



Die Sprache der Proteine – Evolution konservierter Proteinstrukturen

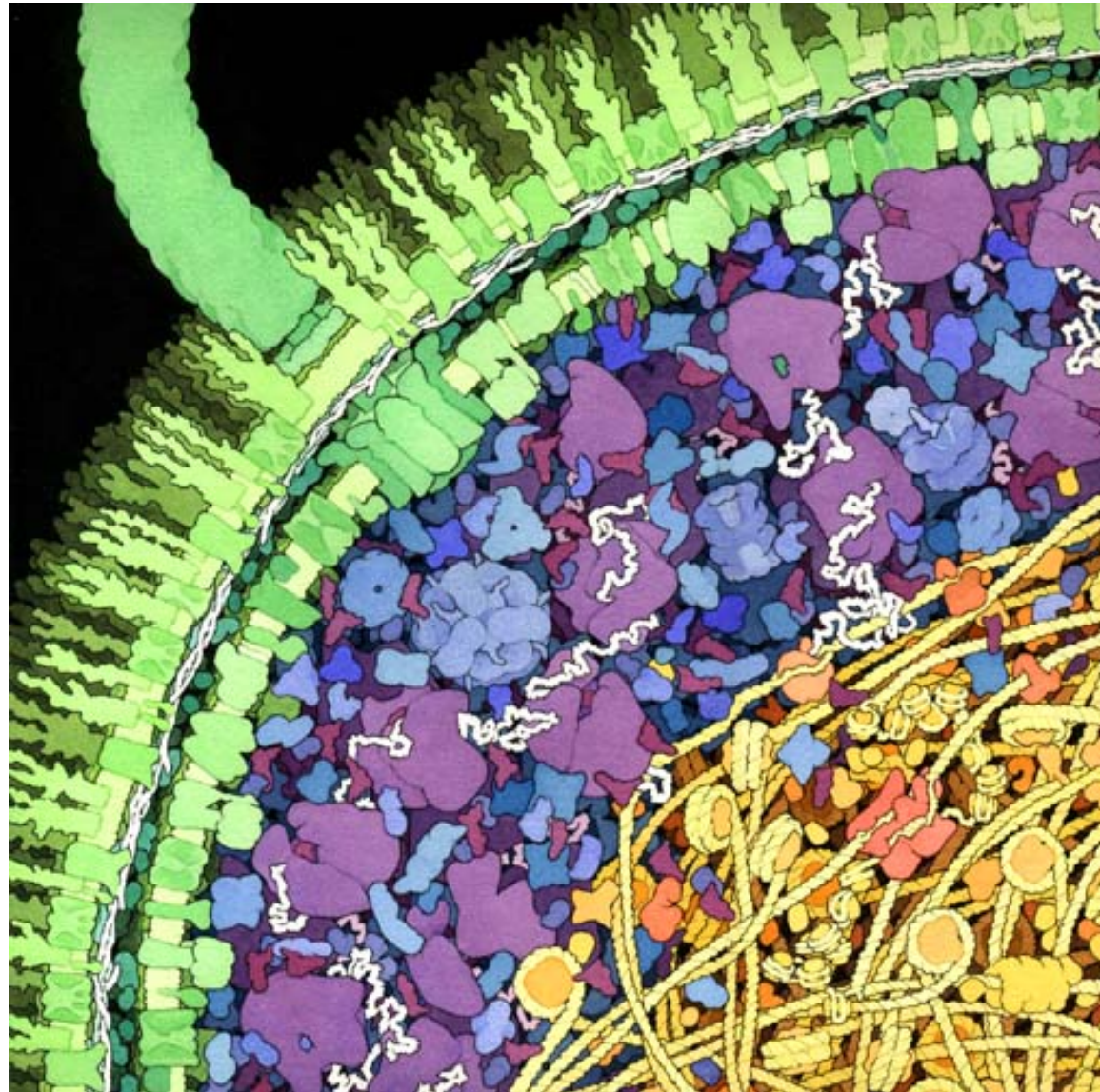
Andrei Lupas

Max Planck Institut für Entwicklungsbiologie
Tübingen

Leben beruht auf der chemischen Aktivität von Proteinen



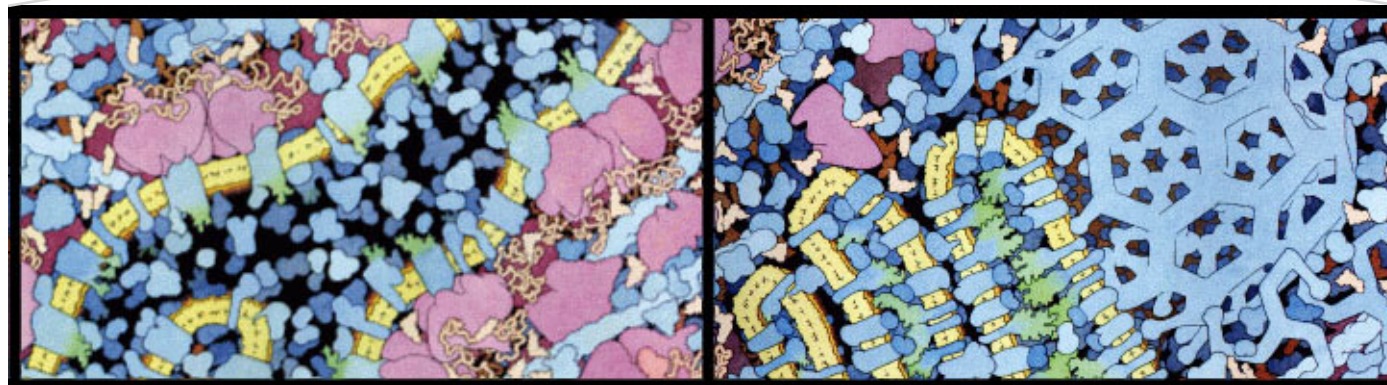
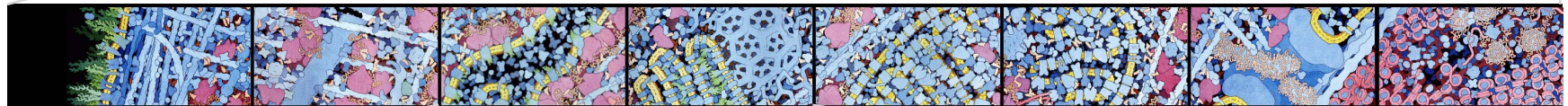
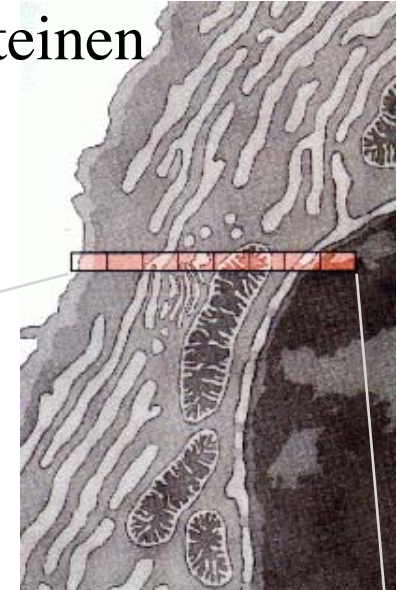
Das Innere eines
Bakteriums



