

Nicht-regelmäßige Sekundärstruktur Sekundärstrukturvorhersage Struktur motive

Wintersemester 2011/12

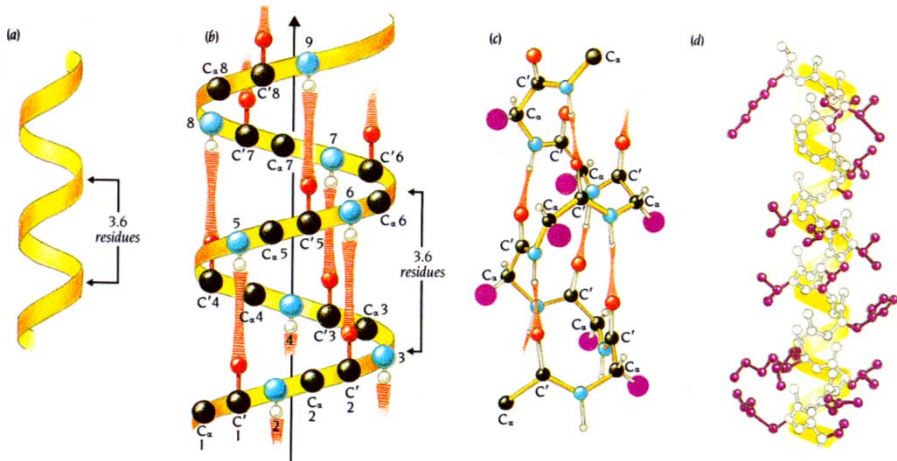
Peter Güntert

Regelmäßige Sekundärstrukturen

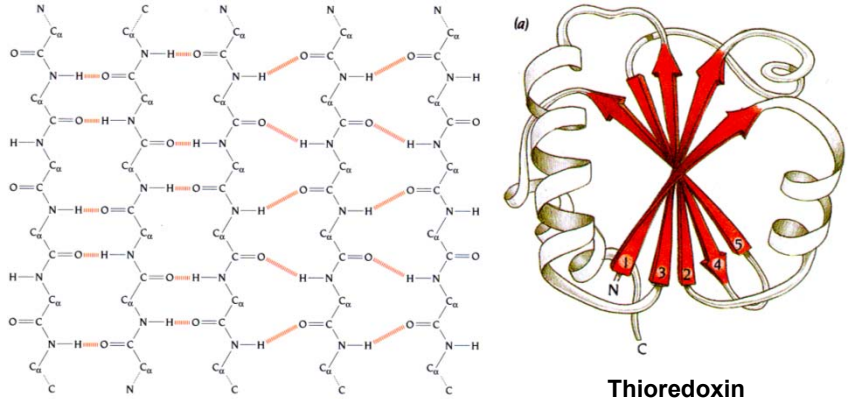
- Eine regelmäßige Sekundärstruktur ist dadurch charakterisiert, dass mehrere aufeinanderfolgende AS die gleiche Rückgratskonformation aufweisen, d.h. dass die Diederwinkel ϕ und ψ jeweils den gleichen Wert haben.

Sekundärstruktur	ϕ	ψ	AS/ Windung	Wasserstoffbrücken
Rechtshändige α -Helix	-57°	-47°	3.6	$i \rightarrow i + 4$
Rechtshändige 3_{10} -Helix	-49°	-26°	3.0	$i \rightarrow i + 3$
Antiparalleles β -Blatt	-139°	135°	2.3	$i \rightarrow j, i + 1 \rightarrow j - 1, \dots$
Paralleles β -Blatt	-119°	113°	2.3	$i \rightarrow j, i + 1 \rightarrow j + 1, \dots$

