-Struktur ermöglichen Aussagen über

- Funktionen
- Wichtigkeit einzelner Aminosäuren
- Liganden (auch Drugs)
- Kinetische Eigenschaften
- andere Interaktionspartner
- Lokalisation

# Proteinfaltung

- Berechnung der Proteinstruktur aus der Sequenz
- Unter Verwendung von Kraftfeldmethoden (Newton'sche Mechanik und Elektrostatik) und Quantenmechanik berechnet man die Struktur mit minimaler Energie
- Extrem aufwändig!
- Heute eher durch vergleichende Analyse
- CASP-Wettbewerb



# MD-Simulationen

- Normalerweise benutzt man MD um Änderungen der Struktur zu
- Zuordnung einer Anfangsgeschwindigkeit zu jedem Atom und Anwendung von Newtons
- Das 2. Newtonsche Gesetz:  $F_i = m_i \cdot a_i$  mit F als Summe aller Kräfte, die auf das i-te Atom wirken. Die Beschleunigung ist die zweite Ableitung der Position nach der Zeit.

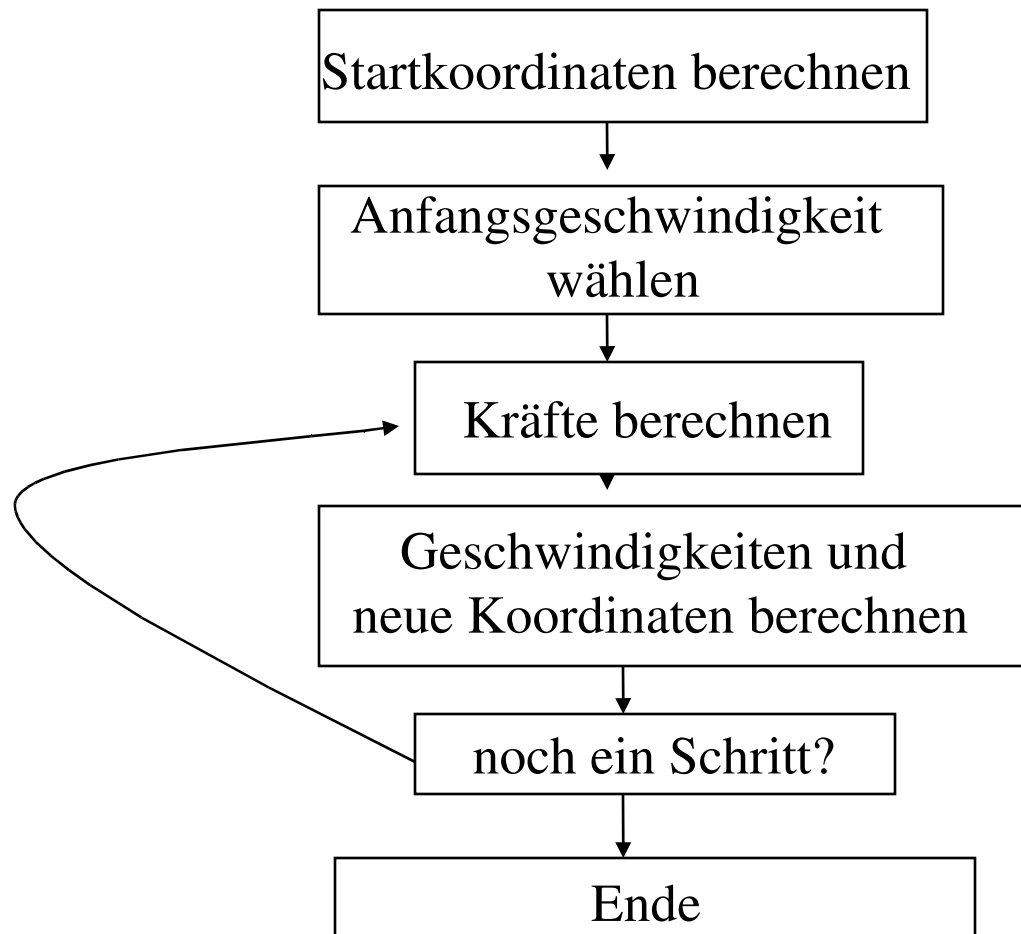
Der 'Leap Frog' - Algorithmus wird oft für die Integration benutzt:

$$\vec{v}_i(t_{n+1/2}) = \vec{v}_i(t_{n-1/2}) + \frac{\vec{F}_i(t_n)}{m_i} \Delta t,$$