David
Goodsell

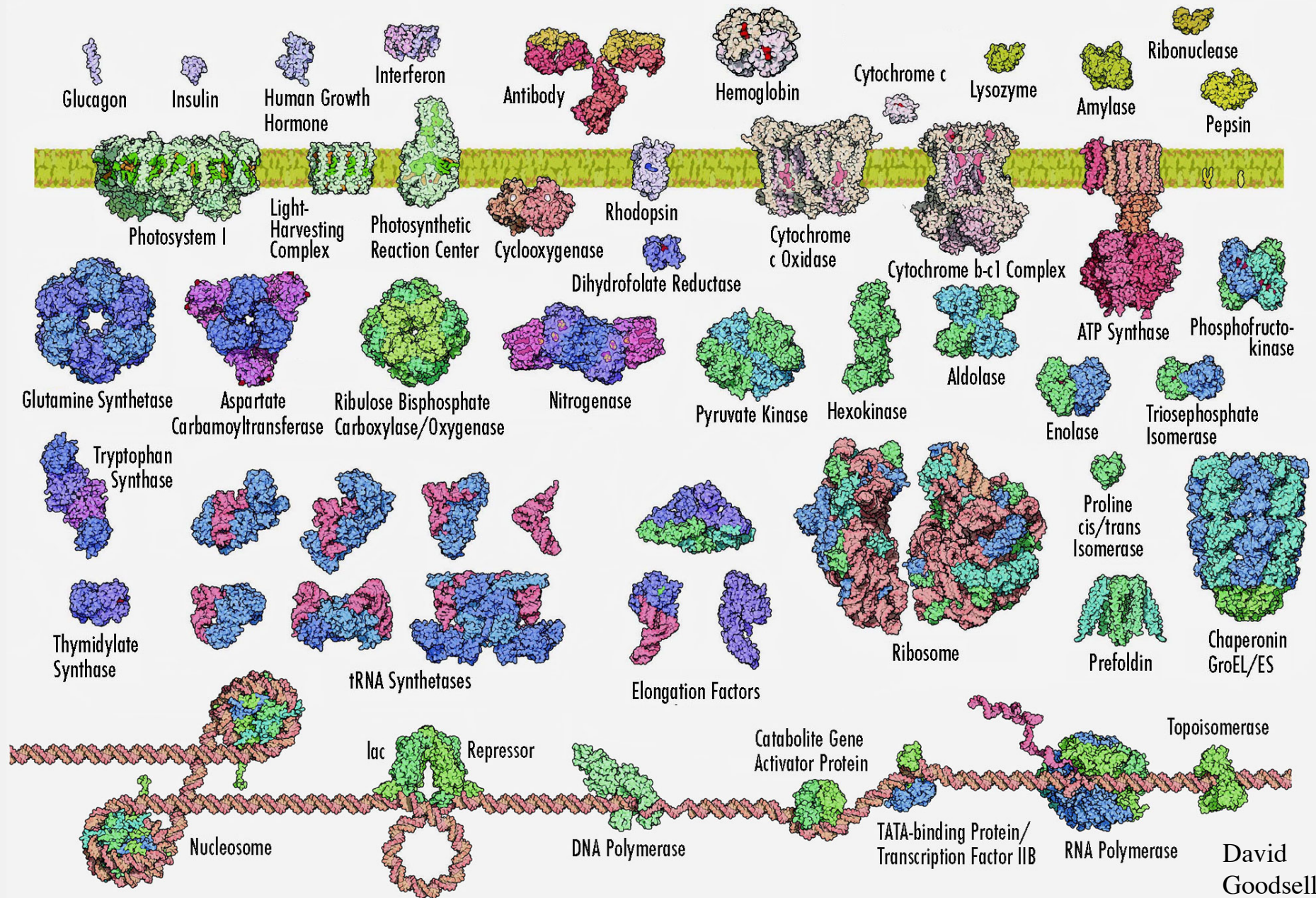
Leben beruht auf der chemischen Aktivität von Proteinen

