

V6 Homologie-basierte Proteinmodellierung

- **Idee:** Sequenzähnlichkeit führt oft zur Ähnlichkeit der 3D-Struktur

Twilight-Zone

- **Lernziele:**
 - (1) verstehe, wie Threading- und Homologiemodelle konstruiert werden
 - (2) wie gut (genau) sind Homologiemodelle?

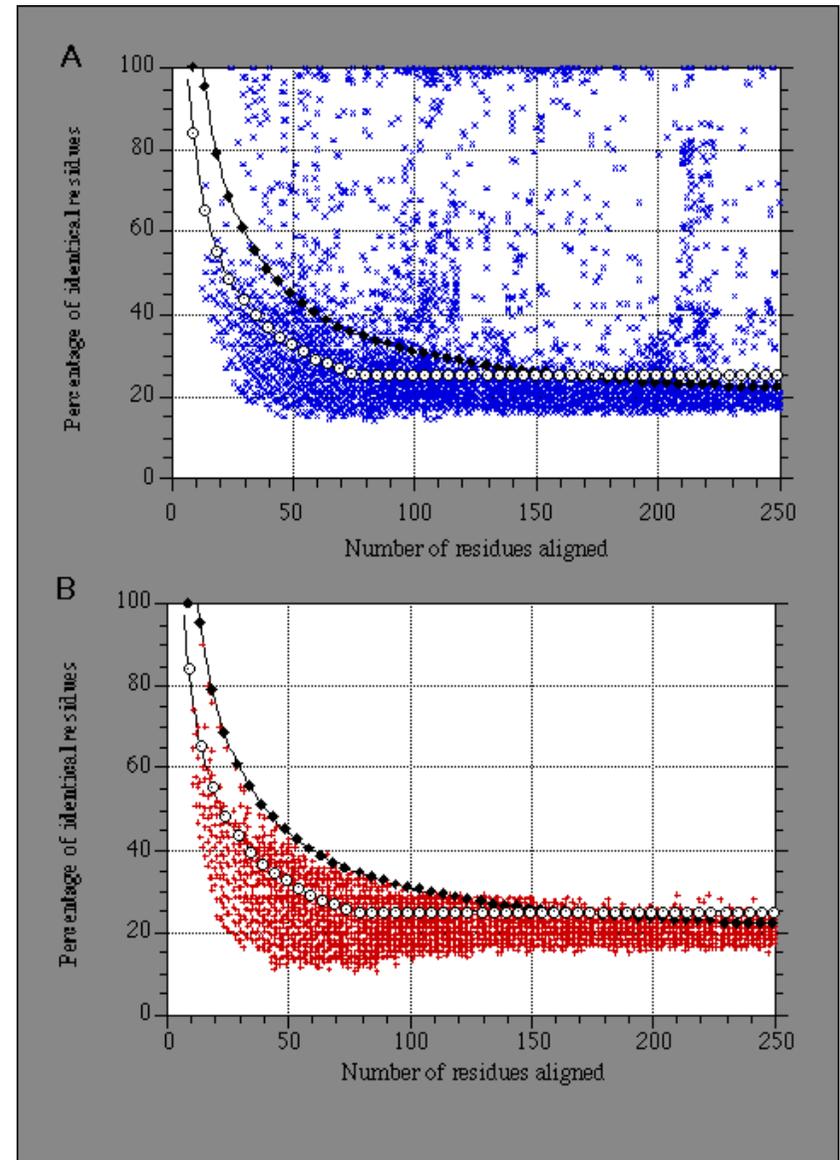
1 Twilight Zone

Die schwarzen Diamant-Symbole kennzeichnen eine Kurve, die als „**Twilight Zone**“ bezeichnet wird.

Paare von Proteinsequenzen mit größerer Sequenz-Identität als die Kurve haben mit Sicherheit eine ähnliche Struktur.

A „true positives“: Proteinpaare mit ähnlicher Struktur liegen sowohl oberhalb und unterhalb der Kurve, können also hohe oder niedrige Sequenzidentität haben.

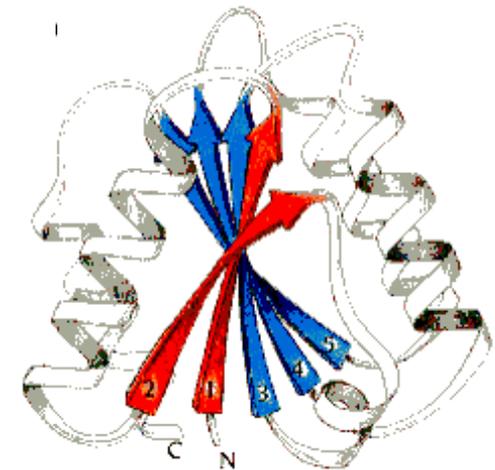
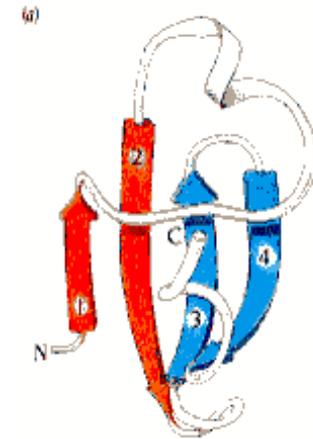
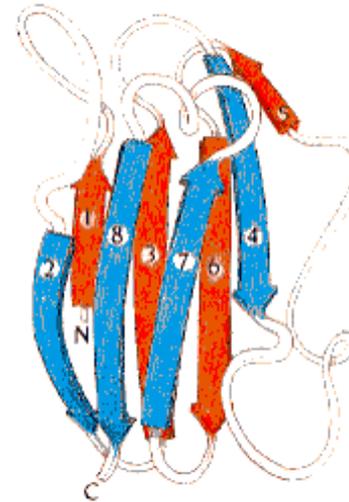
B: „false positives“: Strukturen, die keine bzw. wenig Übereinstimmung aufweisen, liegen stets unter der Kurve.



Rost, Prot. Eng. 12, 85 (1999)

2 Methode zur Fold-Erkennung: Threading

- Gegeben:
 - Sequenz:
IVACIVSTEYDVMKAAR...
 - Ein Datenbank von möglichen Proteinarchitekturen (“**fold**s”)
- Naive Idee: Bilde die Sequenz auf jeden fold ab
- Starte dabei bei jeder möglichen Position
- Bestimme anhand einer energetischen Bewertungsfunktion, welcher Fold am besten zu dieser Sequenz passt.



3 Sequenz-Profil

Profil: Sequenzpositionsspezifische Bewertungsmatrix $M(p,a)$ mit 21 Spalten und N Reihen.

- Reihe p entspricht einer bestimmten Position in den N_R alignierten Inputsequenzen.
- Die ersten 20 Spalten enthalten jeweils die Bewertung dafür, an dieser Position eine der 20 Aminosäuren zu finden.

Eine Extraspalte enthält einen Bestrafungsterm für Insertionen oder Deletionen.

Frequenz $W(p,b)$ für das Auftreten der Aminosäure b an Position p :

$$W(p,b) = c \log (n(b,p) / N_R) \text{ oder } n(b,p) / N_R$$

$n(b,p)$: beobachtete Häufigkeit der Aminosäure b an Position p in den N_R Inputsequenzen;

setze außerdem $n(b,p) = 1$ für jede Aminosäure, die nie in p auftritt.

Berechne $M(p,a)$ aus der Frequenz $W(p,b)$ und einer Austauschmatrix $Y(a,b)$
(PAM/BLOSUM)

$$M(p,a) = \sum_{b=1}^{20} W(p,b) \times Y(a,b),$$

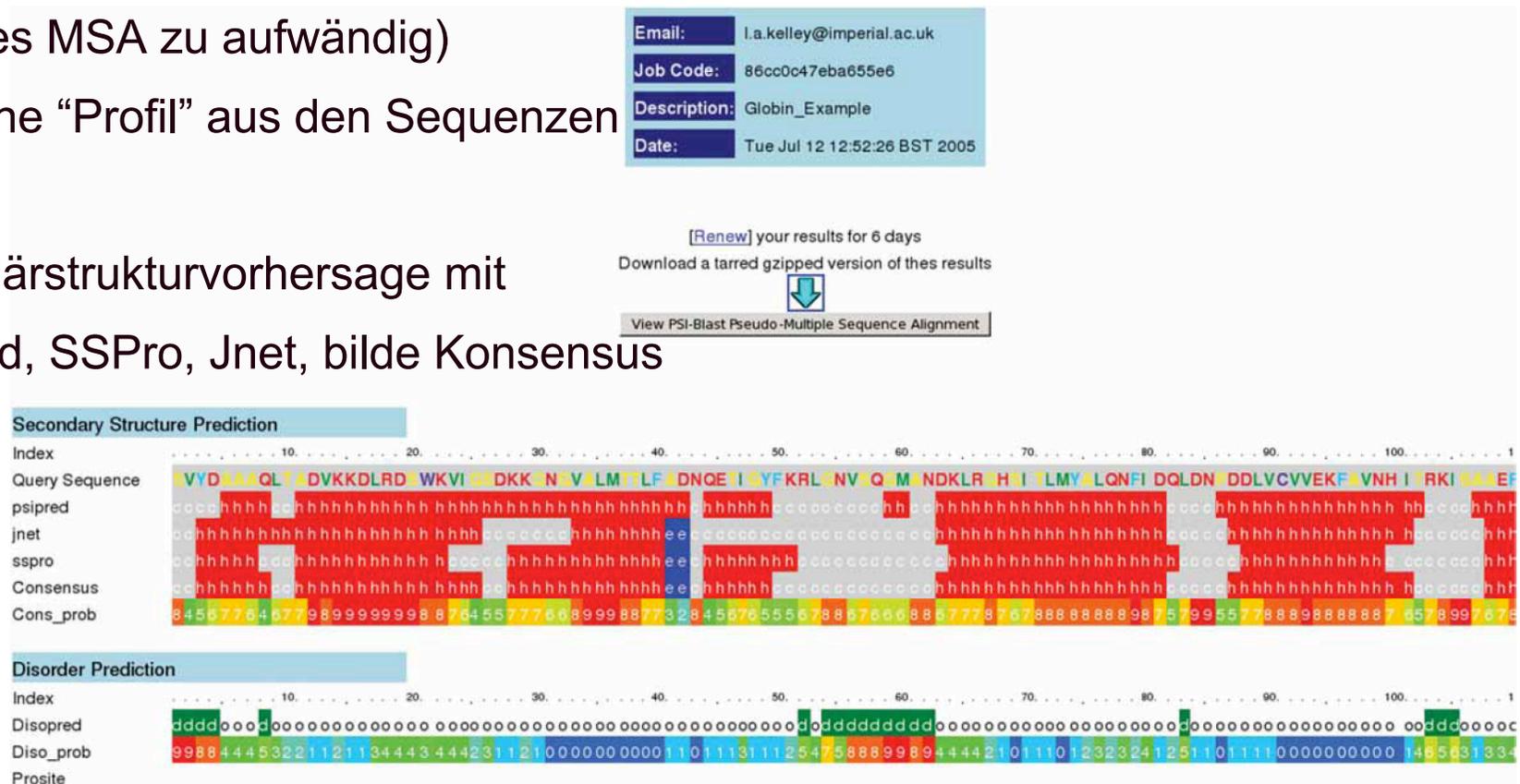
Gribskov, PNAS 84, 4355 (1987)