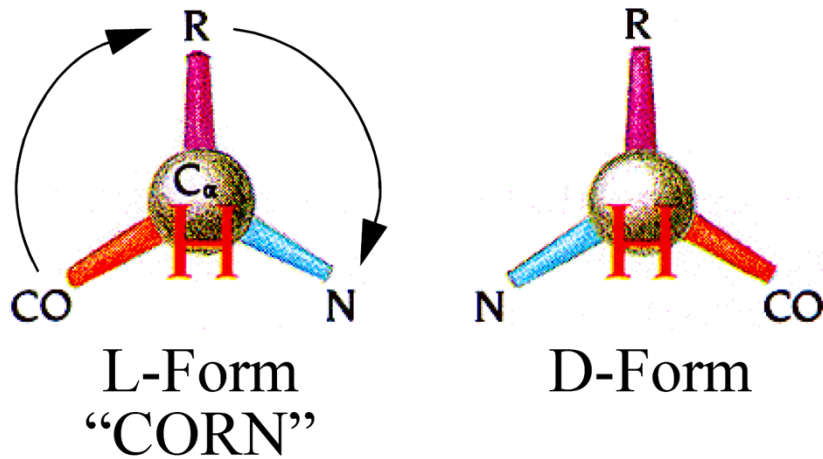
α -Helix



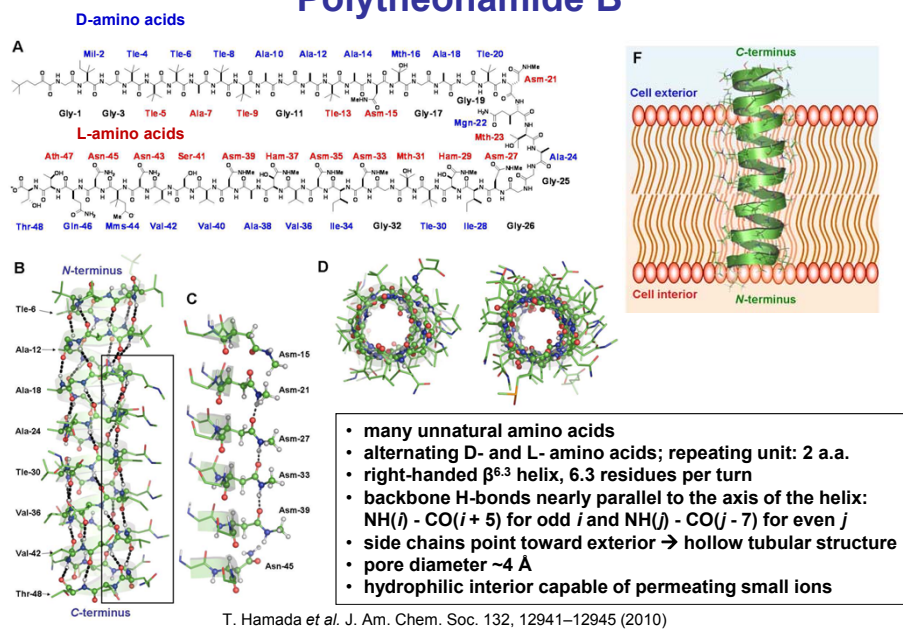
Gemischt antiparalleles/paralleles β -Blatt



L- und D-Aminosäuren



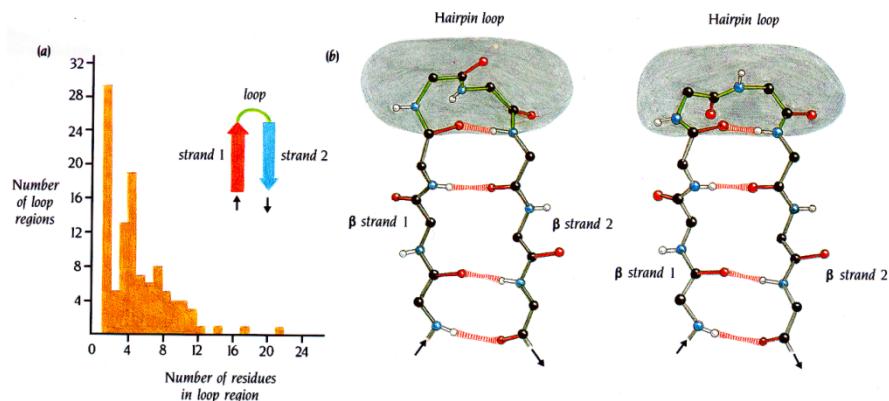
Unusual regular secondary structure: Polytheonamide B