Das Innere einer
menschlichen Zelle



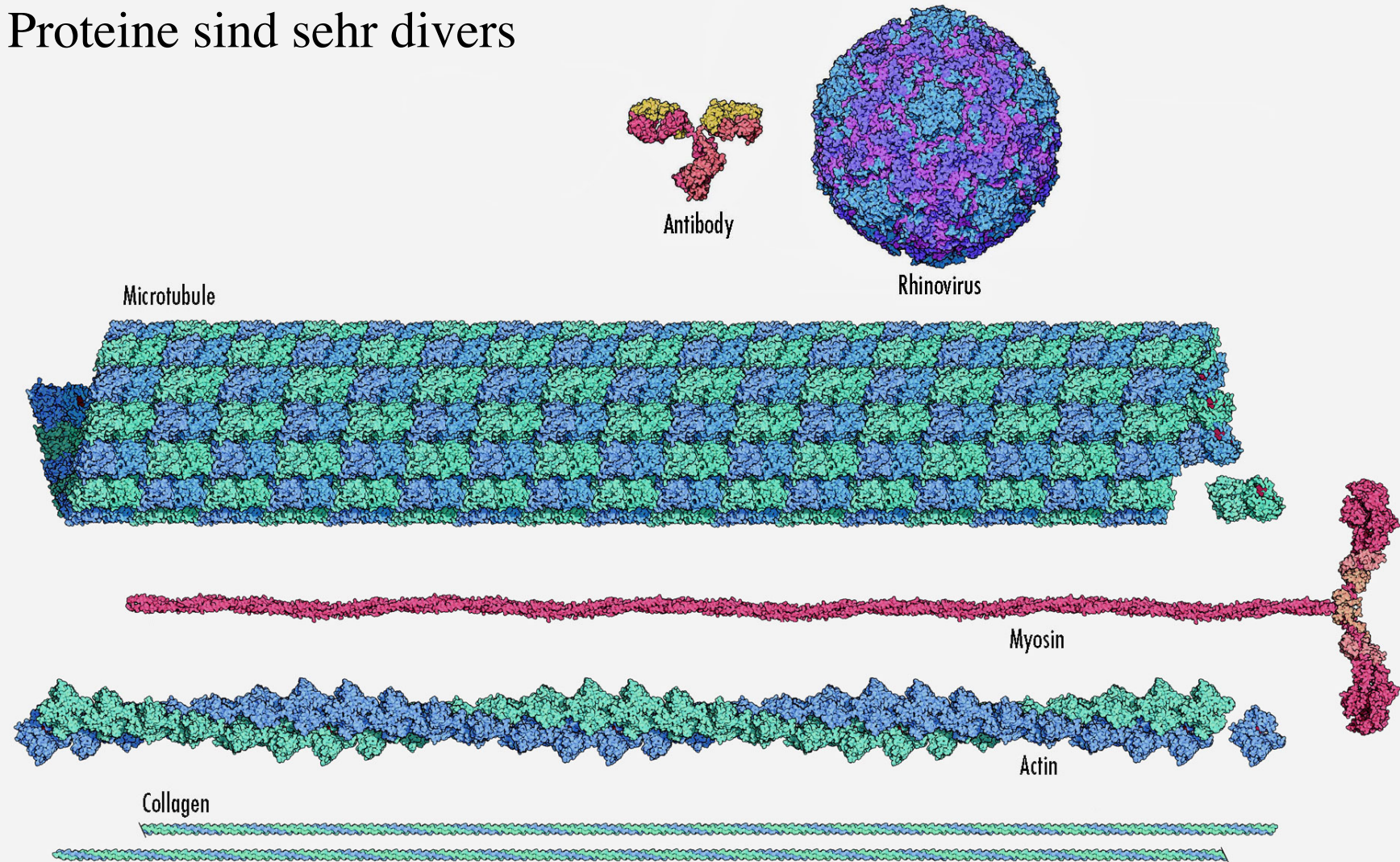
David
Goodsell

Proteine sind sehr divers



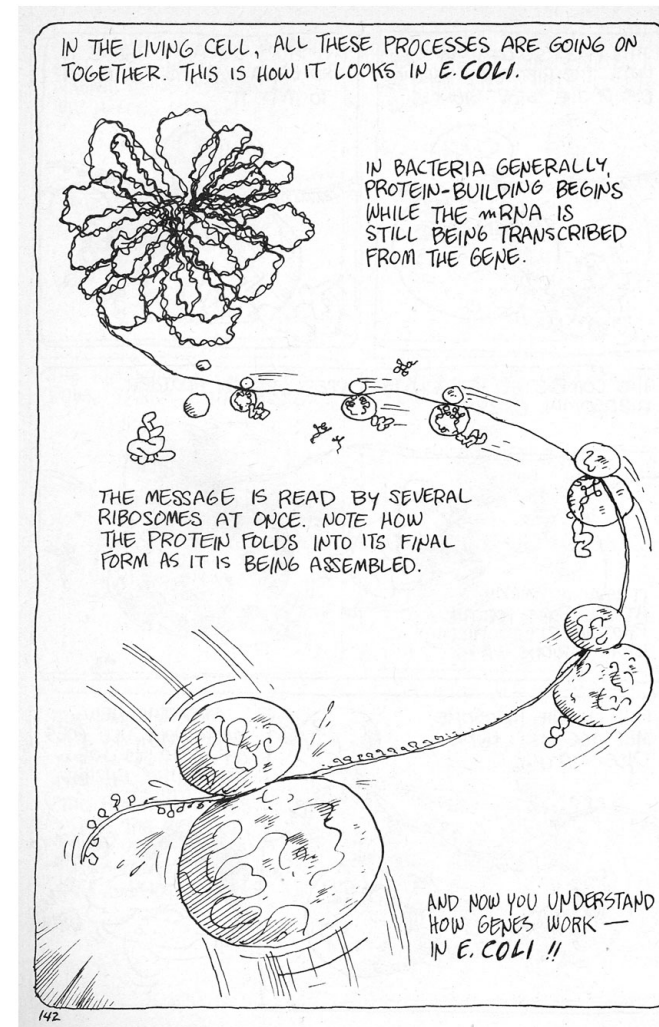
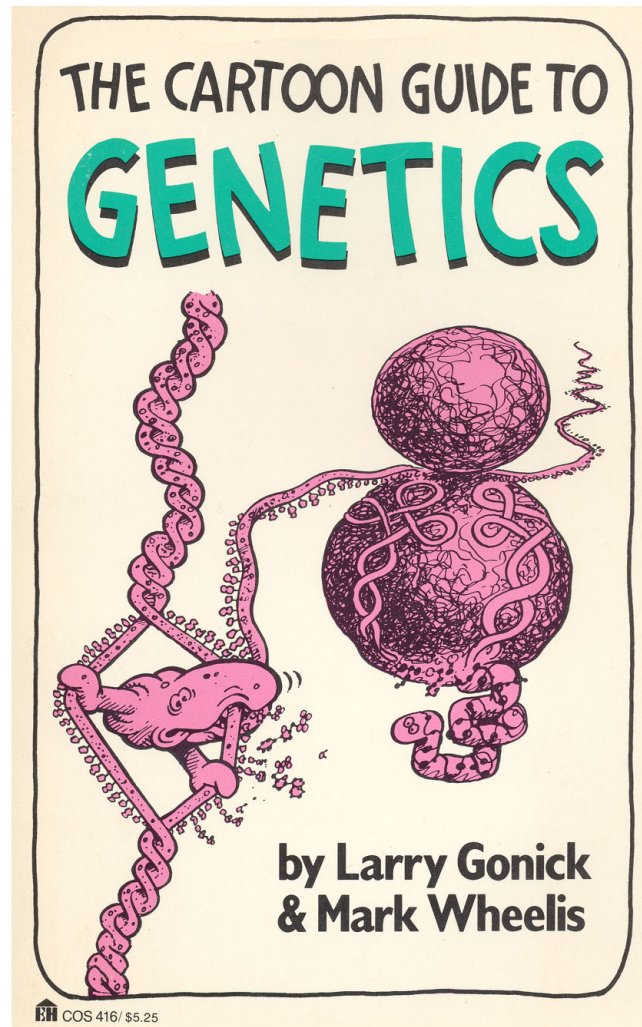
David
Goodsell

Proteine sind sehr divers

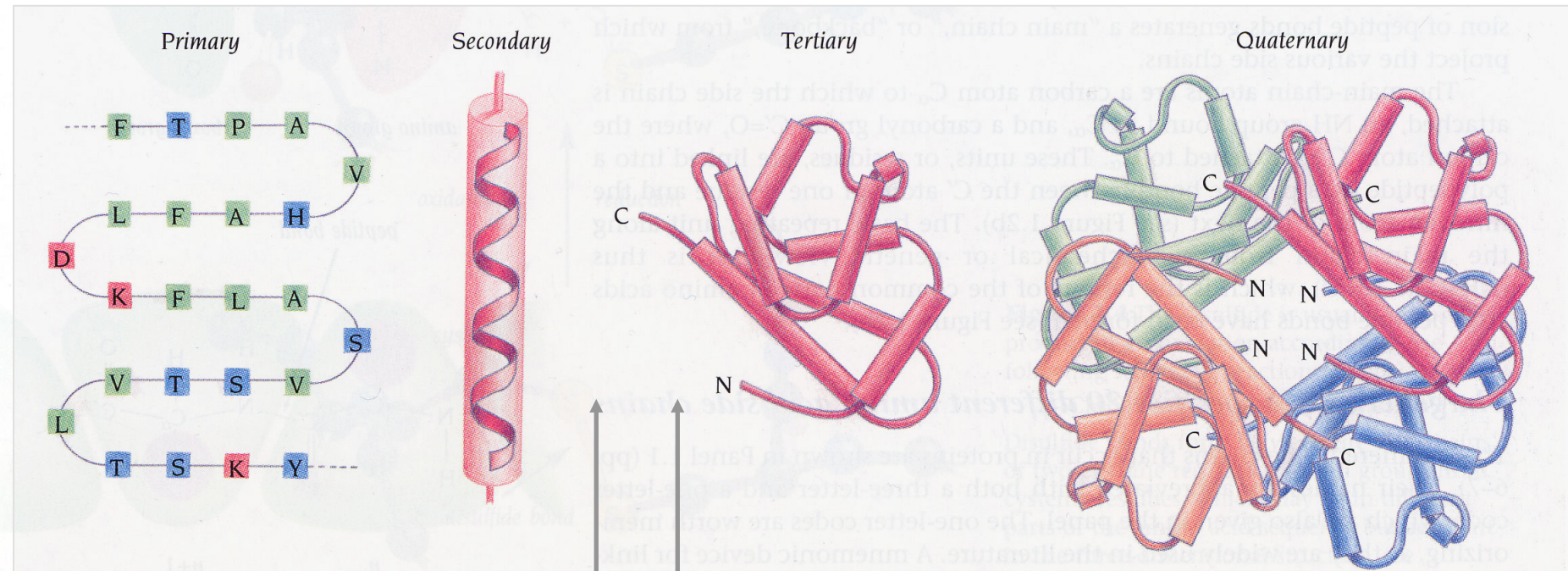


Die meisten Proteine müssen eine definierte 3-D Struktur erreichen um aktiv zu sein