POS	PROBE	CONSENSUS	PROFILE																					
			A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	+/-	
1	EGVL	V	3	-2	3	4	0	4	-1	3	-1	4	4	1	1	-2	1	2	6	-6	-2	9		
2	LLSP	L	2	-2	-2	-1	3	0	-1	3	-1	6	5	-1	3	0	-1	3	1	4	1	-1	9	
3	VVVV	V	2	2	-2	-2	2	2	-3	11	-2	8	6	-2	1	-2	0	2	15	-9	-1	9		
4	KEAT	A	6	-2	5	6	-5	4	1	0	5	-2	0	3	3	1	3	6	0	-6	-4	9		
5	APLP	P	6	-1	0	1	-2	2	0	1	0	2	2	0	8	2	0	2	3	-5	-4	9		
6	GGGG	G	7	1	7	5	-6	15	-1	-3	0	-4	-3	4	3	2	-3	6	4	2	-11	-7	9	
7	SSQE	D	4	-1	7	7	-6	7	2	-2	2	-3	-2	4	3	6	1	6	2	-1	-6	-5	9	
8	SSTP	S	4	4	2	2	-4	4	-1	0	2	-3	-2	2	7	0	1	10	6	0	-2	-4	9	
9	VLVA	V	5	0	-1	-1	3	1	-2	7	-2	7	6	-1	1	-1	-3	0	2	10	-5	-1	9	
10	KRRS	R	0	-1	1	1	-5	0	2	-2	8	-3	1	3	3	3	10	5	1	-2	7	-5	9	
11	MLII	I	0	-2	-3	-2	7	-3	-3	11	-1	11	10	-2	-2	-1	-2	-2	1	9	-3	1	9	
12	SSTS	S	4	6	2	2	-3	5	-1	0	2	-3	-2	3	4	-1	1	12	6	0	0	-4	9	
13	CCCC	C	3	15	-5	-5	-1	2	-1	3	-5	-8	-6	-3	1	-6	-3	7	3	3	-13	10	9	
14	KSQR	K	1	-2	3	3	-6	1	3	-2	7	-3	0	3	3	5	7	4	1	-2	2	-5	9	
15	AAGS	A	10	3	4	3	-5	8	-1	-1	1	-2	-1	3	4	1	-2	7	4	2	-6	-4	9	
16	TSDS	S	4	3	5	4	-5	6	0	0	2	-3	-2	4	3	1	1	9	6	0	-3	-4	9	
17	GGSQ	G	5	1	6	5	-6	9	1	-2	1	-3	-2	4	3	4	0	6	3	0	-6	-6	9	
18	YFLS	F	-1	2	-4	-3	9	-3	0	4	-3	6	3	-1	-3	-3	-3	1	-1	2	7	7	9	
19	TTRL	T	1	-2	0	1	0	0	0	2	2	2	3	1	1	1	3	1	7	2	1	-2	9	
20	FF.L	F	-2	-3	-6	-4	10	-4	-1	6	-4	9	6	-3	-4	-4	-3	-2	-1	3	7	8	4	
21	SS.D	S	3	2	5	4	-4	5	0	-1	2	-3	-2	4	3	1	1	8	2	-1	-2	-3	4	
22	S.SS	S	2	3	1	1	-2	3	-1	0	1	-2	-1	2	2	0	1	8	2	0	1	-2	4	
23	. . .G	G	2	0	2	1	-2	4	0	0	0	-1	-1	1	1	1	-1	2	1	1	-3	-2	4	
24	. . .D	D	1	-1	4	3	-2	2	1	0	1	-1	-1	2	1	2	0	1	1	0	-3	-1	4	
25	. . .G	G	2	0	2	1	-2	4	0	0	0	-1	-1	1	1	1	-1	2	1	1	-3	-2	4	
26	.AGN	A	6	0	4	3	-4	6	1	-1	1	-2	-1	5	2	2	-1	3	3	1	-5	-3	4	
27	YNYT	Y	0	5	0	-1	5	-1	2	1	-1	0	-1	4	-3	-2	-2	0	3	0	3	6	4	
28	EDDY	D	2	-2	9	8	-3	3	4	-1	1	-3	-2	5	-1	4	-1	1	1	-1	-6	0	9	
29	LMAL	L	3	-5	-3	-1	6	-1	-2	6	-1	10	10	-2	0	0	-2	-1	0	6	-1	0	9	
30	YNAW	N	4	1	3	2	0	2	3	-1	1	-1	-1	8	0	1	-1	2	1	-1	-1	2	9	
.
48	SGNS	S	4	3	5	3	-4	7	0	-2	2	-4	-3	6	3	1	0	10	3	0	-2	-4	9	
49	SSNY	S	2	5	2	1	1	2	1	0	1	-2	-2	5	1	-1	0	8	1	-1	3	1	9	

Berücksichtige, dass aus den beobachteten Sequenzen durch Mutation alle 20 AS entstehen könnten. Die Häufigkeit davon wird durch die Austausch-Matrix $Y(a,b)$ ausgedrückt.

4 Methode zur Fold-Erkennung: Phyre2 webserver

- Webserver verwendet repräsentative Bibliothek für bekannte folds
- Lese Eingabesequenz mit unbekannter Struktur
- 5 Iterationen mit PsiBlast; finde nah und fern verwandte Sequenzen (richtiges MSA zu aufwändig)
- Berechne "Profil" aus den Sequenzen
- Sekundärstrukturvorhersage mit Psi-Pred, SSPro, Jnet, bilde Konsensus

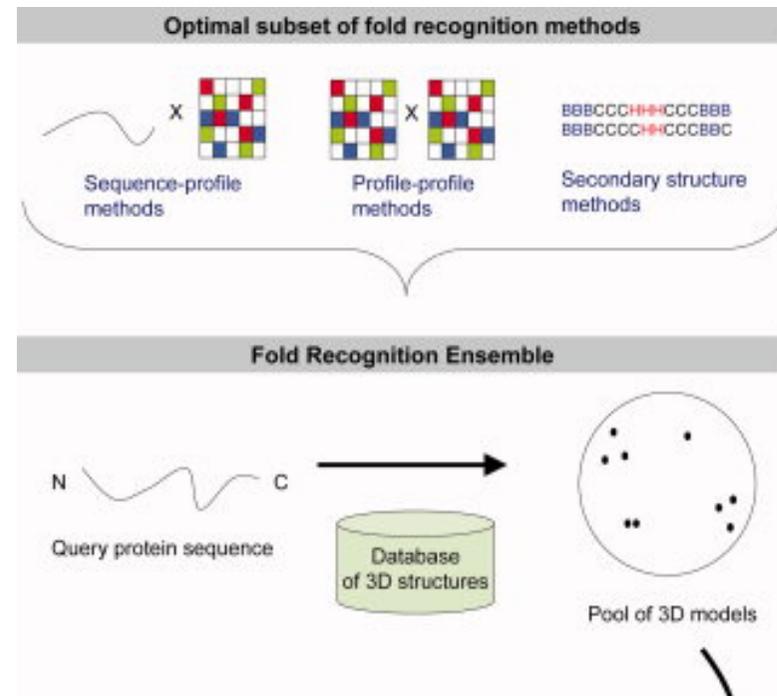


+ Vorhersage ungeordneter Regionen

Kelley, Nature Protocols 4, 363 (2009)

Methode zur Fold-Erkennung: Phyre2 webserver

- **Profile-Profile Alignment** zwischen Profil für Eingabesequenz und Profilen für Strukturfolds
- Berücksichtige auch, wie gut die vorhergesagte Sekundärstruktur zu jeder 3D-Strukturvorlage passt
- Berechne Scores für Passung zu allen 3D-Strukturen in der "fold library"
- Konstruiere komplette Strukturen für die 10 besten Scores
- Ergibt manchmal sehr gute Strukturmodelle bei 15-25% Sequenz-Identität.



Fold Recognition							
View Alignments	SCOP Code	View Model	E-value	Estimated Precision	BioText	Fold/PDB descriptor	Superfamily
	c3baehg (length:145) 100% i.d.		9.3e-20	100 %	0.90 BioText	Globin-like	Globin-like
	c2bk9A (length:153) 23% i.d.		7.7e-17	100 %	0.89 BioText	PDB header:oxygen transport	Chain: A: PDB Molecule:cg9734-pa;

Bennet-Lovsey, Proteins 70, 611 (2008)

5 Homologie-basierte Proteinmodellierung (SwissModel)

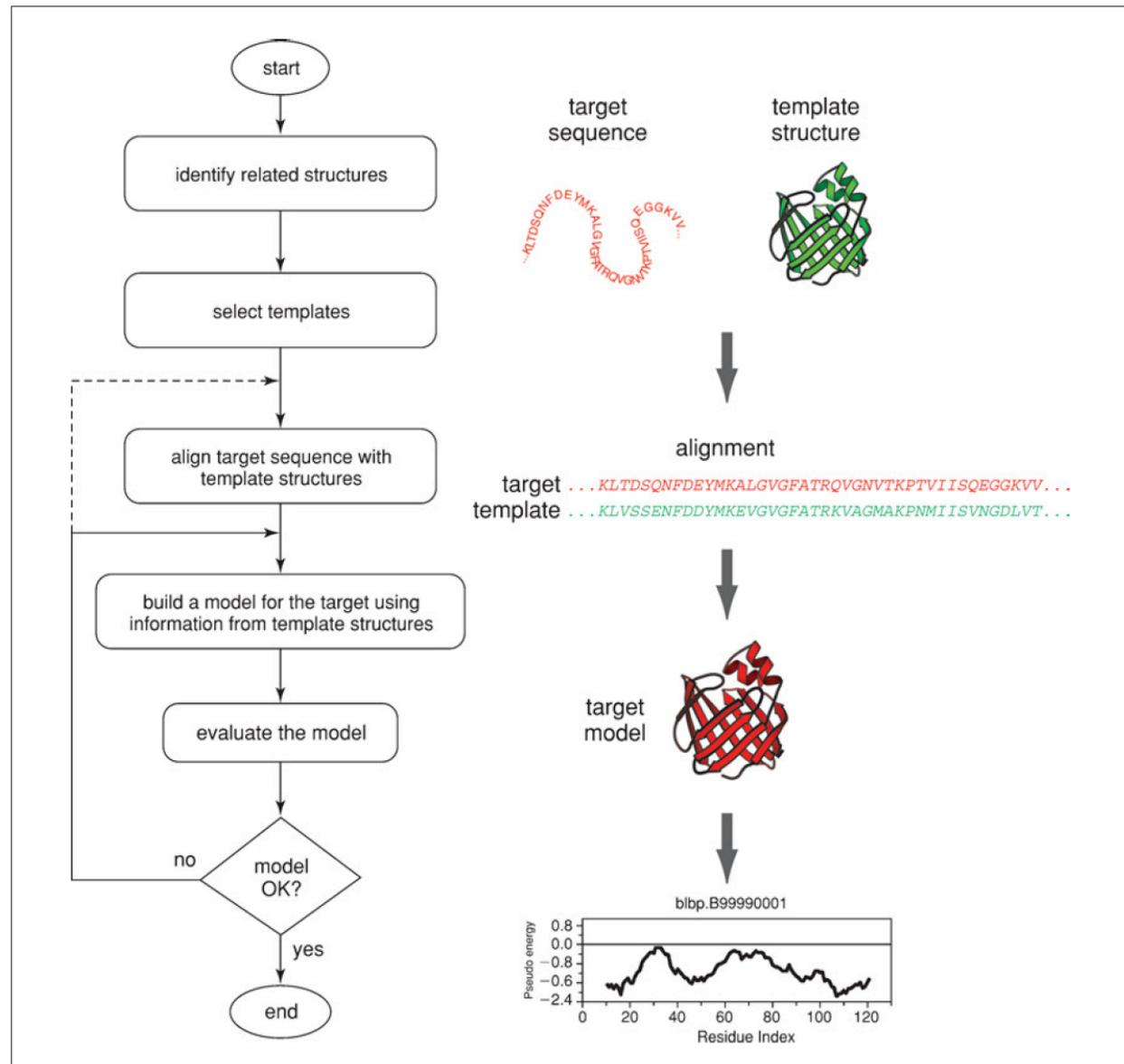
- **Methode:** Ebenfalls wissenschaftlicher Ansatz.
- **Erfordernis: Mindestens** 1 bekannte 3D-Struktur eines verwandten Proteins,
- **Prozedur:**
 - finde Proteine bekannter Struktur, die zu Inputsequenz verwandt sind.
 - Erzeugung eines multiplen Sequenzalignments mit der Zielsequenz.
 - Generierung eines **Frameworks** für die neue Sequenz.
 - Konstruiere fehlende **Loops**.
 - Vervollständige und korrigiere das **Proteinrückgrat**.
 - Korrigiere die **Seitenketten**.
 - Überprüfe die **Qualität** der modellierten Struktur und deren Packung.
 - Strukturverfeinerung durch Energieminimierung und Moleküldynamik.