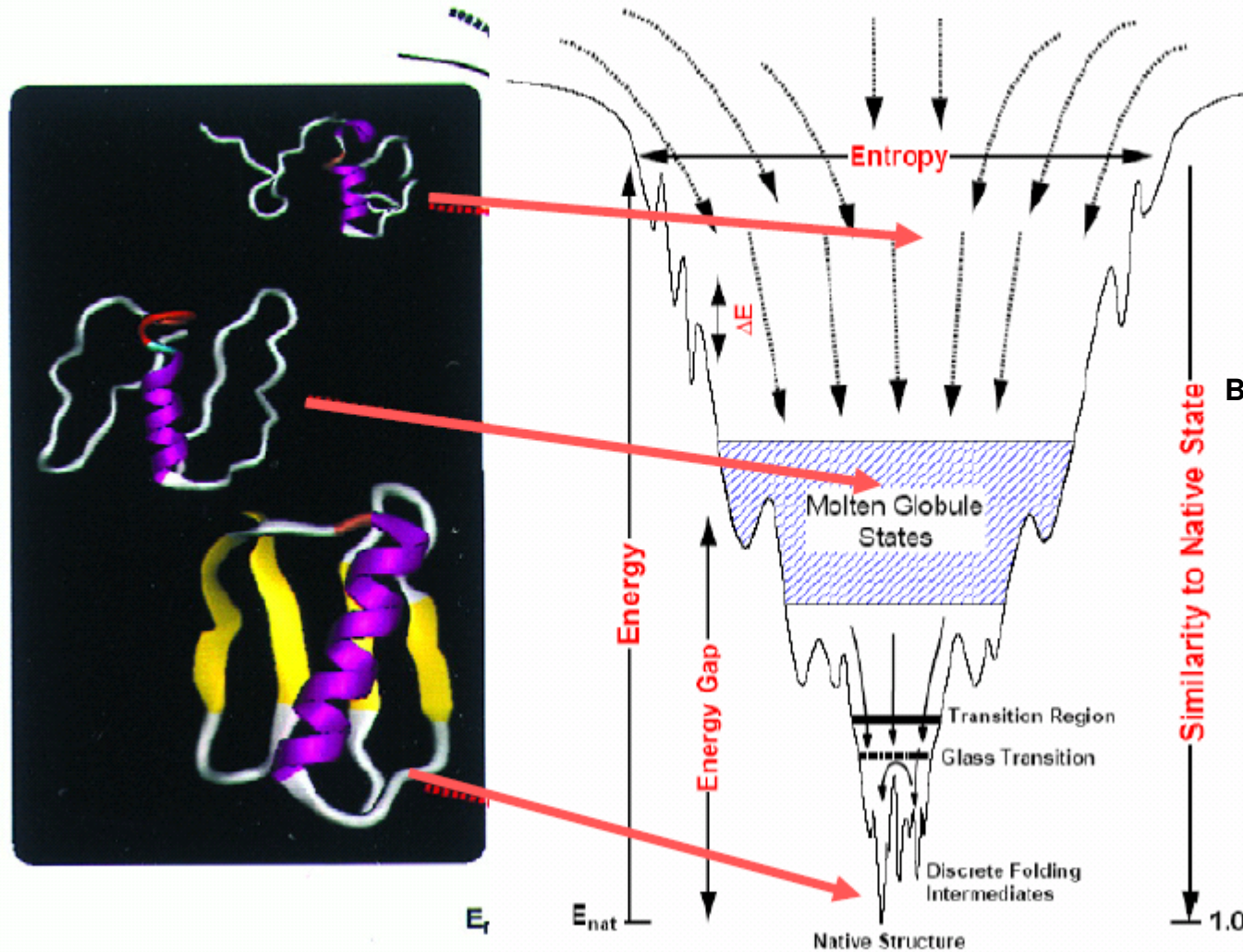
$$\vec{r}_i(t_{n+1}) = \vec{r}_i(t_n) + \vec{v}_i(t_{n+1/2}) \Delta t.$$