Proteine werden als lineare Ketten von Aminosäuren synthetisiert. Um ihre 3-D Struktur zu erreichen, müssen sie falten



Die Strukturhierarchie der Proteine



Branden & Tooze, Introduction to Protein Structure

Supersecondary

Domain

Primärstruktur: Eine Proteinsequenz (FASTA Format)



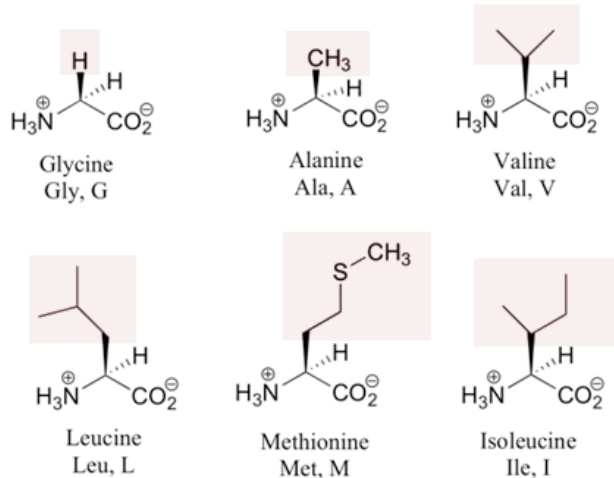
```
>gi|116291|sp|P06143|CHEY_ECOLI CHEMOTAXIS PROTEIN CHEY  
MADKELKFLVVDDFSTMRRIVRNLLKELGFNNVEEAEDGVDALNKLQAGGYGFVISDWNM  
PNMDGLELLKTIRADGAMSALPVLMTAEAKKENIIAAAQAGASGYVVKPFTAATLEEKL  
NKIFEKLG
```

In der Biologie unterscheidet man zwischen Proteinen die von einem gemeinsamen Vorfahren abstammen (Homologen) und Proteinen deren Ähnlichkeit auf einer konvergenten Entwicklung beruht (Analoge).

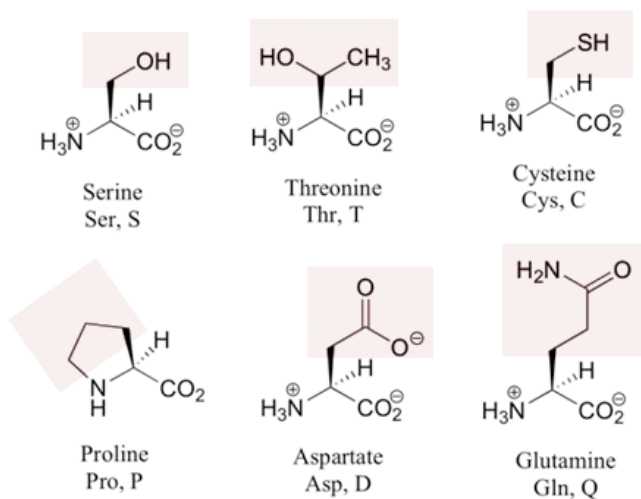
Der effektivste Weg nach Verwandten (Homologen) eines Proteins zu suchen ist über Sequenzähnlichkeit, also aus der Primärstruktur, z.B. mittels BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)