Homologie-basierte Proteinmodellierung (Modeller)



Andrej Sali, UCSF

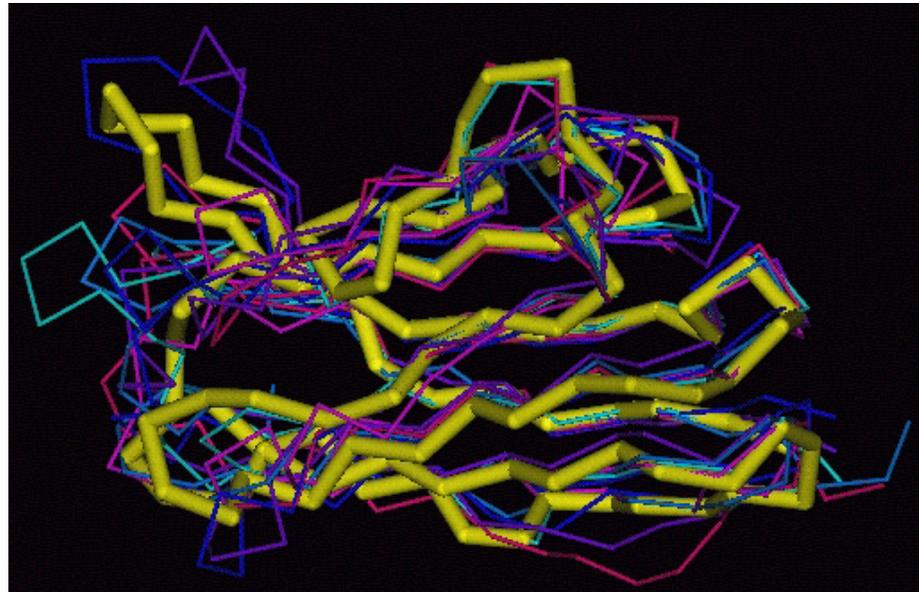
<http://salilab.org>



Eswar, Curr. Protocols in Bioinf. (2006)

3D Framework für die neue Sequenz

- (a) Für alle Atome, die eine ähnliche Position besitzen und vermutlich eine strukturelle Entsprechung in der neuen Struktur besitzen, werden gemittelte Positionen als Framework-Koordinaten bestimmt.
- (b) Seitenketten mit völlig inkorrekturer Geometrie werden entfernt.
- (c) Matrix mit Gewichten für lokale Ähnlichkeit.



www.expasy.org/swissmodel/SWISS-MODEL.html

Konstruktion fehlender Loops

Konformationen für strukturell abweichende Loops zu konstruieren, ist ein ernstes Problem bei der vergleichende Modellierung. Seine Lösung ist (noch) offen.

Dies gilt nicht nur für lange Loops, in denen zahlreiche Mutationen auftraten, sondern auch für kurze Loops im Fall von Insertionen und Deletionen.

Sobald das Alignment von Zielsequenz und der Vorlagesequenz vorliegt, sollte man überprüfen, ob die eingefügten Gaps außerhalb von Sekundärstrukturelementen in der 3D-Struktur der Vorlage liegen.

Ein paar Regeln:

- bei sehr kurzen Loops können wir Daten über beta-turns verwenden

Beta-Turns

Eine Aminosäurekette kann ihre Richtung dadurch umkehren, daß ein „reverse turn“ durch Bildung einer H-Bindung zwischen C=O und H-N gebildet wird.

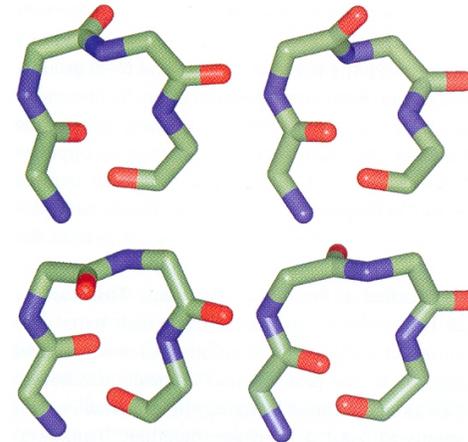


Figure 1.9 The four types of beta turn described in Table 1.1, types I and I' are shown on the top, types II and II' on the bottom.

Wenn dies zwischen zwei antiparallelen beta-Strängen geschieht, nennt man diesen einen beta-Haarnadel (hairpin). Es ergeben sich folgende Diederwinkel:

Table 1.1 Turns are regions of the protein chain that enable the chain to invert its direction. The ϕ and ψ angles of some commonly occurring turns are listed.

Turn type	ϕ_1	ψ_1	ϕ_2	ψ_2
I	-60	-30	-90	0
I'	60	30	90	0
II	-60	120	80	0
II'	60	-120	-80	0

[Tramontano book]

Konstruktion fehlender Loops

Ein paar Regeln:

-falls mittellange Loops kompakte Substrukturen bilden, spielt die Ausbildung von Wasserstoffbrückenbindungen mit den Atomen des Rückgrats die wichtigste Rolle für die Konformation

- falls mittellange Loops ausgedehnte Konformationen haben, ist für ihre Stabilisierung meistens eine hydrophobe Seitenkette verantwortlich, die ins Proteininnere zeigt und zwischen die Sekundärstrukturelemente gepackt ist zwischen denen der Loop liegt.

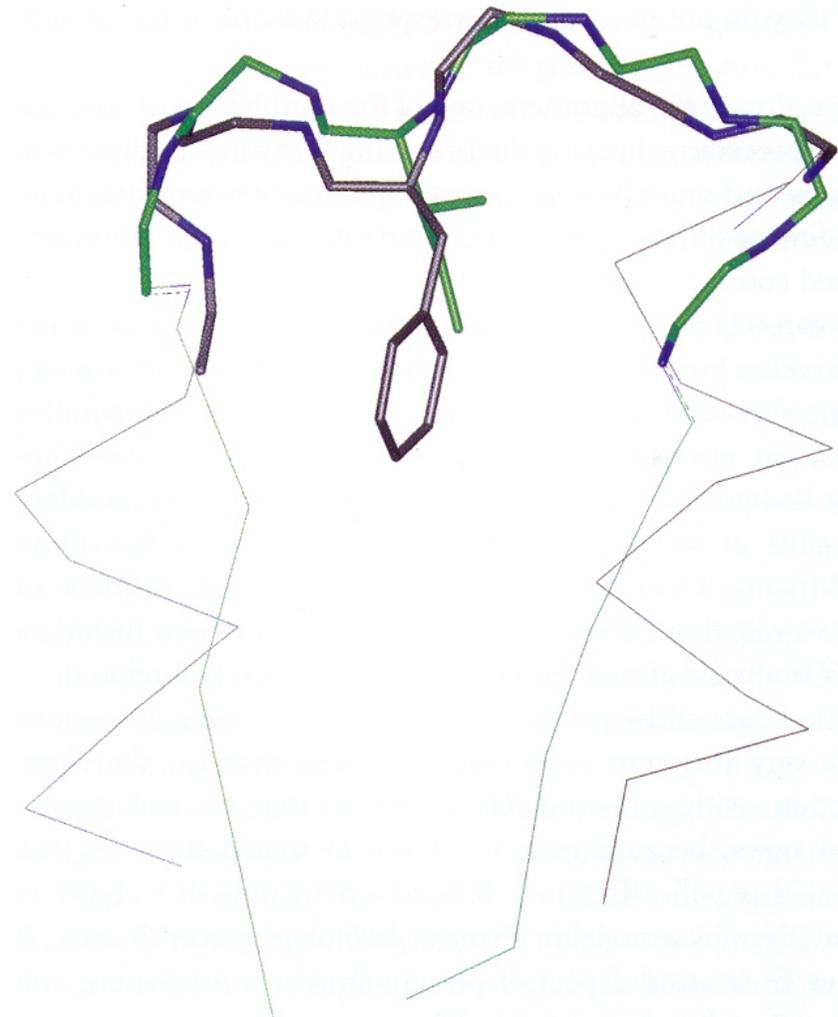


Figure 4.16 The figure shows two loops with similar conformations stabilized by the packing of a central hydrophobic amino acid. Note that one of the loops connects two alpha helices and the other two beta strands.

[Tramontano book]

Konstruktion fehlender Loops

Sehr ähnliche Konformation dreier Loops mit unterschiedlicher Sequenz.

Zwei Loops enthalten ein *cis*-Prolin.

Die stabilisierenden H-Bindungen werden mit sehr unterschiedlichen Proteingruppen ausgebildet.

