

# 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?

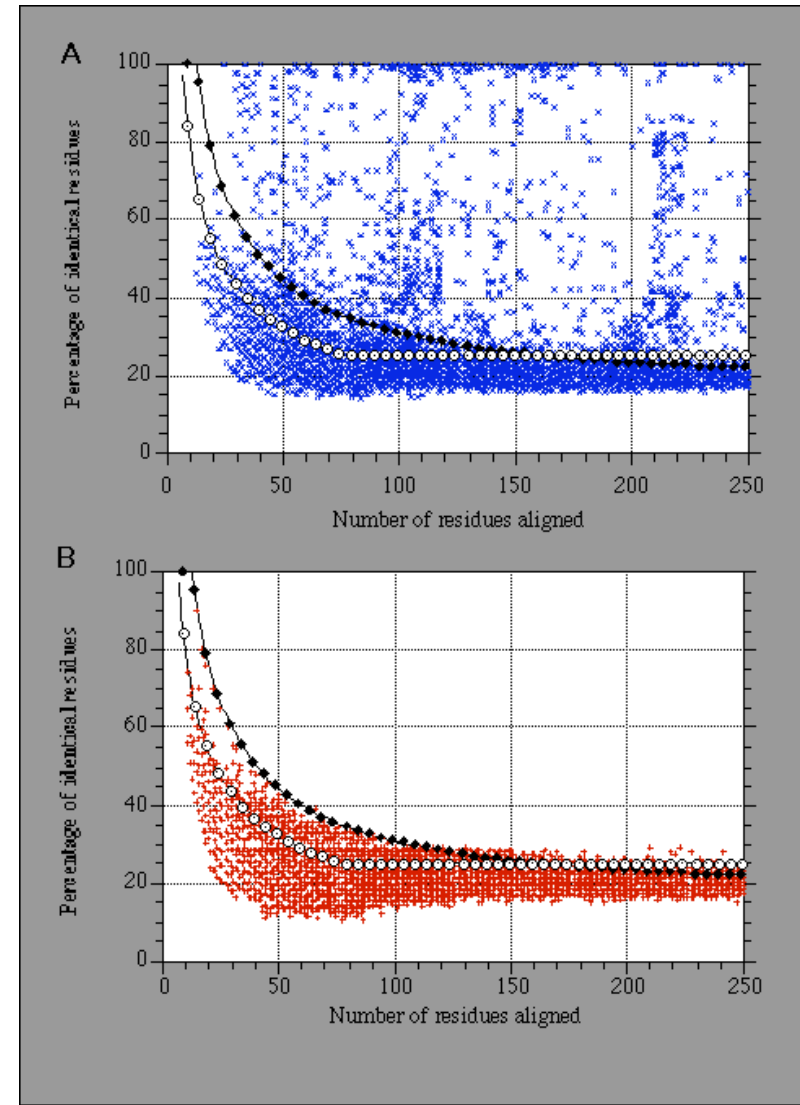
# 1 Twilight Zone

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.

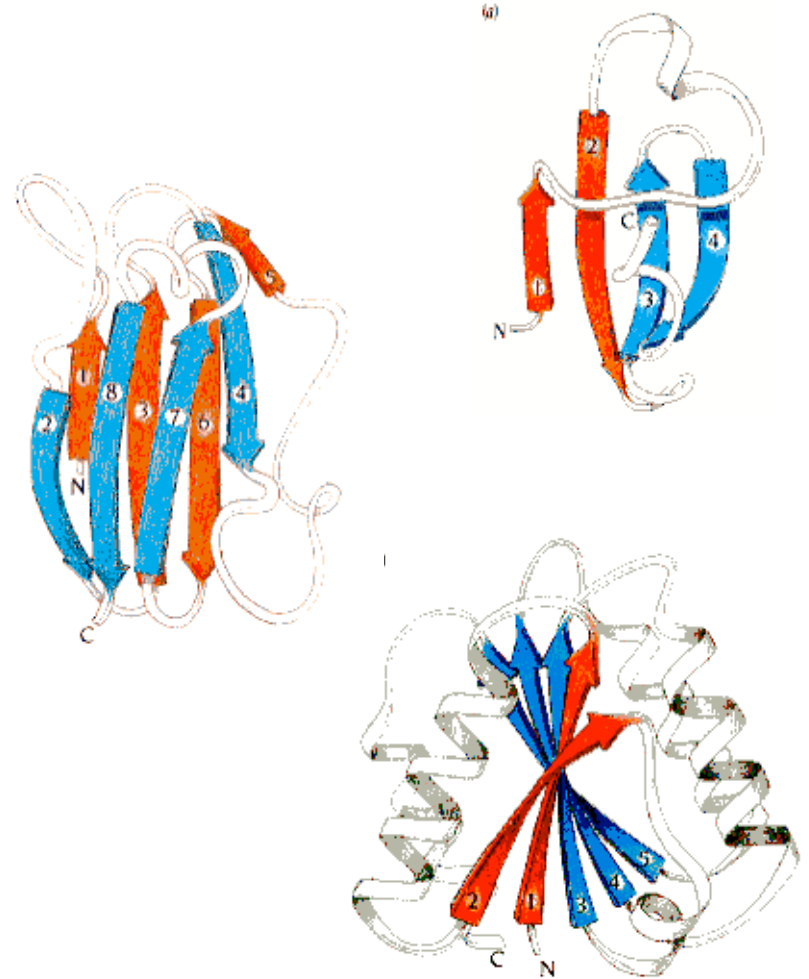
**B**: „false positives“: Strukturen, die keine bzw. wenig Übereinstimmung aufweisen, liegen stets unter der Kurve.



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 |    |    |    |    |   |    |    |    |    |   |    |   |    |    |   |   |    |    |    | Gribskov, PNAS 84, 4355 (1987) |  |
|-----|---------|-----------|---------|----|----|----|----|---|----|----|----|----|---|----|---|----|----|---|---|----|----|----|--------------------------------|--|
|     |         |           | A       | C  | D  | E  | F  | G | H  | I  | K  | L  | M | N  | P | Q  | R  | S | T | V  | W  | Y  | +/-                            |  |
| 1   | E G V L | V         | 3       | -2 | 3  | 4  | 0  | 4 | -1 | 3  | -1 | 4  | 4 | 1  | 1 | 1  | -2 | 1 | 2 | 6  | -6 | -2 | 9                              |  |
| 2   | L L S P | L         | 2       | -2 | -2 | -1 | 3  | 0 | -1 | 3  | -1 | 6  | 5 | -1 | 3 | 0  | -1 | 3 | 1 | 4  | 1  | -1 | 9                              |  |
| 3   | V V V V | V         | 2       | 2  | -2 | -2 | 2  | 2 | -3 | 11 | -2 | 8  | 6 | -2 | 1 | -2 | -2 | 0 | 2 | 15 | -9 | -1 | 9                              |  |
| 4   | K E A T | A         | 6       | -2 | 5  | 6  | -5 | 4 | 1  | 0  | 5  | -2 | 0 | 3  | 3 | 3  | 1  | 3 | 6 | 0  | -6 | -4 | 9                              |  |
| 5   | A P L P | P         | 6       | -1 |    |    |    |   |    |    |    |    |   |    |   |    |    |   |   |    |    |    |                                |  |





# 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.





# 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















# 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.

