Primärstruktur: Die Seitenketten der Aminosäuren

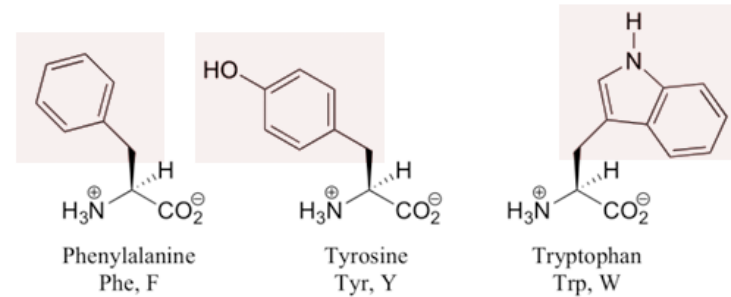
Nonpolar, aliphatic side groups



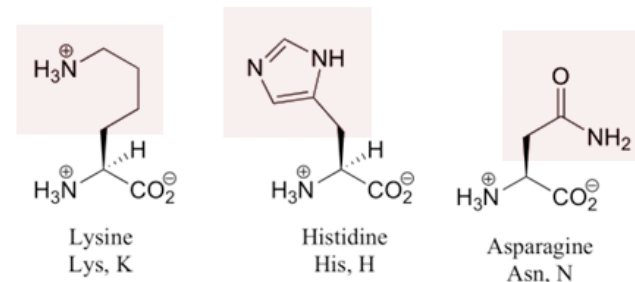
Polar, uncharged side groups



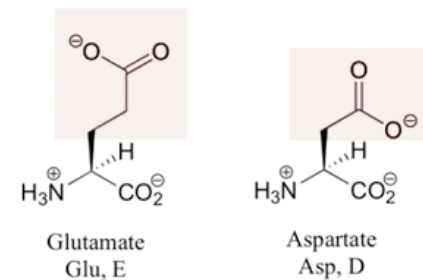
Aromatic side groups



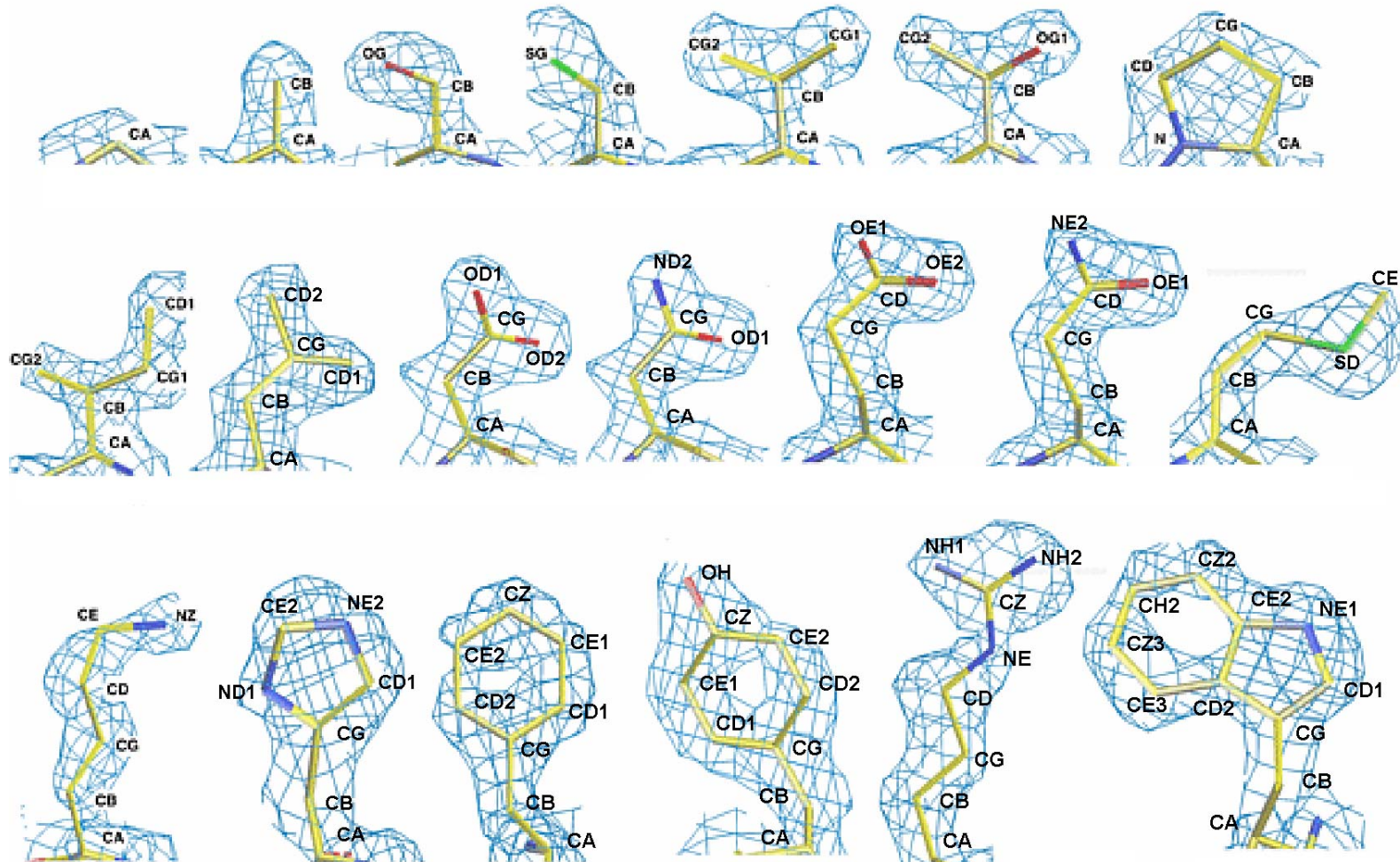
Positively charged side groups