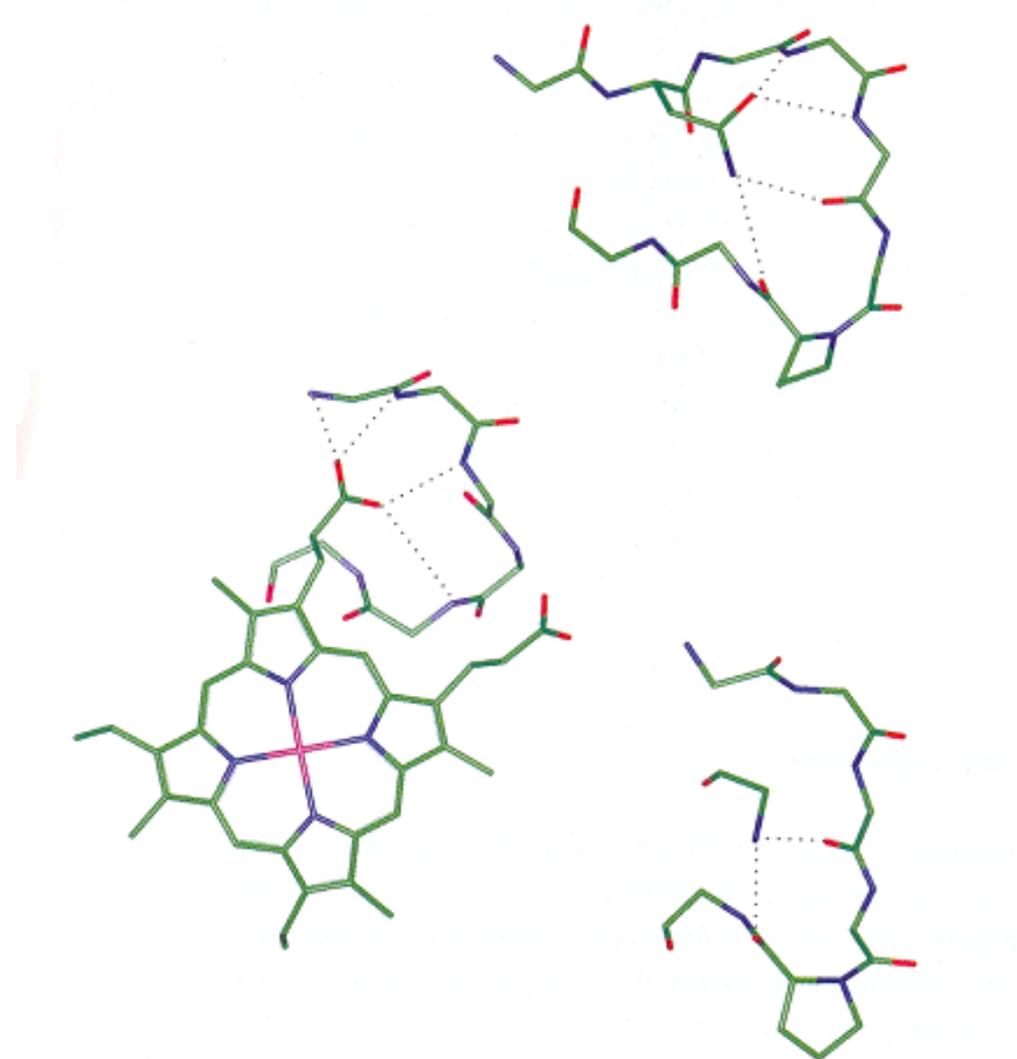


Figure 4.17 The three loops shown in the figure are very similar and stabilized by hydrogen-bonds, however the partners of these interactions are different in the three different proteins (an immunoglobulin, a viral protein, and a cytochrome).

[Tramontano book]

Konstruktion fehlender Loops

Basierend auf den Verankerungen der Loops

(a) wird entweder eine Datenbank bekannter Loopfragmente in der PDB-Datenbank durchsucht.

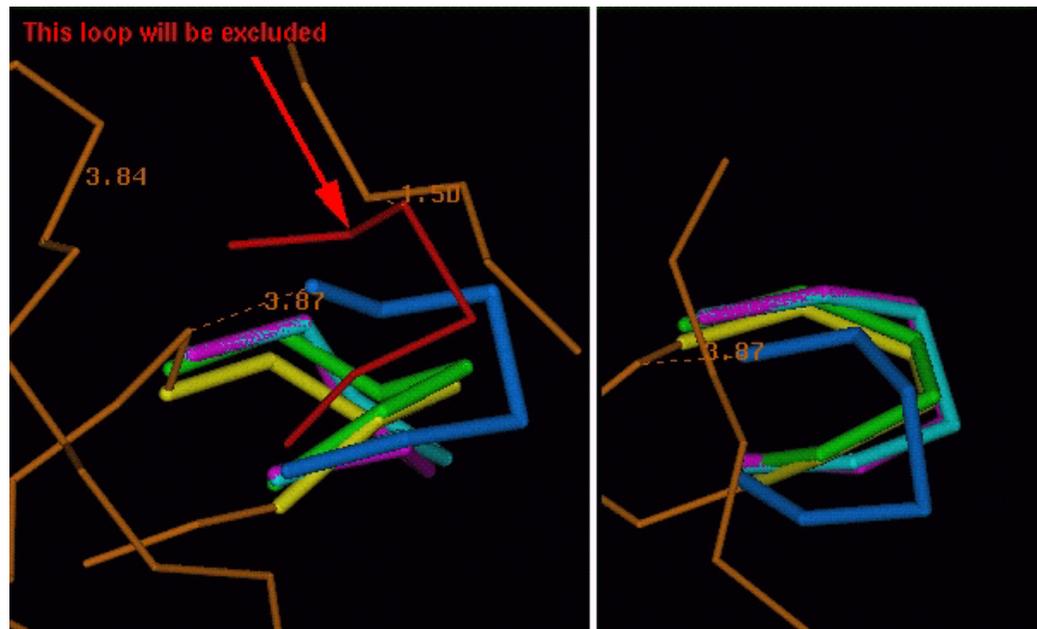
Für den neuen Loop verwendet man dann entweder das am besten passende Fragment oder ein Framework aus den 5 besten Fragmenten.

(b) oder es wird der Torsionsraum der Loopresiduen durchsucht

- 7 erlaubte Kombinationen der Φ - Ψ Winkel

- benötigter Raum für den gesamten Loop

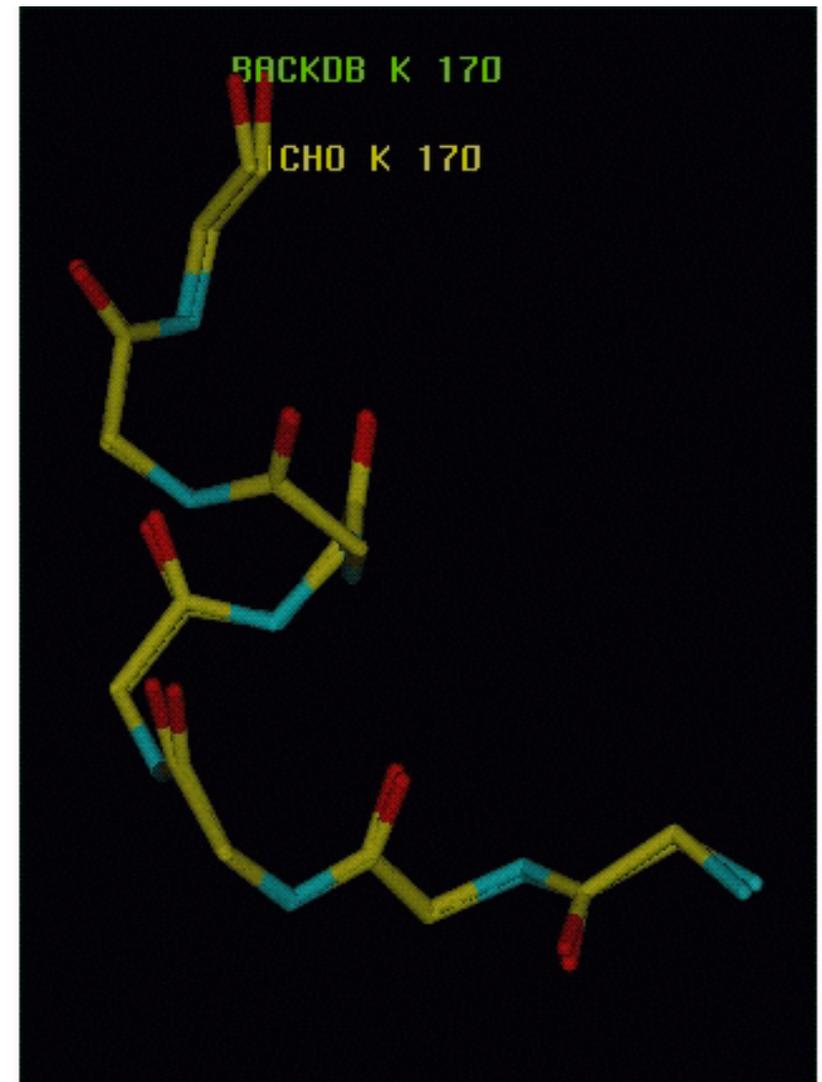
www.expasy.org/swissmodel/SWISS-MODEL.html



Rekonstruktion von fehlendem Proteinrückgrat

Das Rückgrat wird auf der Grundlage von C_{α} -Positionen konstruiert.

- 7 Kombinationen der Φ - Ψ Winkel sind erlaubt.
- Durchsuche Datenbank für Backbone-Fragmente mit Fenster aus 5 Residuen, Verwende die Koordinaten der 3 zentralen Residuen des am besten passenden Fragments.



www.expasy.org/swissmodel/SWISS-MODEL.html

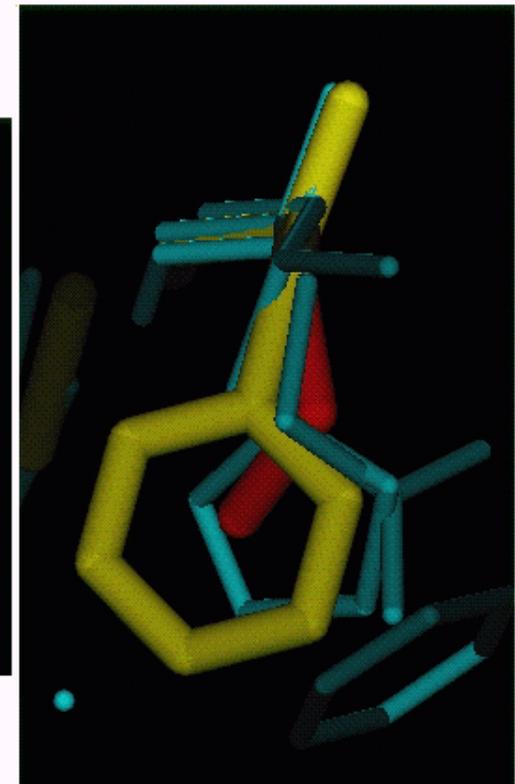
Konstruktion unvollständiger/fehlender Seitenketten

Ponder & Richards (1987): einige Aminosäuren bevorzugen bestimmte Winkelbereiche für ihre Seitenkettenwinkel → Rotamerbibliotheken.

Verwende Bibliothek erlaubter Seitenketten-Rotamere geordnet nach der Häufigkeit des Auftretens in der PDB-Datenbank.