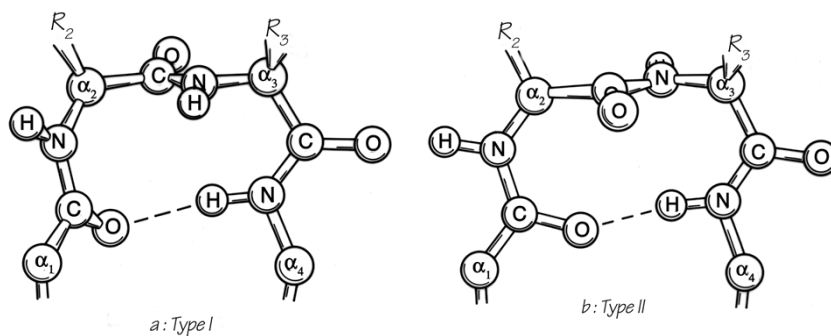
Nicht regelmäßige Sekundärstrukturen

- Regelmässige Sekundärstruktur:
 $\phi_i = \phi_{i+1} = \phi_{i+2} = \dots$ und $\psi_i = \psi_{i+1} = \psi_{i+2} = \dots$
- Nicht regelmässige Sekundärstruktur:
 Häufig auftretende lokale Strukturen mit bestimmten Werten der Torsionswinkel
 (ϕ_i, ψ_i) , (ϕ_{i+1}, ψ_{i+1}) , (ϕ_{i+2}, ψ_{i+2}) , (ϕ_{i+3}, ψ_{i+3})
- Insbesondere Turns (tight-turns)

Hairpin Loops



Type I and Type II Turns



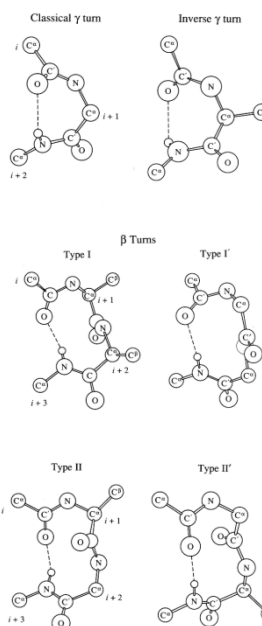
Tight turns

Table 6.2 Structural Features of γ and β Turns

Bend type	Dihedral Angles of Central Residues (deg) ^a			
	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
γ				
classical	70 to 85	-60 to -70		
inverse	-70 to -85	60 to 70		
β				
I	-60	-30	-90	0
I'	60	30	90	0
II	-60	120	80	0
II'	60	-120	-80	0
III	-60	-30	-60	-30
III'	60	30	60	30
IV	Any bend with two or more angles differing by $>40^\circ$ from those given here			
V	-80	80	80	-80
V'	80	-80	-80	80
VIa ^b	-60	120	-90	0
VIb ^b	-120	120	-60	0
VII	Kink in chain created by $\psi_2 \approx 180^\circ$, $ \phi_3 < 60^\circ$; or $ \psi_2 < 60^\circ$, $\phi_3 \approx 180^\circ$			
VIII	-60	-30	-120	120

^a The central residue of a γ turn is numbered $i+1$; the two central residues of a β turn are $i+1$ and $i+2$.

^b The peptide bond between residues $i+1$ and $i+2$ is cis, and residue $i+2$ is Pro.



Aminosäuren in Sekundärstrukturen

Allgemein:

- Pro: keine H-Brücke, stark eingeschränkter ϕ Winkel
- Gly: konformativ freier im ungefalteten Zustand (Entropie!)
- C $^{\beta}$ -verzweigte AS: konformative Einschränkung
- H-Brückenbildung mit Rückgrat im ungefalteten Zustand ungünstig
- α -Helix:
 - häufig: A, E, L >> H, M, Q, W, V, F
 - selten: G, P, Y, N
- β -Faltblatt:
 - häufig : M, V, I >> C, Y, F, Q, L, T, W
 - selten : E << P, N, H, S, K
- Turns:
 - häufig : N, G, P, S
 - selten : I, V, L, M, F