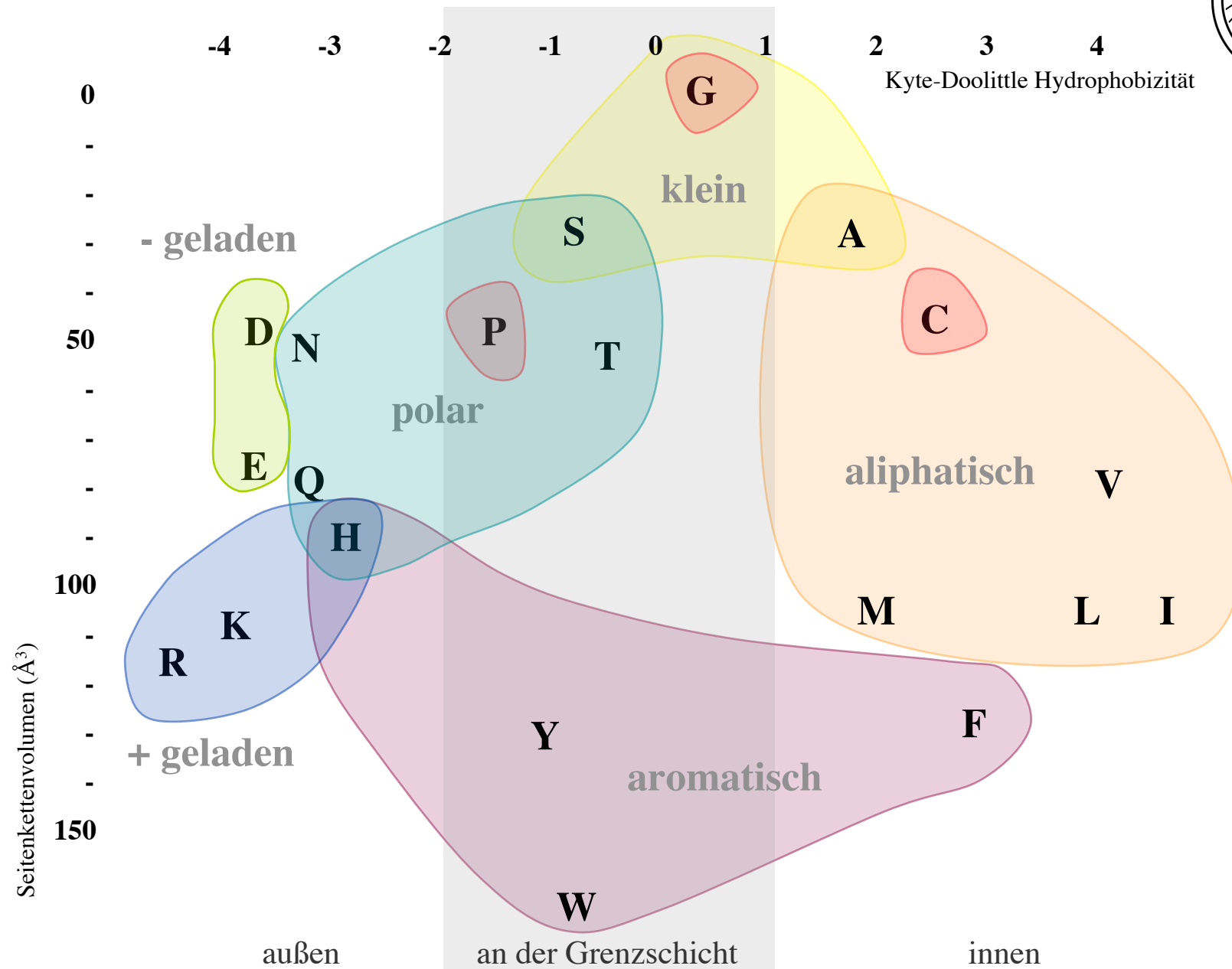
Negatively charged side groups



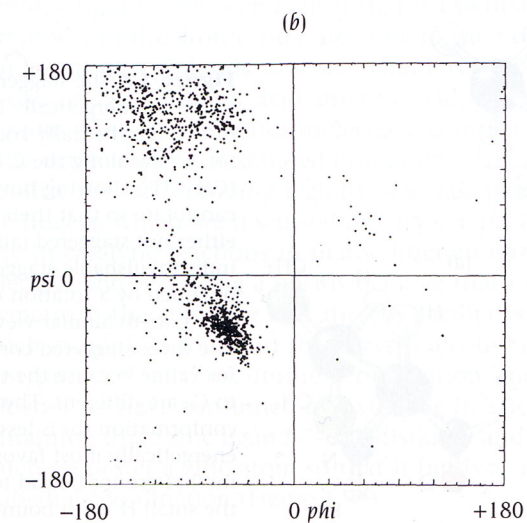
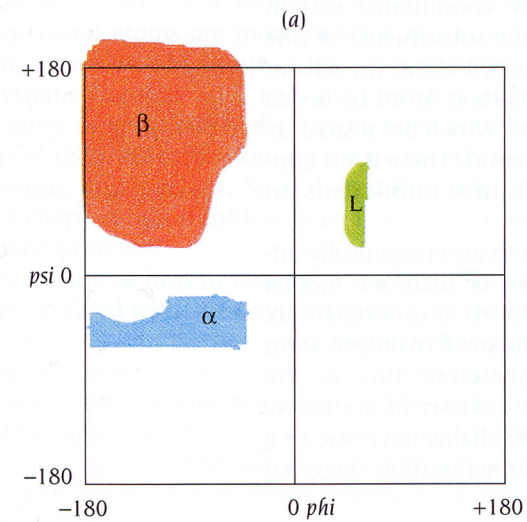
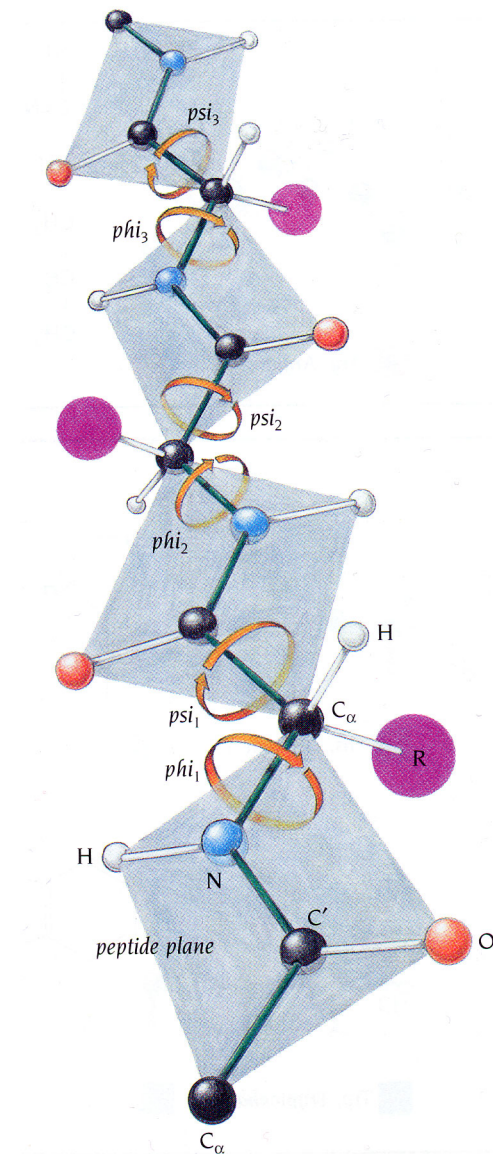
Primärstruktur: Die Seitenketten der Aminosäuren



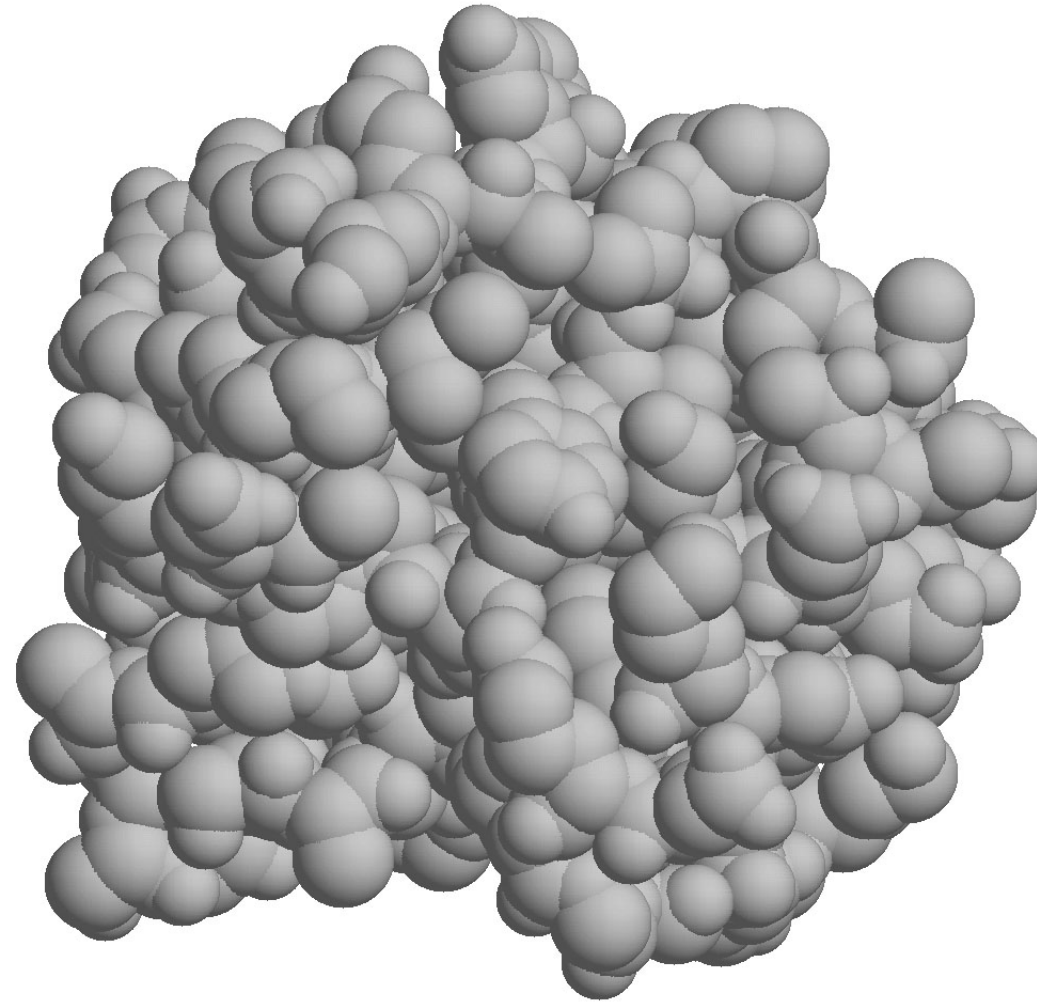
Primärstruktur: Die Eigenschaften der Aminosäuren