- Erst werden verdrehte (aber komplette) Seitenketten korrigiert.
- fehlende Seitenketten werden aus der Rotamer-Bibliothek ergänzt.

Teste dabei, ob van-der-Waals Überlapps auftreten und ob die Torsionswinkel in erlaubten Bereichen liegen.



www.expasy.org/swissmodel/SWISS-MODEL.html

Rotamer-Bibliotheken: günstige Diederwinkel

Günstige χ_1 -Drehwinkel der Valin-Seitenkette:

beobachtete Häufigkeit der Rotamere

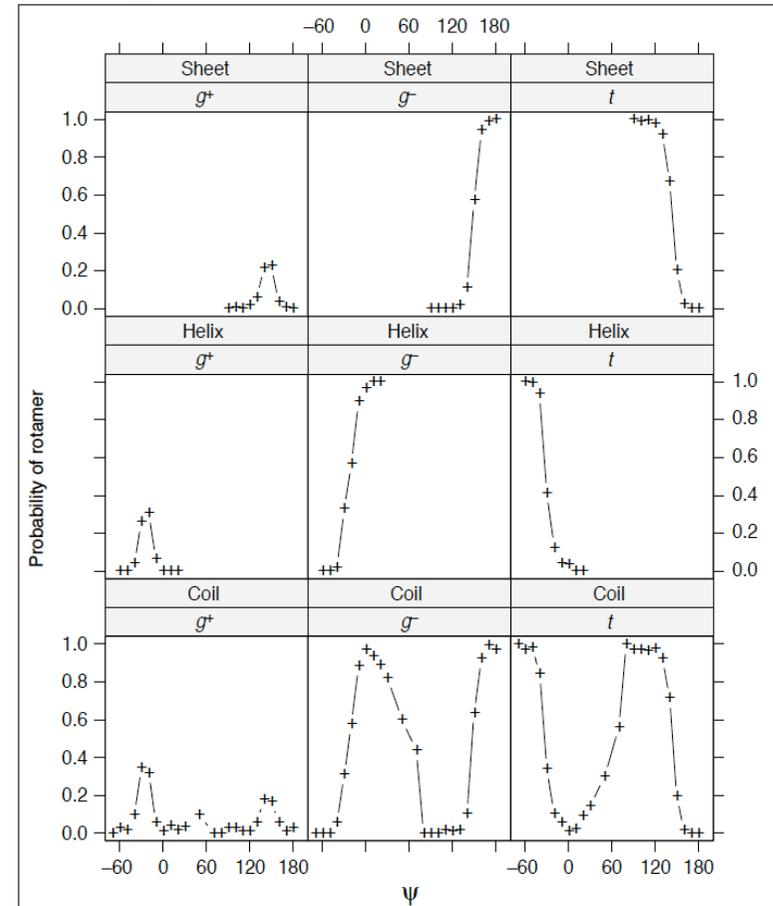
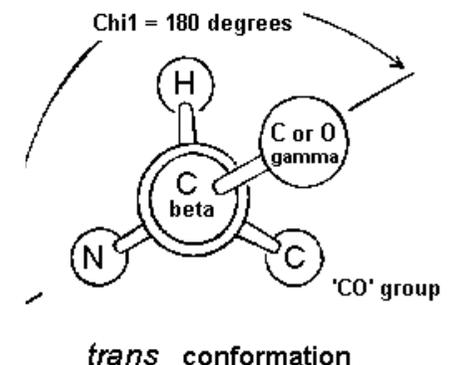
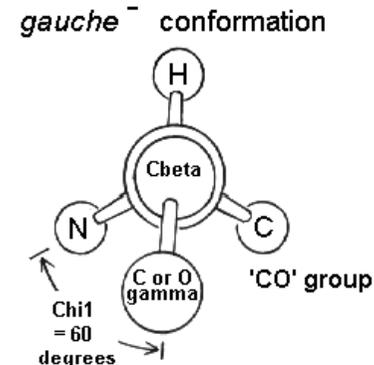
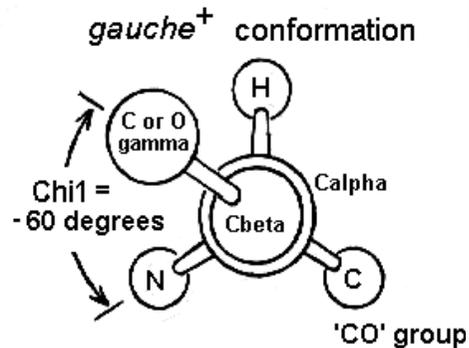
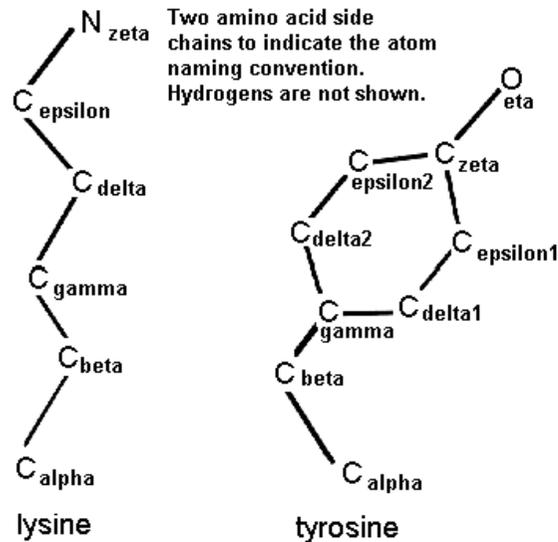
gauche⁺ ($\chi_1 \sim +60^\circ$)

gauche⁻ ($\chi_1 \sim -60^\circ$)

trans ($\chi_1 \sim 180^\circ$)

in verschiedenen Sekundärstrukturen

als Funktion des Rückgratsdiederwinkels Ψ .



R. Dunbrack (2002) Curr.Opin.Struct.Biol. 12, 431

<http://swissmodel.expasy.org/course/text/chapter3.htm>

Paarungs-Präferenz von Aminosäuren

Bei der Orientierung der Seitenketten wird üblicherweise jede für sich betrachtet (Rotamer-Bibliothek berücksichtigt zwar die Konformation des Rückgrats, aber nicht die Umgebung).

Aminosäuren nehmen jedoch je nach Umgebung unterschiedliche Konformationen ein.

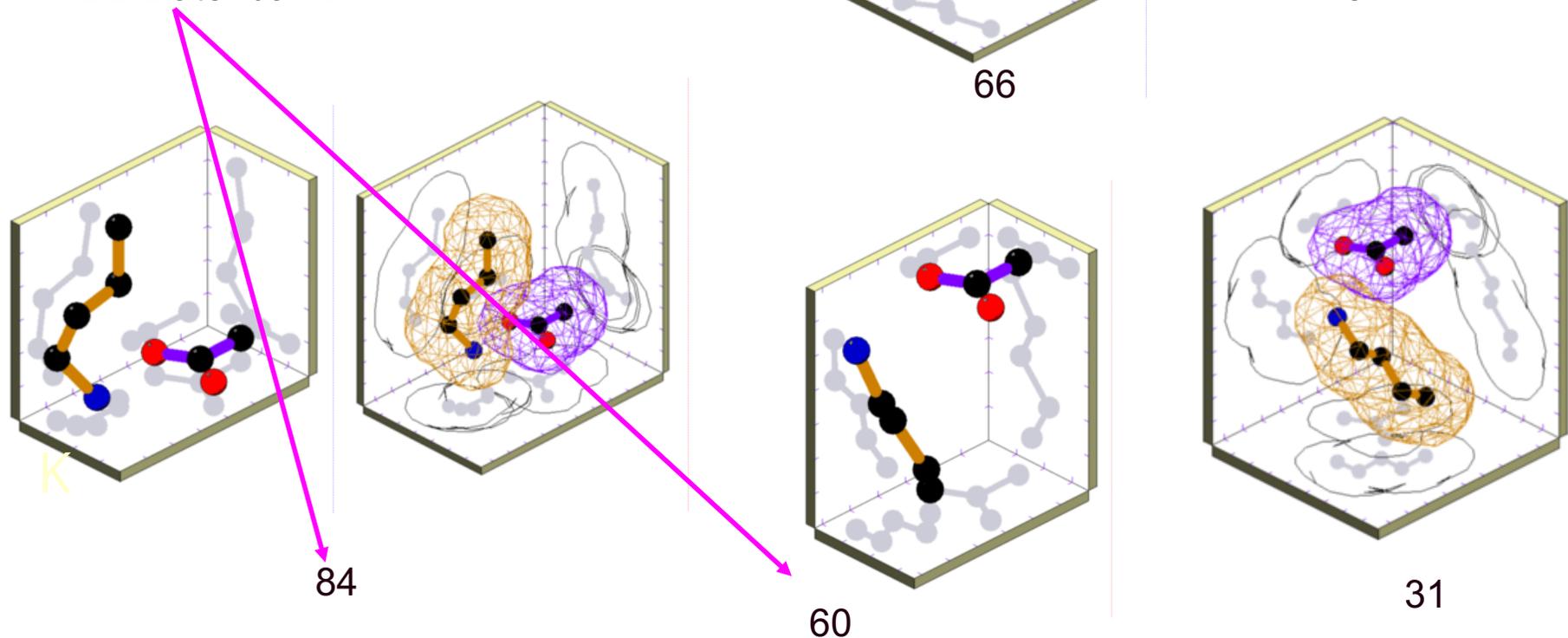
Diese “Packungseffekte” können ebenfalls für komparative Modellierung berücksichtigt werden.

Salzbrücken

Datenbank für statistische Präferenz für die Orientierung von Aminosäure-Seitenketten in der PDB-Datenbank:

<http://www.biochem.ucl.ac.uk/bsm/sidechains/index.html#>

Häufigkeit von 1845 Asp-Lys-Kontakte in PDB-Datenbank



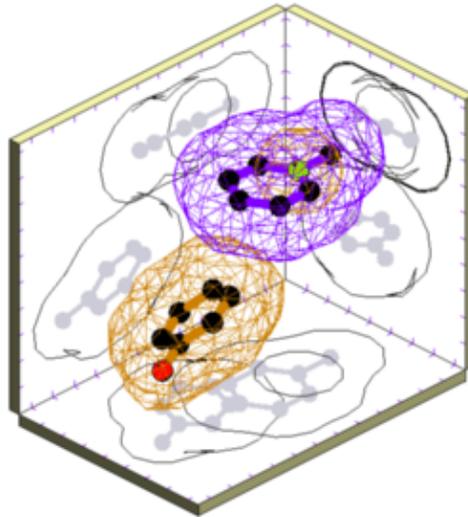
Π -stacking von aromatischen Ringen

Aromatische Ringe (z.B. Phenol, Benzol, Seitenketten von Tyrosin, Phenylalanin, Tryptophan, Histidin ...) besitzen delokalisiertes Elektronensystem ausserhalb der Ringebene.