# Moleküldynamik

- Berechnung von Molekülbewegungen, z.B. Docking



# Proteinfaltungsvorhersage



Bryngelson, Wolynes, PNAS (1987)

# MD-Simulationen

Kann man einen Faltungsprozess mit MD-Simulationen simulieren?

1998 -> Simulation der 36-Residuen des Villin-Fragments

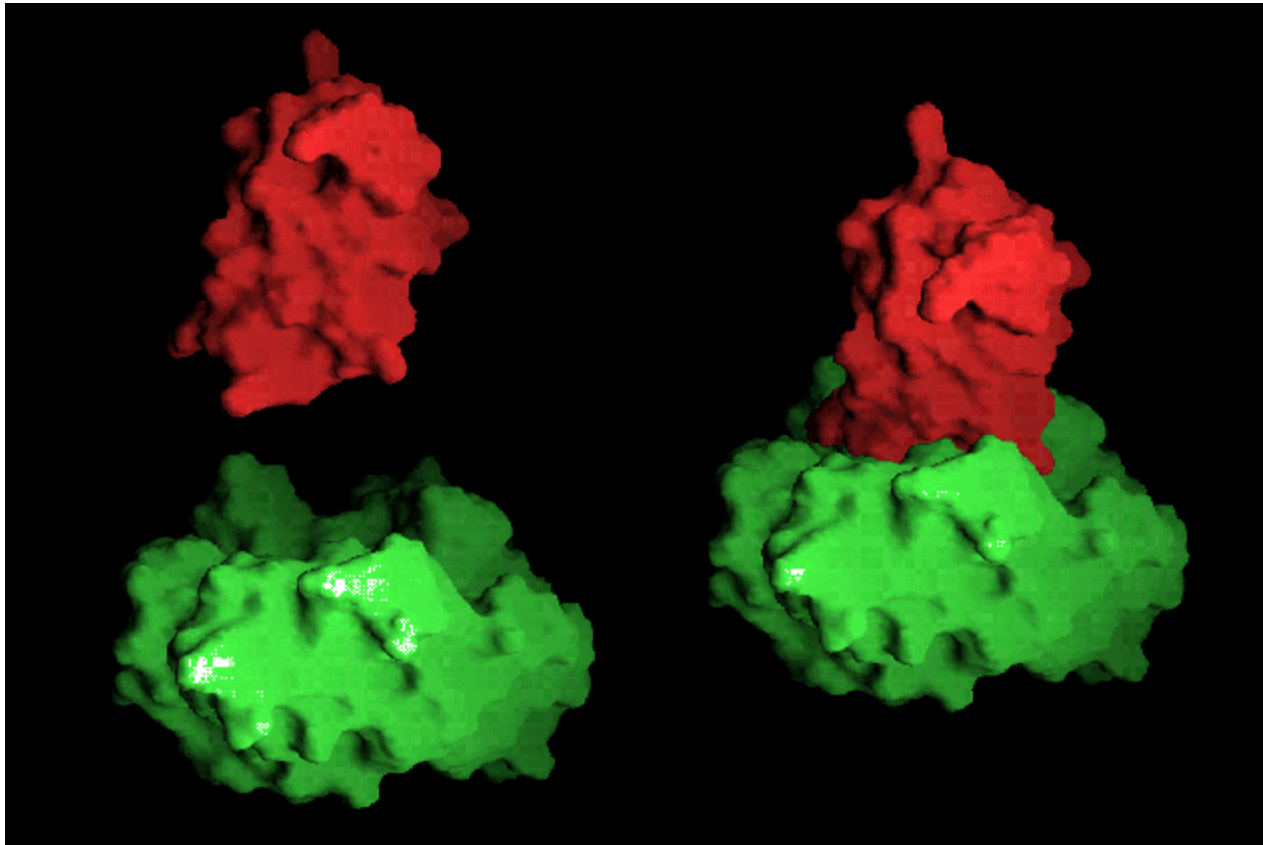
exp. Faltungszeit: zwischen 10 – 100 s,  $T_m = 70^\circ \text{C}$