Sekundärstruktur: Die Eigenschaften der Hauptkette

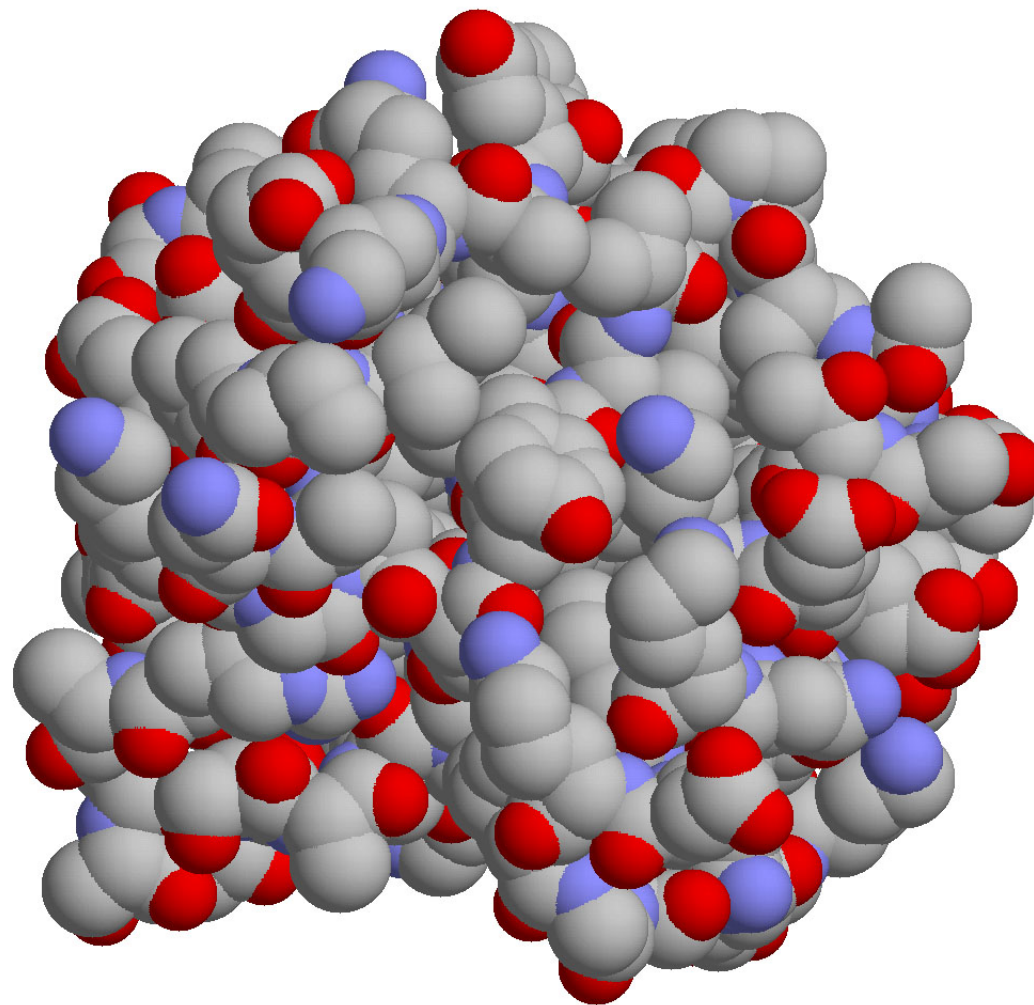


Tertiärstruktur: schematische Darstellungen



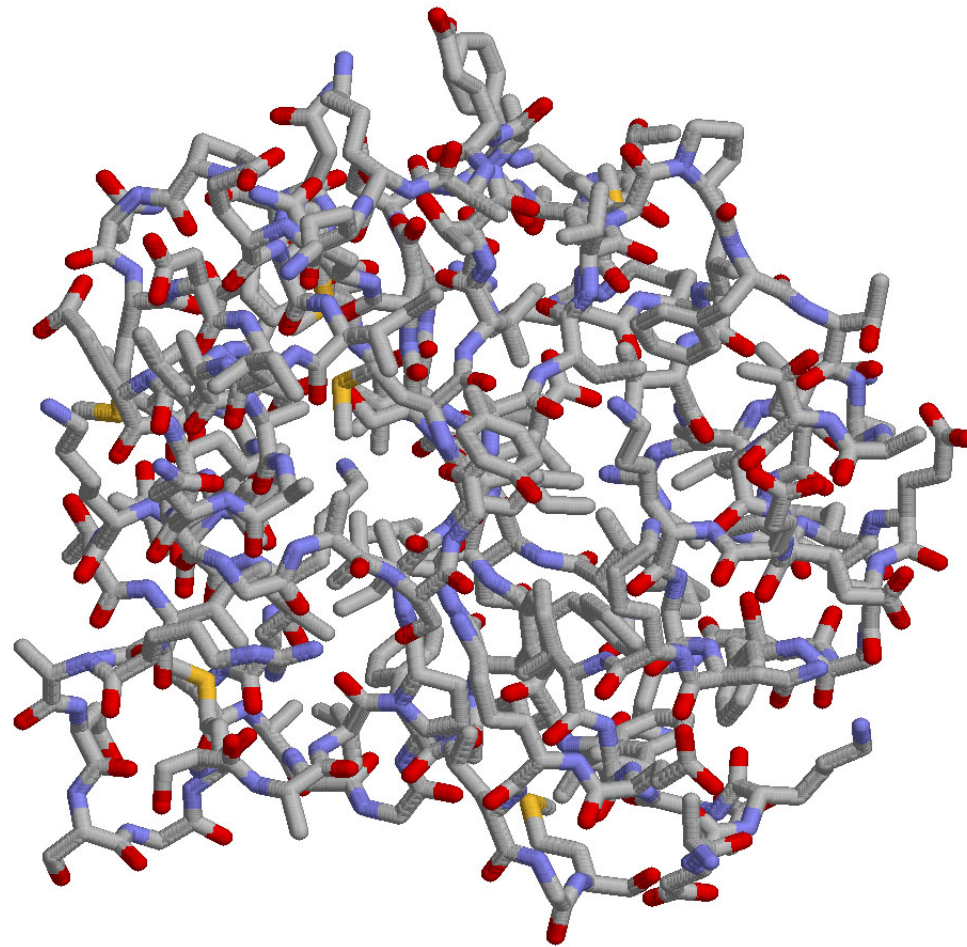
space-filling

Tertiärstruktur: schematische Darstellungen



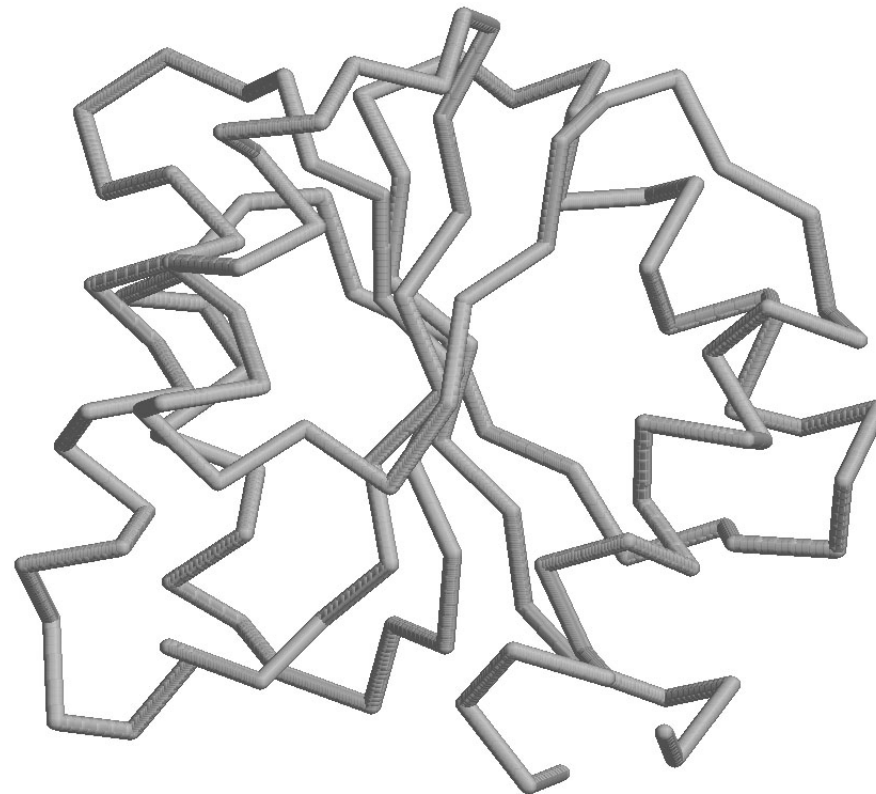
space-filling with CPK colors

Tertiärstruktur: schematische Darstellungen