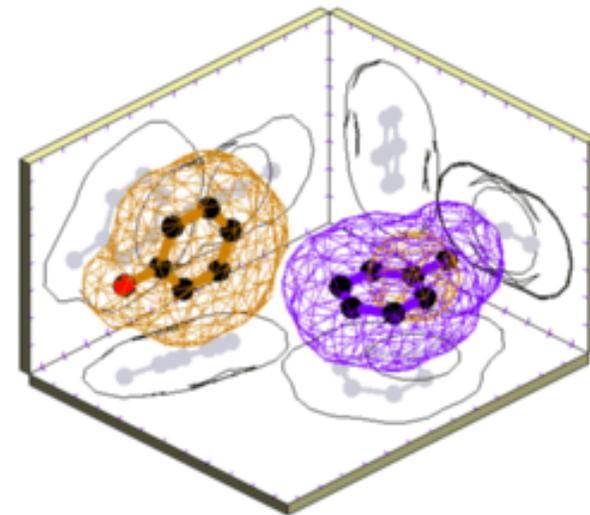
Mehrere dieser Ringe "packen" gerne aufeinander bzw. senkrecht zueinander.

Cluster Phe-Tyr

1



4

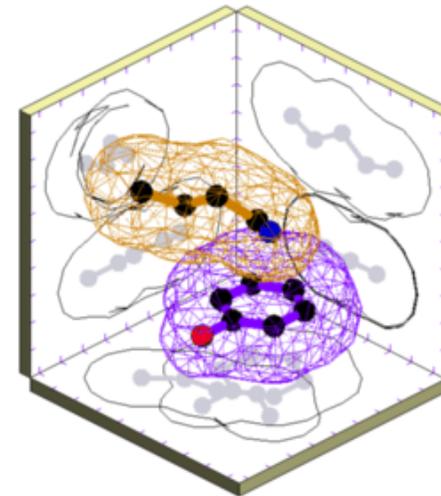
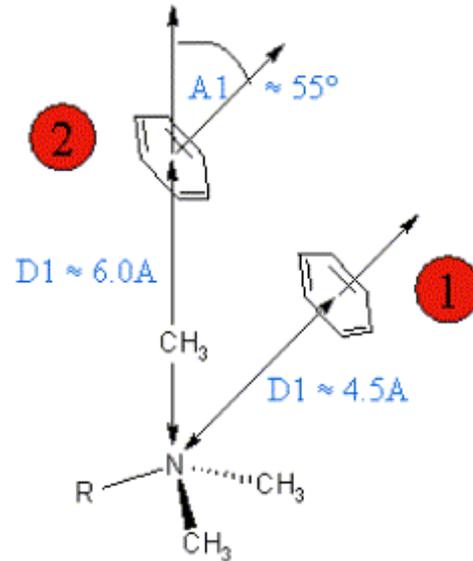


Kationen- π -Wechselwirkung

Die gleichen **aromatischen Ringe** wechselwirken gerne senkrecht zur Ringebene mit positiv geladenen Gruppen.

Beispiele: Acetylcholin in Bindungstasche von Acetylcholinesterase

K



Tyr-Lys
Cluster 6

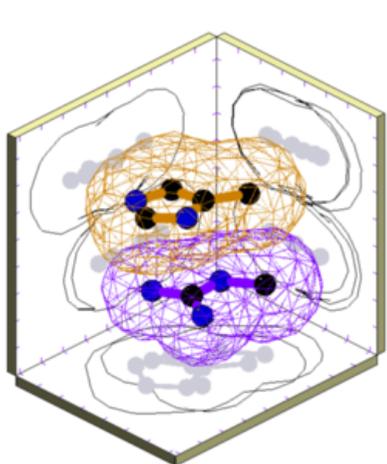
Bevorzugte Geometrien für die Wechselwirkung von Trimethyl-Ammoniumgruppen mit Phenyl-Ringen
Gohlke & Klebe, JMB 2000

Kationen- Π -Wechselwirkung

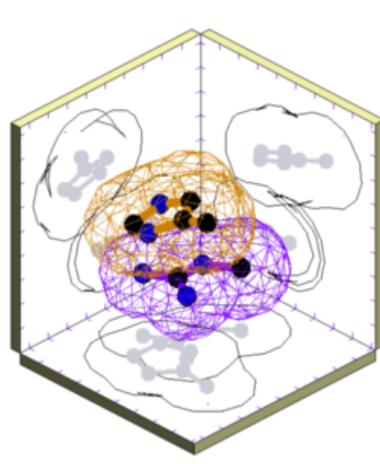
Wechselwirkung der positiv geladenen Guanidinium-Gruppe von Arg mit dem Π -Elektronensystem von His.

Fast immer planare Packung. Nur in Cluster 3 Ausbildung einer Wasserstoffbrücke N-H ... N

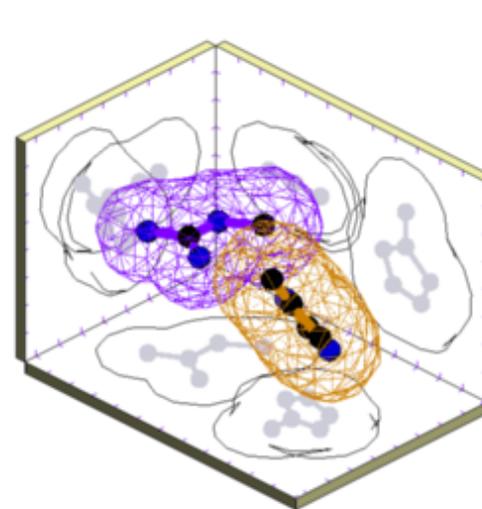
1



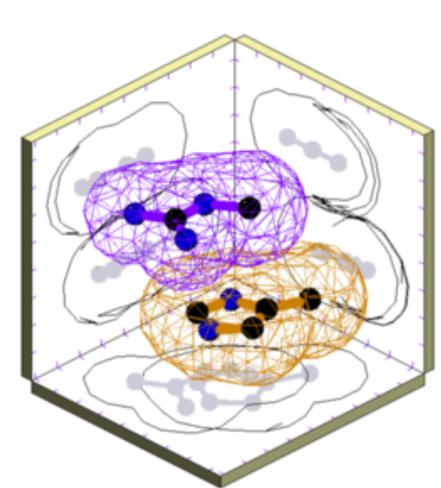
2



3



4



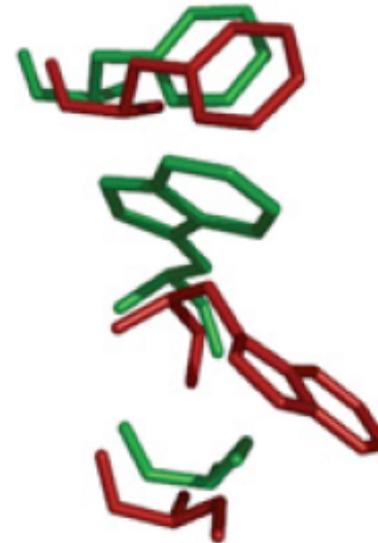
K

Typische Fehler bei Homologie-Modellierung (I)

(1) Fehlerhafte Packung der Seitenketten.

In rot gezeigt ist die Kristallstruktur des cellular retinoic acid binding protein I (CRAB1) aus Maus.

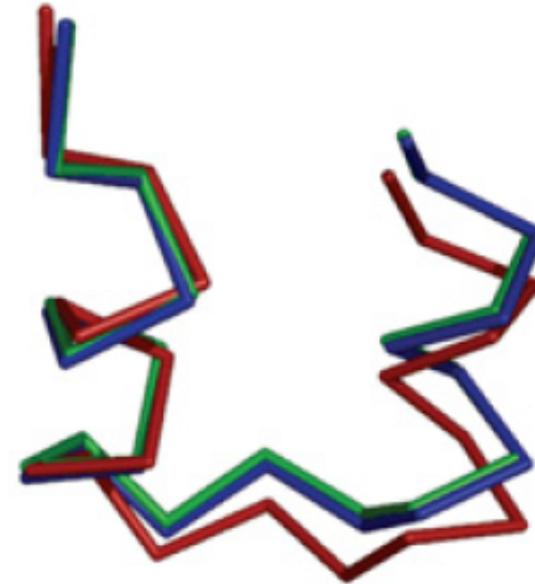
Die modellierte Struktur der Tryptophan Residue 109 (Mitte) ist in grün gezeigt.



Typische Fehler bei Homologie-Modellierung (II)

(B) Verschiebungen in korrekt alignierten Regionen.

Hier ergeben sich leichte Abweichungen des Modells des CRAB1 Proteins (grün) von der Kristallstruktur des CRAB1 (rot) entsprechend der Kristallstruktur des fatty acid binding protein (blau), das als Vorlage benutzt wurde.



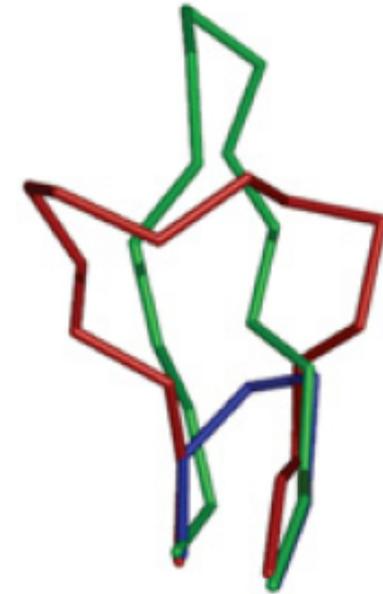
Eswar, Curr. Protocols in Bioinf. (2006)

Typische Fehler bei Homologie-Modellierung (III)

(C) Fehler in Regionen ohne Vorlage.

Gezeigt ist die Verbindung zwischen den $\text{C}\alpha$ -Atomen der Schleife 112–117 für

- die Kristallstruktur des menschlichen eosinophil neurotoxin (rot),
- dessen Modell (grün), und
- die Vorlagestruktur Ribonuclease A (blau).



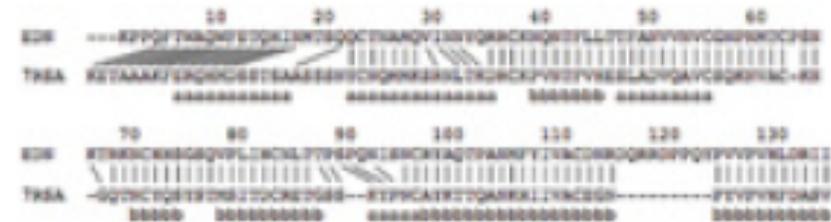
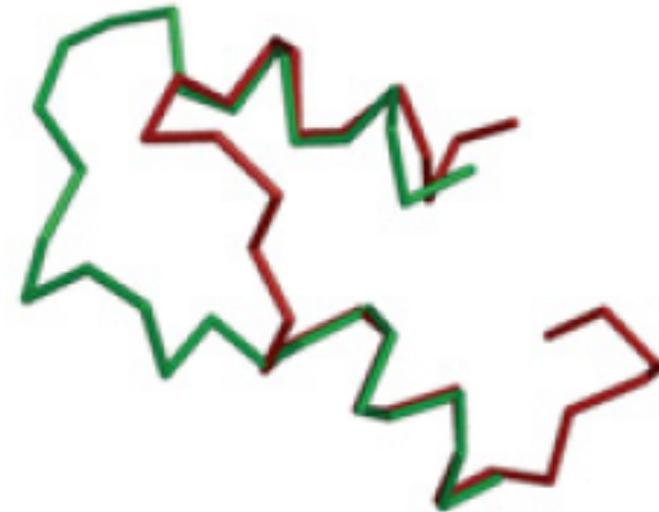
Eswar, Curr. Protocols in Bioinf. (2006)

Typische Fehler bei Homologie-Modellierung (IV)

(D) Fehler durch Misalignment.