Relative Häufigkeiten der AS in Sekundärstrukturen

	Name	P (a)	P (b)	P (turn)
Relative Häufigkeiten der AS in den drei Sekundärstrukturklassen α-Helix, β-Blatt und Turn. Erlauben zusammen mit empirischen Regeln (z.B. Chou-Fasman) Vorhersage der Sekundärstruktur Zuverlässigkeit 50-60%	Alanine	142	83	66
	Arginine	98	93	95
	Aspartic Acid	101	54	146
	Asparagine	67	89	156
	Cysteine	70	119	119
	Glutamic Acid	151	037	74
	Glutamine	111	110	98
	Glycine	57	75	156
	Histidine	100	87	95
	Isoleucine	108	160	47
	Leucine	121	130	59
	Lysine	114	74	101
	Methionine	145	105	60
	Phenylalanine	113	138	60
	Proline	57	55	152
	Serine	77	75	143
Threonine	83	119	96	
Tryptophan	108	137	96	
Tyrosine	69	147	114	
Valine	106	170	50	

Secondary Structure Prediction

- Given a protein sequence $a_1a_2\dots a_N$, secondary structure prediction aims at defining the state of each amino acid as being either H (helix), E (extended = strand), or O (other) (Some methods have 4 states: H, E, T for turns, and O for other).
- The quality of secondary structure prediction is measured with a “3-state accuracy” score, or Q_3 . Q_3 is the percent of residues that match “reality” (experimentally determined structure).

Q_3 score for secondary structure prediction

ALHEASGPSVILFGSDVTVPPASNAEQAK Amino acid sequence
hhhhhoooo**eeee**oooo**eeee**oooo**hhhhh** Actual secondary structure

o**hhh**oooo**eeee**oooo**eeee**ooo**hhhhh** $Q_3 = 22/29 = 76\%$
 (useful prediction)

hhhhhoooo**hhhh**ooo**hhh**oooo**hhhhh** $Q_3 = 22/29 = 76\%$
 (terrible prediction)

- Q_3 for random prediction is 33%.
- Secondary structure assignment in real proteins has ~10% uncertainty. Therefore, a “perfect” prediction would have $Q_3 = 90\%$.

Chou-Fasman Parameter

AS	P_α	Klasse	AS	P_β	Klasse	AS	P_α	Klasse	AS	P_β	Klasse
Glu	1.53	H_α	Met	1.67	H_β	Ile	1.00	I_α	Ala	0.93	I_β
Ala	1.45		Val	1.65		Asp	0.98	i_α	Arg	0.90	i_β
Leu	1.34		Ile	1.60		Thr	0.82		Asp	0.80	
His	1.24	h_α	Cys	1.30	h_β	Ser	0.79	b_α	Lys	0.74	b_β
Met	1.20		Tyr	1.29		Arg	0.79		Ser	0.72	
Gln	1.17		Phe	1.28		Cys	0.77		His	0.71	
Trp	1.14		Gln	1.23		Asn	0.73	Asn	0.65		
Val	1.14		Leu	1.22		Tyr	0.61	Pro	0.62		
Phe	1.12		Thr	1.20		Pro	0.59	Glu	0.26		
Lys	1.07		I_α	Trp		1.19	Gly	0.53	B_α	B_β	

Chou-Fasman rules for secondary structure prediction (simplified)

Predicting helices:

1. Find nucleation site: 4 out of 6 contiguous residues with $P(\alpha) > 1$.
2. Extension: extend helix in both directions until a set of 4 contiguous residues has an average $P(\alpha) < 1$ (breaker).
3. If average $P(\alpha)$ over whole region is > 1 , it is predicted to be helical.