- enthält 3 kurze Helices (NMR), die durch Loop und Schleife verbunden sind
- dicht gepackter hydrophober Kern

-> 4 Monate CPU Zeit auf 256 Prozessor Cray T3D und T3E



# Docking



# *Homologie-Modeling*



- Bei bekannter Struktur eines/mehrerer verwandter Proteine
- Anpassung der Struktur an die veränderten Aminosäuren
- Funktioniert ziemlich gut und ist viel schneller als herkömmliche Methoden
- Ganze Datenbanken wurden damit generiert

# Swiss-Model

<http://swissmodel.expasy.org/repository/>

## SWISS-MODEL REPOSITORY

[Home](#) | [Advanced Search](#) | [>>Swiss-Model](#) | [HELP](#)

UniProt AC or ID:

### Welcome to the SWISS MODEL Repository

The SWISS-MODEL Repository is a database of annotated three-dimensional comparative protein structure models generated by the fully automated homology-modelling pipeline SWISS-MODEL. The repository is developed at the Biozentrum Basel within the Swiss Institute of Bioinformatics.

#### REPOSITORY STATUS

|              |           |
|--------------|-----------|
| Models:      | 1341793   |
| UniProt:     | 10.4      |
| PDB:         | 7.5.2007  |
| Last update: | 30.5.2007 |



**NOTE:** The SWISS-MODEL repository contains theoretically calculated models, which may contain significant errors.

Jürgen Kopp & Torsten Schwede

[swissmodel.expasy.org/repository](http://swissmodel.expasy.org/repository)

[\[Disclaimer\]](#) [\[HELP\]](#) [\[Contact\]](#)

# Swiss-Model

```
      250      260      270      280      290      300
ACYFETASAV TKDPAVAARG SALTPISMES GNFDISKYRFV LPTTKFDLDI DDASLNKGQQ

      310      320      330      340      350      360
ALEKMISGMV LGEIARRVIV HLSSINCLPA ALQTALGNRG SFESRFAGMI SADRMPGLQF

      370      380      390      400      410      420
TRSTIQKVCV VDVQSIEDLR IIRDVCLVR GRAAQLSASF CCAPLVKTQT QGRATIAIDG

      430      440      450      460      470
SVFEKIPSPR RVLQDNINRI LGPECDVRAV LAKGGSGVGA ALISAIVADG K
```

Q38C41 in FASTA format

[View entry in original UniProtKB/TrEMBL format](#)

[View entry in raw text format \(no links\)](#)

[Request for annotation of this UniProtKB/TrEMBL entry](#)

**BLAST**

BLAST submission on ExPASy/SIB  
or at NCBI (USA)



Sequence analysis tools: ProtParam, ProtScale, Compute pI/Mw,  
PeptideMass, PeptideCutter, Dotlet (Java)



ScanProsite, MotifScan



Submit a homology modeling request to SWISS-MODEL

**NPS@**

NPSA Sequence analysis tools



[ExPASy Home page](#)

[Site Map](#)

[Search ExPASy](#)

[Contact us](#)

[Swiss-Prot](#)

Hosted by SIB Switzerland | Mirror sites: | [Australia](#) | [Brazil](#) | [Canada](#) | [China](#) | [Korea](#) |

# *Datenbanken im Netz*



- Protein Data Bank (PDB)  
<http://www.wwpdb.org/>
- Protein Structure Database (PSdb)  
<http://www.psc.edu/~deerfiel/PSdb>  
Java viewer
- Molecular Modeling Database (MMDB)  
<http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml>
- EBI Macromolecular Structure Database (MSD)  
<http://www.ebi.ac.uk/msd/index.html>  
viele Tools