N-terminale Region der Kristallstruktur von menschlichem eosinophil neurotoxin (rot) im Vergleich mit dem Modell (grün).

Der Fehler resultiert aus dem ungünstigen Alignment mit der Vorlage Ribonuclease A (unten).

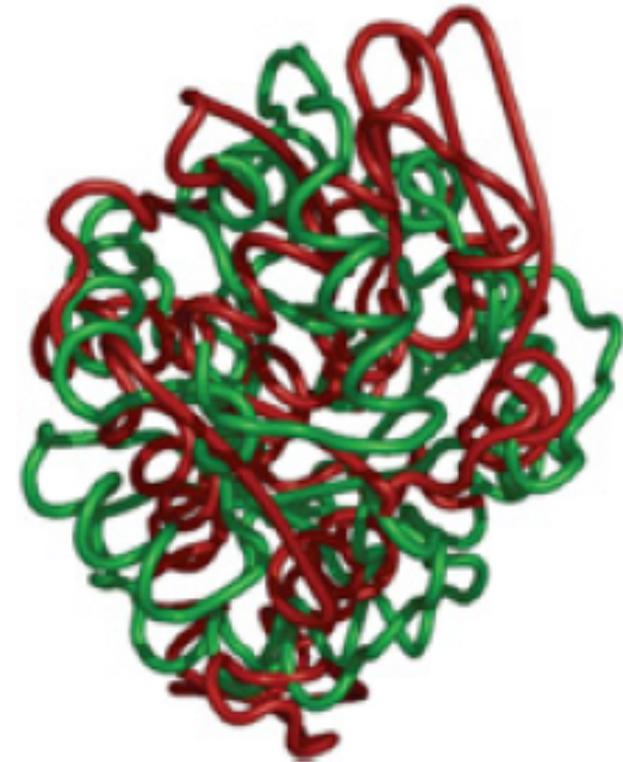


Eswar, Curr. Protocols in Bioinf. (2006)

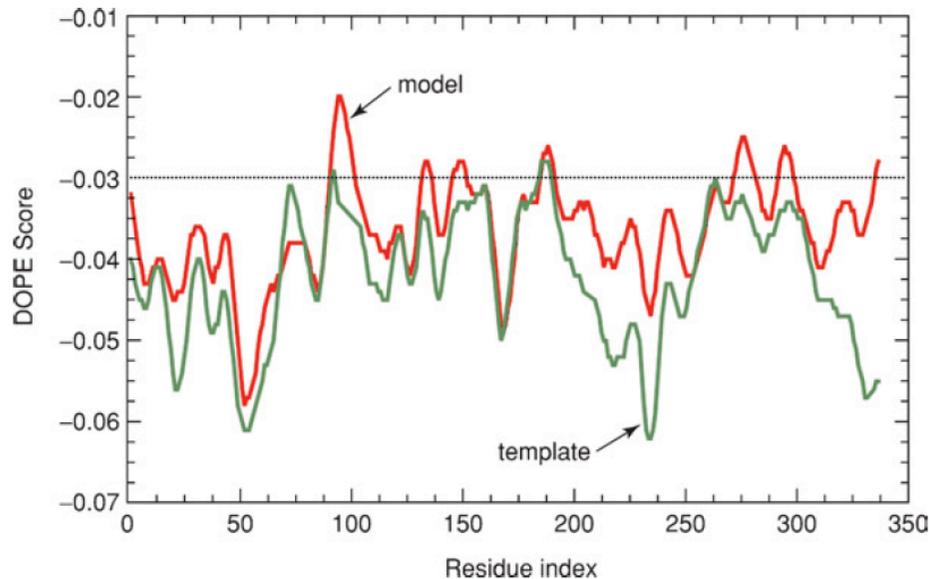
Typische Fehler bei Homologie-Modellierung (V)

(E) Fehler durch inkorrekte Vorlage.

Vergleich der Kristallstruktur für α -trichosanthin (rot) mit dem Modell (grün), das mit Indol-3-Glycerophosphat-Synthase als Vorlage erzeugt wurde..



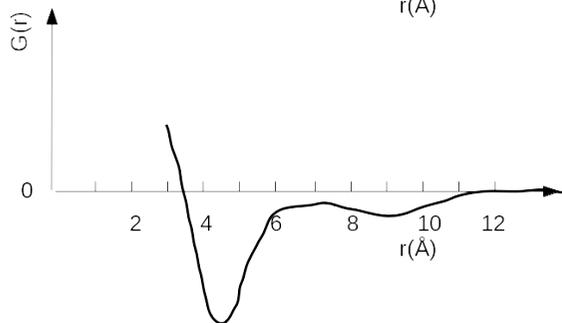
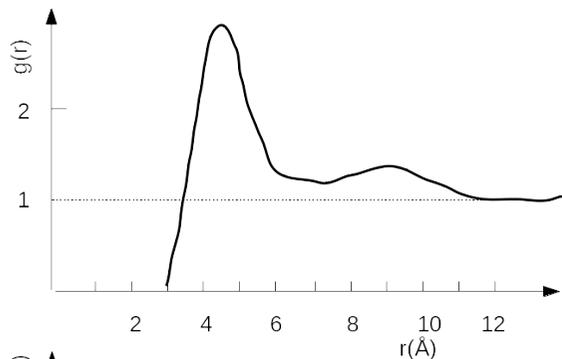
Bewertung von Strukturmodellen (Modeller)



Modeller verwendet das statistische DOPE-Potential (Discrete Optimized Protein Energy) zur Bewertung von Strukturmodellen.

Niedrigere Energien sind besser.

DOPE ist ein statistisches Potential für die Wahrscheinlichkeiten, wie häufig bei einem bestimmten Abstand das Atompaar $i - j$ in den bekannten Proteinstrukturen auftritt.



Statistisches Potential:

Aus den in Proteinstrukturen beobachteten radialen Häufigkeiten $g(r)$ für Aminosäure-Paare (-> *bei welchem Abstand gibt es mehr als erwartet?*), berechnet man durch „Boltzmann-Inversion“ deren effektive Wechselwirkungsstärke.

$$\frac{p_1}{p_2} = e^{-\frac{E_1 - E_2}{kT}} \longrightarrow G(r) = -k_B T \ln p(r)$$

Eswar, Curr. Protocols in Bioinf. (2006)

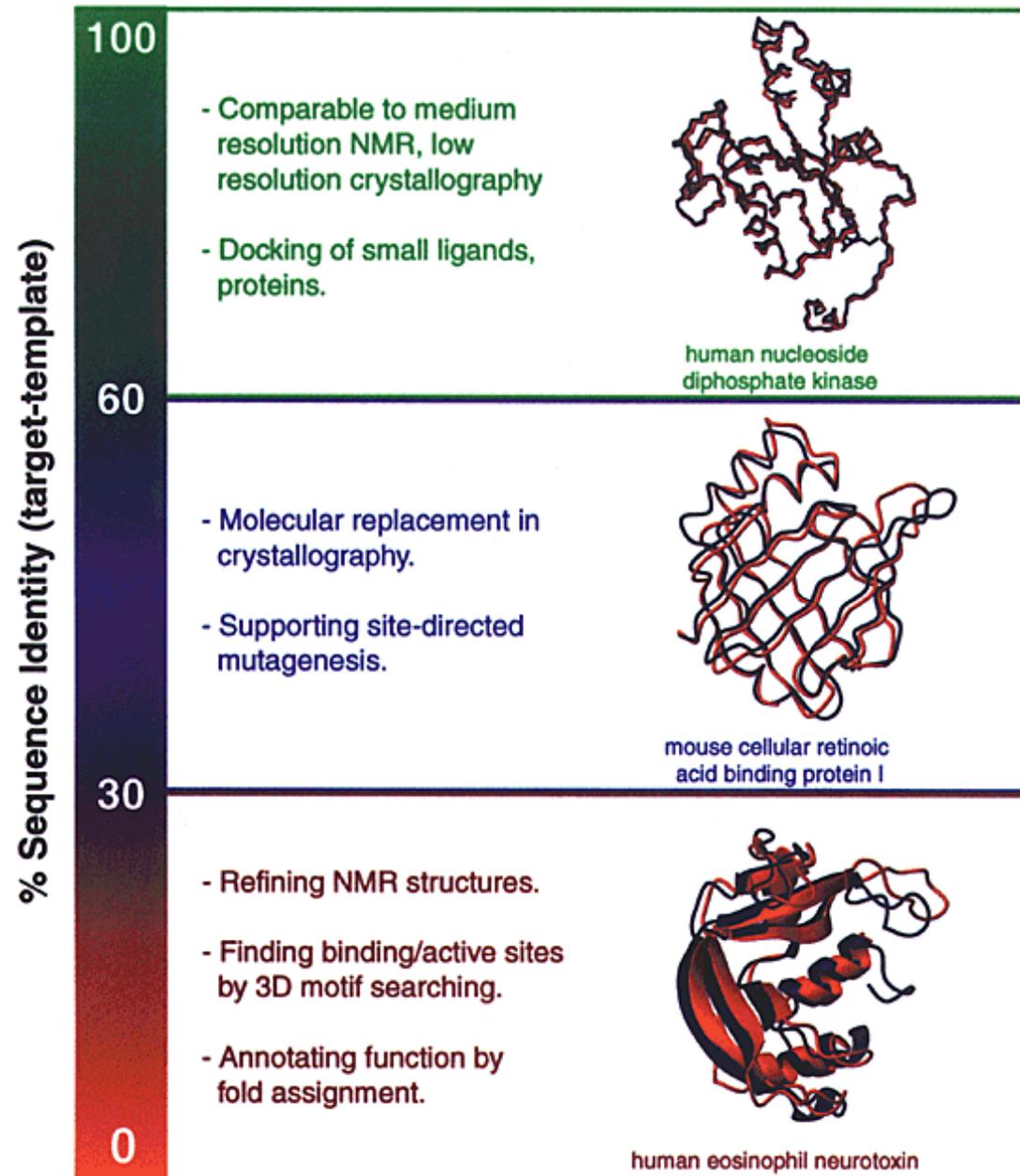
Homologie/Komperative Modellierung

Qualität der Modellierung hängt von Sequenzidentität mit Vorlage ab.

Man sollte stets beachten, dass die Vorlage nicht aus der Twilight Zone stammt.

Protein structure modeling for structural genomics.