Predicting strands:

1. Find nucleation site: 3 out of 5 contiguous residues with $P(\beta) > 1$.
2. Extension: extend strand in both directions until a set of 4 contiguous residues has an average $P(\beta) < 1$ (breaker).
3. If average $P(\beta)$ over whole region is > 1 , it is predicted to be a strand.

http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=misc1

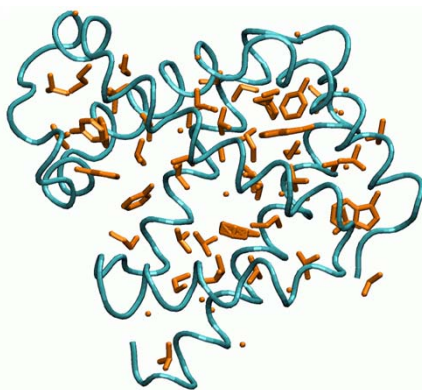
P.Y. Chou and G.D. Fasman (1978). Empirical predictions of protein conformation. *Ann. Rev. Biochem.* 47: 251-276

Chou-Fasman Regeln für Sekundärstrukturvorhersage

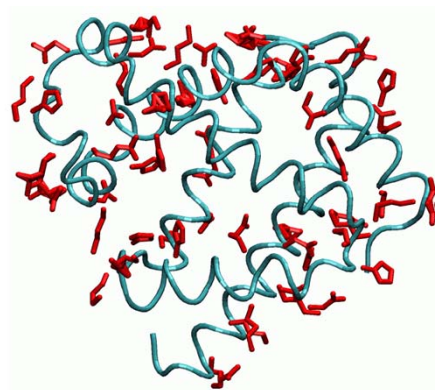
1. A cluster of four helical residues (H_α or h_α) out of six residues along the protein sequence will nucleate a helix, with weak helical residues (l_α) counting as 0.5 h_α . The helical segment is extended in both directions until α -tetrapeptide breakers with $\{P_\alpha\} < 1.00$ are reached. Pro cannot occur in the inner helix or at the C-terminal helical end. Pro, Asp, Glu and His, Lys, Arg are incorporated, respectively, at the N- and C-terminal ends. Any segment with $\{P_\alpha\} \geq 1.03$ as well as $\{P_\alpha\} > \{P_\beta\}$ is predicted as helical.
2. A cluster of three β formers (H_β or h_β) out of five residues along the protein sequence will nucleate a β -sheet. The β -sheet is extended in both directions until β -tetrapeptide breakers with $\{P_\beta\} < 1.00$ are reached. Any segment with $\{P_\beta\} \geq 1.05$ as well as $\{P_\beta\} > \{P_\alpha\}$ is predicted as β -sheet.
3. When regions in proteins contain both α - and β -forming residues, the overlapping region is helical if $\{P_\alpha\} > \{P_\beta\}$, or β -sheet if $\{P_\beta\} > \{P_\alpha\}$.
4. The probability of bend occurrence at residue i is calculated from $p_i = f_i \times f_{i+1} \times f_{i+2} \times f_{i+3}$ with the aid of a Table of the conformational parameters for β -turn formation. The average probability of β -turn occurrence is $\{p_i\} = 0.55 \times 10^{-4}$. Tetrapeptides with $p_i > 0.75 \times 10^{-4}$ as well as $\{P_\beta\} > 1.00$ and $\{P_\alpha\} < \{P_\beta\} > \{P_\beta\}$ are selected as probable bends.

P.Y. Chou and G.D. Fasman (1978). Empirical predictions of protein conformation. *Ann. Rev. Biochem.* 47: 251-276

Verteilung der Aminosäuretypen



hydrophobe AS

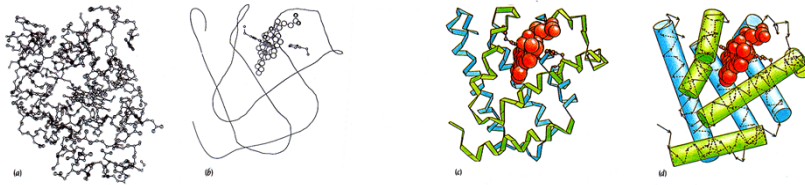


geladene AS

Darstellung von Proteinstrukturen

Methoden:

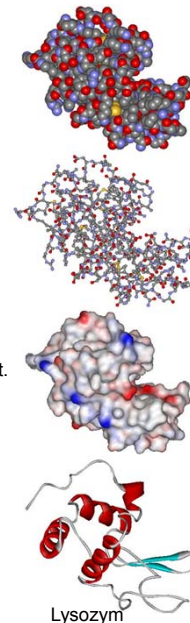
- Dreidimensionale Modelle aus Metall, Kunststoff usw.:** Guter 3D Eindruck, aber sehr aufwendig zu bauen. Unflexibel.
- Interaktive Computergrafik mit Stereoeffekt:** Guter 3D Eindruck, aufwendige Technik.
- 2D Computergrafik:** 3D Eindruck durch Drehung des Moleküls in Echtzeit, Tiefenstaffelung, Schattierung, Perspektive usw.
- Zweidimensionale Darstellungen auf Papier:** Statisch. Stereoeffekt möglich mit Hilfe zweier nebeneinander stehender Bilder, von denen eines vom rechten, das andere vom linken Auge betrachtet wird.



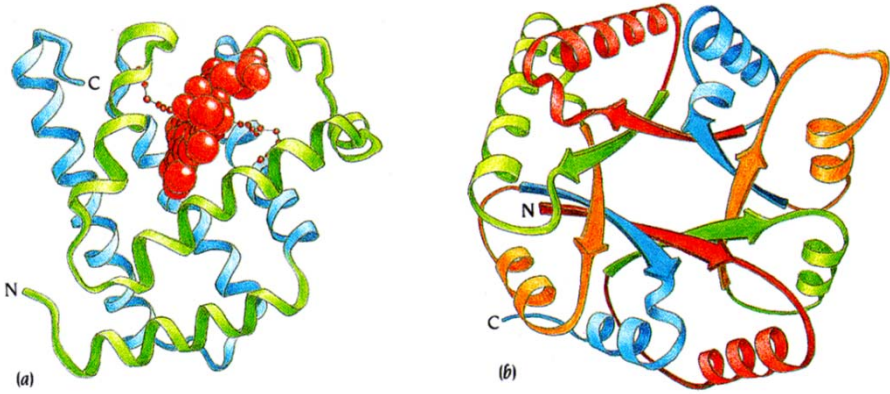
Darstellung von Proteinstrukturen

Was wird dargestellt?

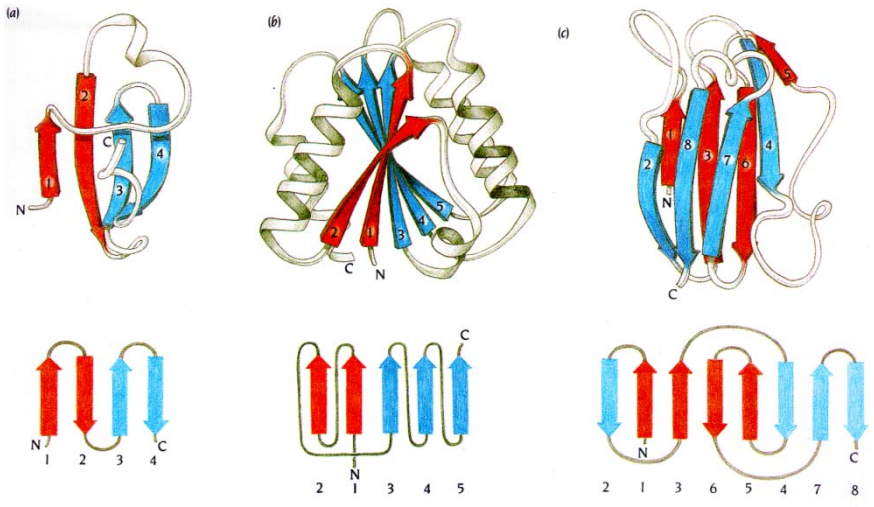
- CPK Modell:** Atome als Kugeln mit Radius, der dem van der Waals Radius entspricht. Guter Eindruck für Rauffüllung und Oberflächen, aber kein Einblick ins Innere des Proteins; Verlauf der Polypeptidkette ist schwierig bis unmöglich zu verfolgen.
- Ball-and-Stick Modell:** Atome als kleine Kugeln, chemische Bindungen als dünne Zylinder. Detailgenauigkeit, wenn ein kleiner Teil des Proteins dargestellt wird; für ganze Proteine zu komplexe Bilder.
- Linienmodell:** Chemische Bindungen als Linien. Einfach, "durchsichtig" (Proteininneres sichtbar), bei Darstellung aller Atome unübersichtlich, sinnvoll z.B. bei Darstellung nur des Rückgrats, geeignet für Überlagerungen bei Strukturvergleichen.
- Oberflächendarstellung:** Wasserzugängliche Oberfläche, die erzeugt wird, indem eine Kugel mit der Größe eines Wassermoleküls auf den Oberflächenatomen des Proteins abgerollt wird. Oft entsprechend des elektrostatischen Potenzials eingefärbt. Erlaubt Identifikation von Bindungstaschen und positiv/negativ geladener „Patches“.
- Schematische Diagramme:** z.B. Helices als Spiralen oder Zylinder, β -Stränge als flache Pfeile, Loops als "Schläuche". Eingeführt von Jane Richardson 1981; ursprünglich von Hand gezeichnet, heute mit dem Computer herstellbar. Guter Überblick über die Sekundärstruktur und die Faltung auch ohne Stereoeffekt.
- Topologiediagramme:** Zweidimensionale Darstellung der β -Blatt Topologie. Einfache, schematische Wiedergabe des Aufbaus von β -Blättern: Anzahl β -Stränge, deren relative Richtung, parallel oder antiparallel, und deren Verknüpfung untereinander. Keine Information über die 3D Struktur.