# PDB files

```
SCALE3      0.000000  0.000000  0.009027      0.000000
ATOM        1  N   LYS   658      23.864  20.736  75.930  1.00  46.26      N
ATOM        2  CA  LYS   658      22.609  20.563  75.140  1.00  45.18      C
ATOM        3  C   LYS   658      22.886  20.966  73.686  1.00  44.83      C
ATOM        4  O   LYS   658      23.720  21.835  73.418  1.00  45.53      O
ATOM        5  CB  LYS   658      21.479  21.423  75.733  1.00  45.82      C
ATOM        6  CG  LYS   658      20.081  21.038  75.278  1.00  46.76      C
ATOM        7  CD  LYS   658      19.708  19.661  75.801  1.00  49.57      C
ATOM        8  CE  LYS   658      18.512  19.070  75.072  1.00  50.09      C
ATOM        9  NZ  LYS   658      18.818  18.805  73.631  1.00  52.67      N
ATOM       10  N   SER   659      22.216  20.293  72.757  1.00  43.34      N
ATOM       11  CA  SER   659      22.372  20.550  71.331  1.00  41.62      C
ATOM       12  C   SER   659      21.759  21.903  70.925  1.00  39.27      C
ATOM       13  O   SER   659      22.464  22.779  70.422  1.00  37.49      O
ATOM       14  CB  SER   659      21.727  19.408  70.543  1.00  41.75      C
ATOM       15  OG  SER   659      20.345  19.319  70.850  1.00  43.69      O
ATOM       16  N   GLU   660      20.460  22.070  71.179  1.00  36.37      N
ATOM       17  CA  GLU   660      19.735  23.293  70.851  1.00  36.27      C
ATOM       18  C   GLU   660      20.379  24.556  71.424  1.00  34.20      C
ATOM       19  O   GLU   660      20.221  25.643  70.876  1.00  34.42      O
ATOM       20  CB  GLU   660      18.300  23.197  71.359  1.00  37.83      C
ATOM       21  CG  GLU   660      17.487  22.088  70.719  1.00  46.15      C
ATOM       22  CD  GLU   660      16.303  21.620  71.589  1.00  51.57      C
ATOM       23  OE1 GLU   660      16.115  22.157  72.706  1.00  54.44      O
ATOM       24  OE2 GLU   660      15.562  20.700  71.160  1.00  54.28      O
ATOM       25  N   ALA   661      21.123  24.403  72.515  1.00  32.01      N
ATOM       26  CA  ALA   661      21.764  25.529  73.186  1.00  28.22      C
ATOM       27  C   ALA   661      23.194  25.870  72.748  1.00  27.00      C
ATOM       28  O   ALA   661      23.590  27.030  72.774  1.00  27.86      O
ATOM       29  CB  ALA   661      21.722  25.306  74.695  1.00  27.66      C
ATOM       30  N   LEU   662      23.996  24.875  72.398  1.00  23.53      N
ATOM       31  CA  LEU   662      25.368  25.155  72.008  1.00  21.70      C
ATOM       32  C   LEU   662      25.532  25.257  70.494  1.00  20.79      C
ATOM       33  O   LEU   662      26.466  25.892  70.005  1.00  22.55      O
ATOM       34  CB  LEU   662      26.292  24.078  72.563  1.00  17.40      C
ATOM       35  CG  LEU   662      26.106  23.784  74.040  1.00  18.80      C
ATOM       36  CD1 LEU   662      26.848  22.528  74.386  1.00  21.15      C
```



## *Tools*

- Online visualization <http://aismig.dkfz.de/service.php>
- Protein Structure Prediction Server  
<http://bmerc-www.bu.edu/psa/index.html>
- Molecular Visualization Freeware  
<http://www.umass.edu/microbio/rasmol>
- Swiss Pdb Viewer  
<http://www.expasy.org/spdbv>
- PDBsum (einfache und schnelle Visualisierung von PDB-Einträgen)  
<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>