R. Sánchez *et al.* Nat. Struct. Biol. 7, 986 - 990 (2000)



Bewertung der Qualität eines Homologiemodells - Allgemeine Gesichtspunkte

- Ein Modell wird als **falsch** angesehen, wenn mindestens eines seiner strukturellen Elemente gegenüber dem Rest des Modells falsch angeordnet ist. Dies kann durch ein falsches Sequenzalignment entstehen. Das Modell kann dennoch korrekte Stereochemie besitzen.
- Man kann ein Modell als **ungenau** ansehen wenn seine atomare Koordinaten mehr als 0.5 Å von einer experimentellen Kontrollstruktur abweichen.
- Ungenauigkeiten können auch in der Stereochemie (Bindungslängen und –winkel auftreten). Dies kann leicht mit **WhatCheck** überprüft werden.
- **Statistische Paarpotentiale** für die Verteilung von Aminosäuren in bekannten Proteinen erlauben manchmal die Aufspürung von fehlerhaften Modellen.

www.expasy.org/swissmodel/SWISS-MODEL.html

Proteinkern und Loops

Fast jedes Proteinmodell enthält nicht-konservierte Loops, die als die am wenigsten zuverlässigen Teile des Proteinmodells angesehen werden können.

Andererseits sind diese Bereiche der Struktur oft auch am flexibelsten – hohe Temperaturfaktoren in Kristallstrukturen oder hohe Unterschiede zwischen verschiedenen (gleichsam gültigen) NMR-Strukturen.

Die Residuen im Proteinkern werden gewöhnlich fast in der identischen Orientierung wie in experimentellen Kontrollstrukturen modelliert.

Residuen an der Proteinoberfläche zeigen größere Abweichungen.

www.expasy.org/swissmodel/SWISS-MODEL.html

Vergleich zweier Strukturen: RMSD

Root mean square deviation:

$$RMSD_{1,2} = \sqrt{\frac{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2}{n}}$$

Man vergleicht zwei Proteinstrukturen 1 und 2 durch die Berechnung des mittleren quadratischen Abstands der Koordinaten der n sich entsprechenden Atome.

Dann nimmt man noch die Wurzel daraus.

Werte unterhalb von 0.2 nm oder 2 Å kennzeichnen eine hohe strukturelle Ähnlichkeit.

Zum Vergleich: die Länge einer C-C Bindung beträgt 0.15 nm.

Die Distanzen aller Atome weichen also höchstens etwa um eine Bindungslänge voneinander ab.

Test für die Zuverlässigkeit von SwissModell

3DCrunch-Projekt von Expasy zusammen mit SGI.

Idee: Generiere „Homologie-Modelle“ für Proteine mit bekannter 3D-Struktur um zu überprüfen, wie genau die mit Homologie-Modellierung erzeugten Strukturmodelle sind.

Die Vorlagen besaßen 25 – 95 % Sequenzidentität mit dem Zielprotein.

1200 Kontrolle-Modelle wurden erstellt.

Grad der Identität [%]	Modell innerhalb von x Å RMSD zur Vorlage					
	< 1	< 2	< 3	< 4	< 5	> 5
25-29	0	10	30	46	67	33
30-39	0	18	45	66	77	23
40-49	9	44	63	78	91	9
50-59	18	55	79	86	91	9
60-69	38	72	85	91	92	8
70-79	42	71	82	85	88	12
80-89	45	79	86	94	95	5
90-95	59	78	83	86	91	9

www.expasy.org/swissmodel/SWISS-MODEL.html

CAMEO: Continuous Automated Model Evaluation

Wettbewerb: Für die innerhalb von 5 Wochen neu bei der PDB eingereichten Kristallstrukturen von Proteinen werden aufgrund von deren Sequenzen mit automatisierten Homology-Modeling-Webservern Strukturmodelle erzeugt und mit den experimentellen Strukturen verglichen.

CAMEO verwendet für die Bewertung keine RMSD-Abweichungen, da deren Werte stark durch Rotationen zwischen Proteindomänen beeinflusst werden.

Stattdessen werden IDDT-scores (local distance difference test) berechnet:

- Betrachtet werden alle paarweisen Distanzen zwischen Atomen ($< 15 \text{ \AA}$).
- Getestet wird, wie viele Distanzen um weniger als 0.5 \AA , 1 \AA , 2 \AA und 4 \AA abweichen.
- Daraus wird der durchschnittliche Prozentanteil gebildet.

Haas et al. Proteins
86, 387-398 (2017)

CAMEO: Continuous Automated Model Evaluation

TABLE 2 CAMEO IDDT based ranking for the time frame 2016-05-01 to 2016-07-30^a

Server Name	IDDT (all Targets)
Robetta	65.3 ± 16.45
RaptorX	63.8 ± 16.57
IntFOLD3-TS ^c	62.2 ± 17.53
IntFOLD4-TS	62.0 ± 16.32
SWISS-MODEL	56.5 ± 22.49
SPARKS-X	56.3 ± 18.25
Princeton_template	55.6 ± 15.68
IntFOLD2-TS ^c	55.1 ± 17.47
HHpredB ^b	47.6 ± 17.41
M4T ^c	45.1 ± 16.74
Phyre2 ^c	44.5 ± 23.135
NaiveBLAST ^c	43.3 ± 25.59
RBO Aleph ^b	38.6 ± 16.28

TABLE 4 Average response time from submission of the sequence to reception of structure prediction by CAMEO^a

Server name	Avg. response time (hh:mm:ss)
HHpredB ^b	00:21:38
SWISS-MODEL	00:27:35
SPARKS-X	01:29:17
Phyre2 ^c	02:14:08
Princeton_template	03:22:15
NaiveBLAST ^c	04:07:31
M4T ^c	08:20:48
RaptorX	13:53:17
RBO Aleph ^b	16:12:39
IntFOLD2-TS ^c	28:09:28
IntFOLD3-TS ^c	28:23:14
Robetta	29:03:28
IntFOLD4-TS	36:42:55

Robetta macht die besten Modelle, benötigt dafür jedoch aufgrund von aufwändigem Konformationssampling sehr lange.

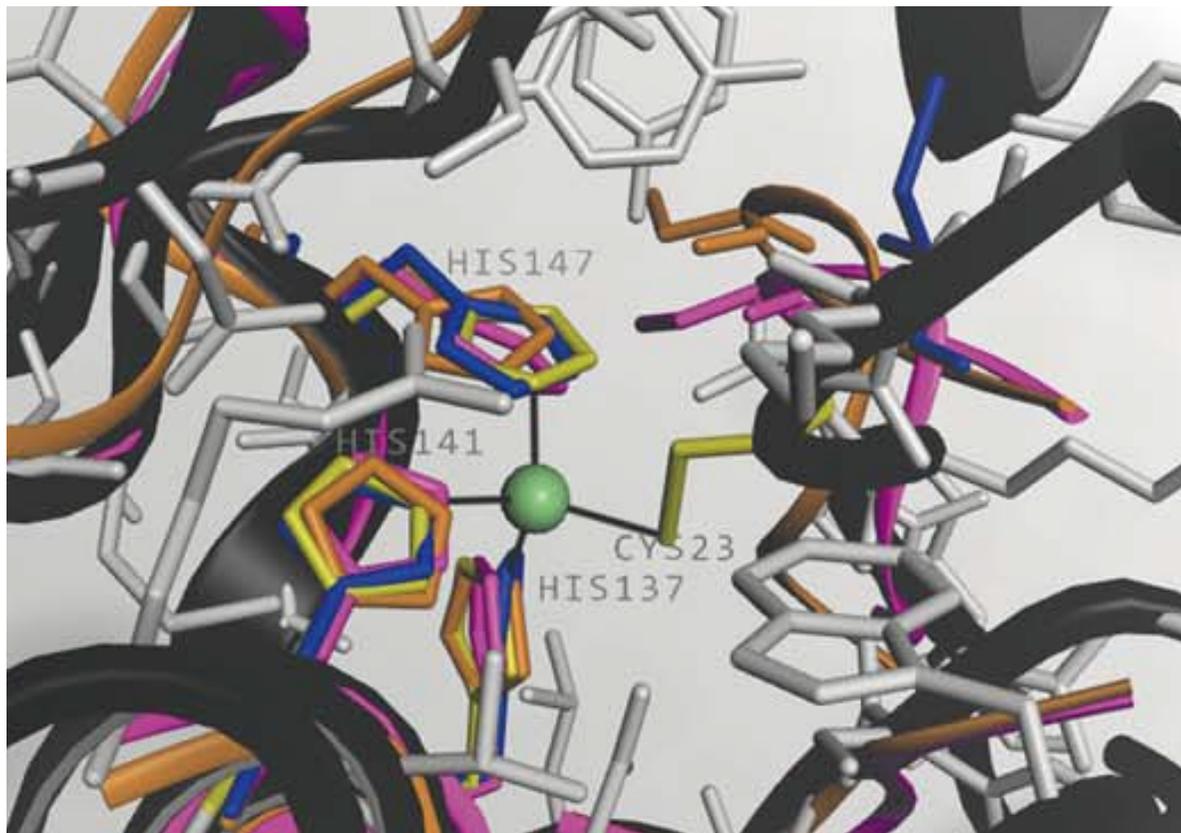
Haas et al. Proteins
86, 387-398 (2017)

CAMEO: Continuous Automated Model Evaluation

Strukturmodelle für Myroilysin, PDB ID 5CZW, schwarz).

In der Mitte liegt ein Zink-Ion (grüne Kugel), das von 4 Residuen (3 His, 1 Cys) koordiniert wird (gelb - Kristallstruktur).

Homologiemodelle mit SWISS-MODEL (orange), IntFOLD4-TS (blau), Sparks-X (magenta).



Alle Tools machen gute Vorhersagen für die drei Histidin-Residuen 137, 141 und 147, die das Zink-Ion koordinieren, weichen jedoch für Cystein 23 voneinander ab.

Haas et al. Proteins
86, 387-398 (2017)

Ligandendocking in Homologiemodelle ??

- Homologiemodelle können zwar recht gut sein, aber nicht immer für Ligandendocking geeignet sein
- Grund: falsche Seitenkettenrotamere in Bindungstasche
- Ansatz1: verwende flexibles Docking, wo auch Teile des Proteins flexibel sind
- Ansatz2: verwende zusätzliches experimentelles Wissen, verlangt manuelles Vorgehen
- Ansatz3: erstelle Homologiemodell in Anwesenheit eines modellierten Liganden, dessen Position z.B. aus Modell-Vorlage stammt

Zusammenfassung – Homologiemodellierung

- Gemeinsamer Kern von Proteinen mit 50% Sequenzidentität besitzt ca. 1 Å RMSD
- Dies gilt sogar für absolut identische Sequenzen.
- Der zuverlässigste Teil eines Proteinmodells ist der Sequenzabschnitt, den es mit der Vorlage gemeinsam hat. Die größten Abweichungen liegen in den konstruierten Schleifen.
- Die Wahl der Modellvorlage ist entscheidend!
Die An- oder Abwesenheit von Ko-faktoren, anderen Untereinheiten oder Substraten kann Proteinkonformation sehr beeinflussen und somit alle Modelle, die von ihnen abgeleitet werden.
- Jeder Fehler im Alignment produziert falsche Modelle!
Solche Alignment-Fehler treten bei Sequenzidentität unter 40% auf.