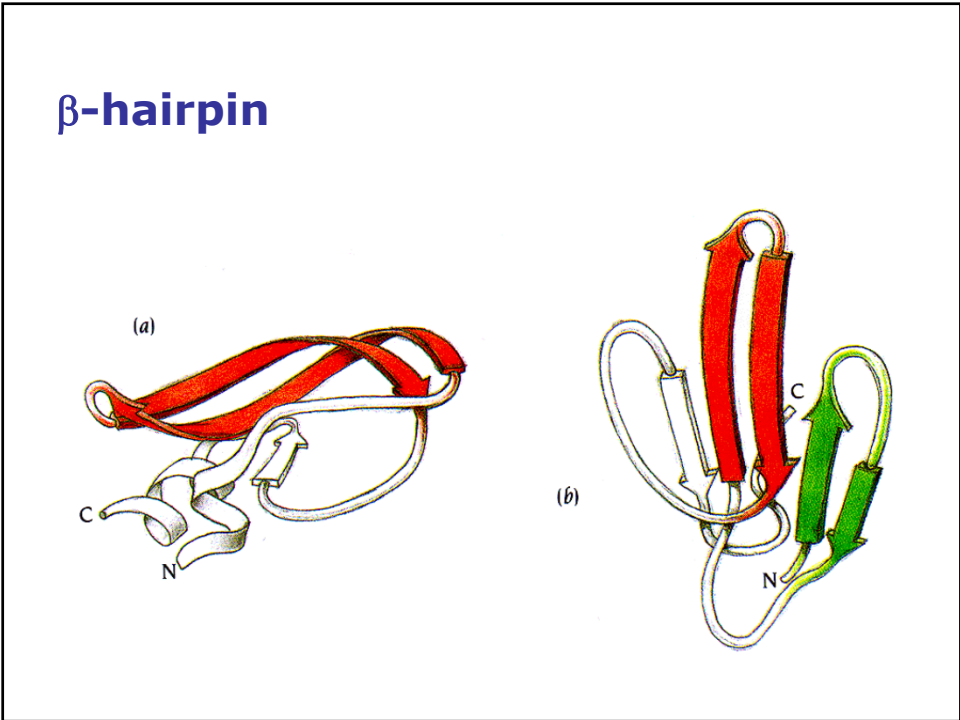
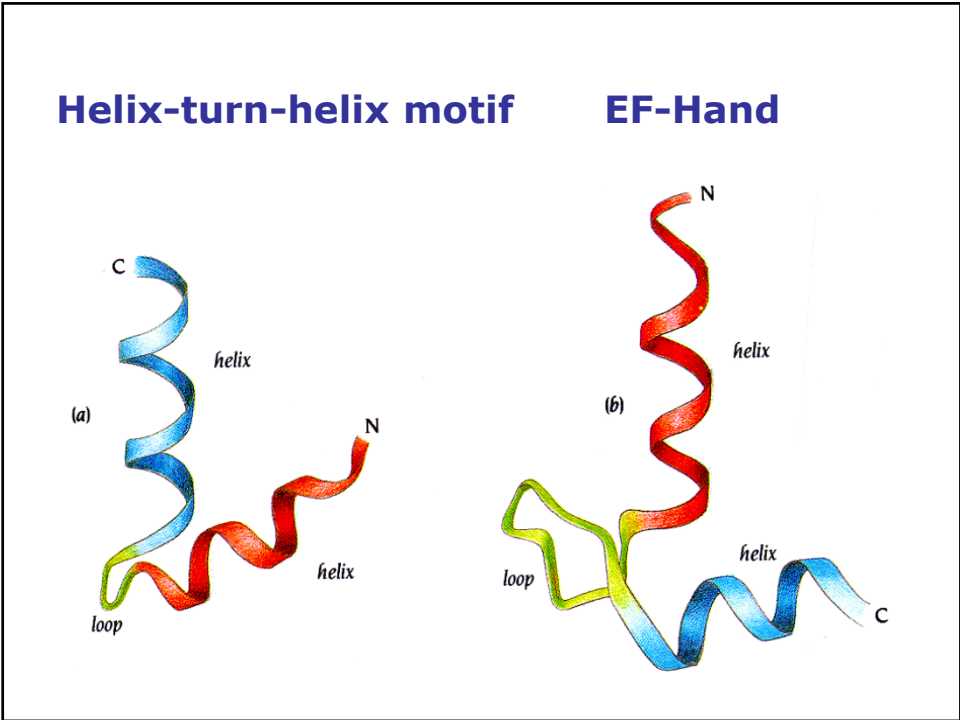


Schematische Diagramme

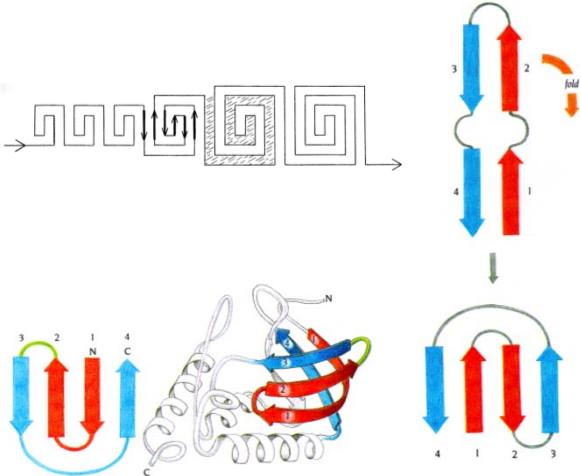


Topologiediagramme





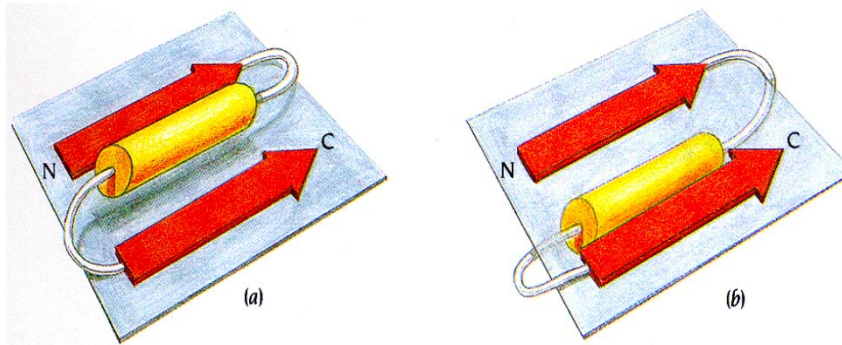
Greek key motif



β - α - β -Motiv



Rechts- und linkshändiges β - α - β Motiv



Literatur

- C. Branden & J. Tooze, *Introduction to Protein Structure*, Garland, 21999.
- M. Williamson, *How Proteins Work*, Garland, 2012
- T. E. Creighton, *Proteins, Structures and Molecular Properties*, Freeman, 21993