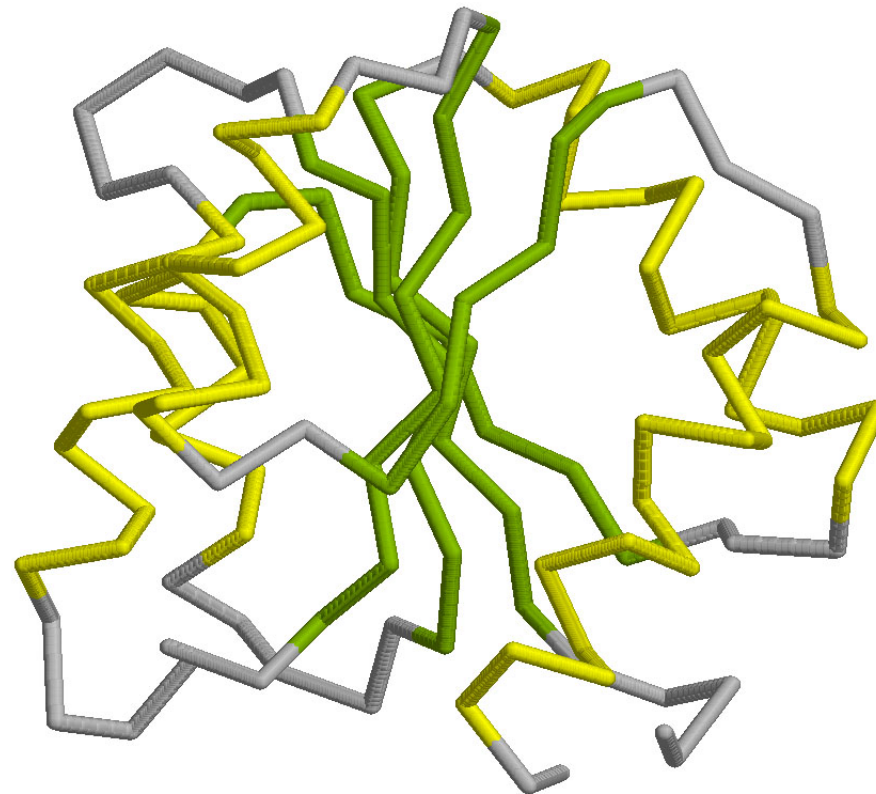
sticks with CPK colors

Tertiärstruktur: schematische Darstellungen



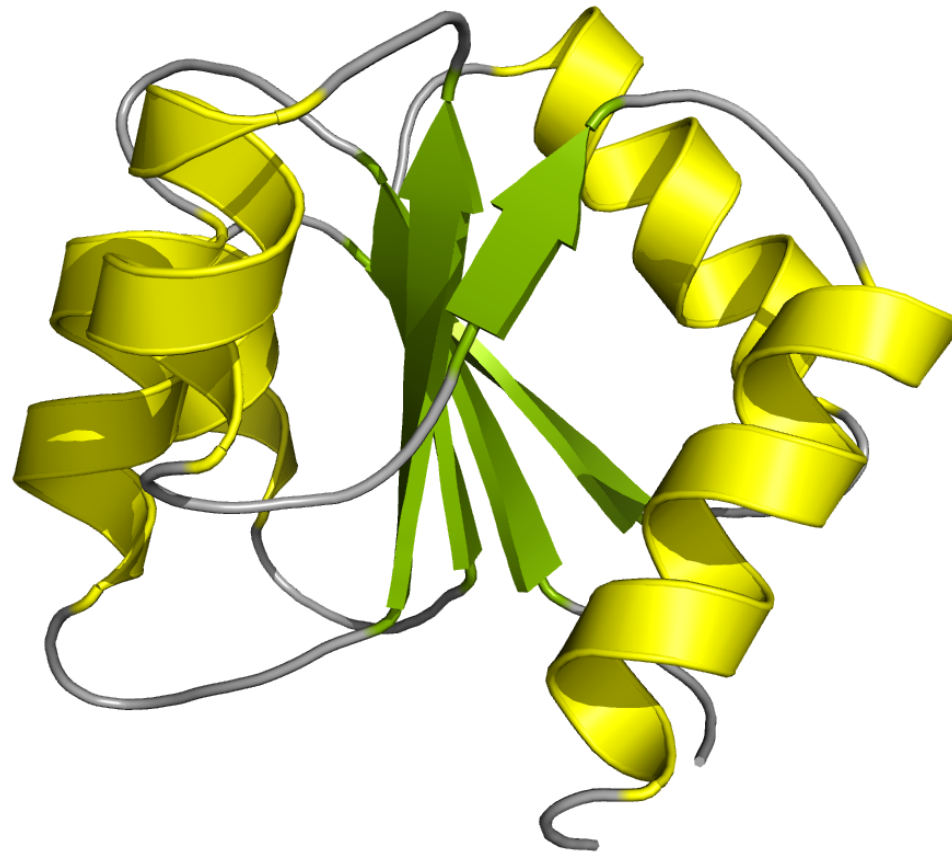
C-alpha trace

Tertiärstruktur: schematische Darstellungen



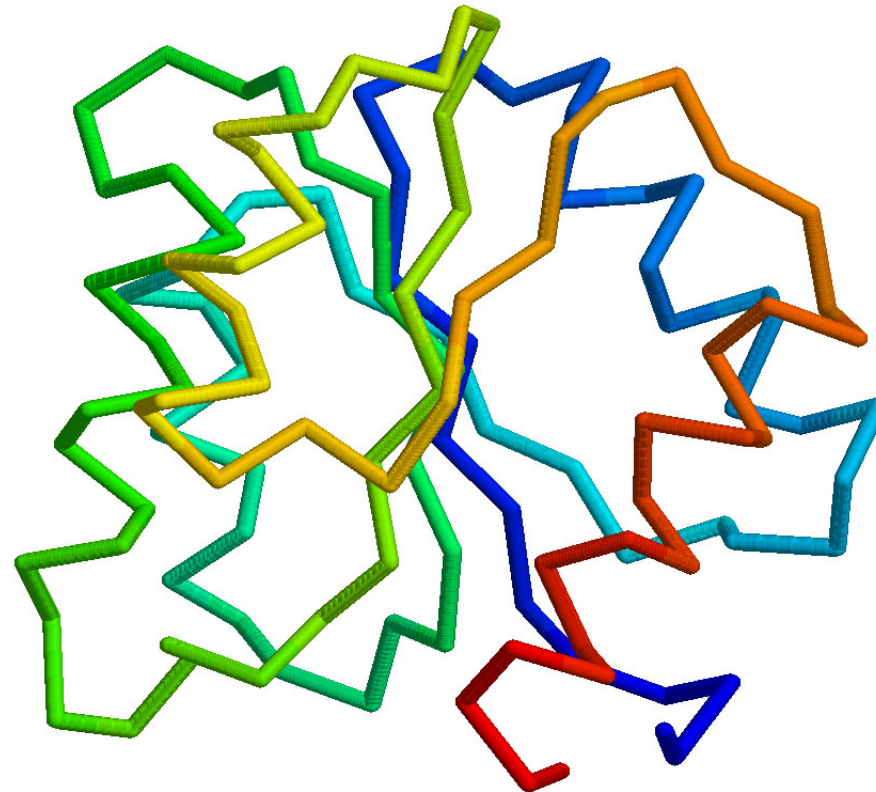
C-alpha trace with colored secondary structure

Tertiärstruktur: schematische Darstellungen

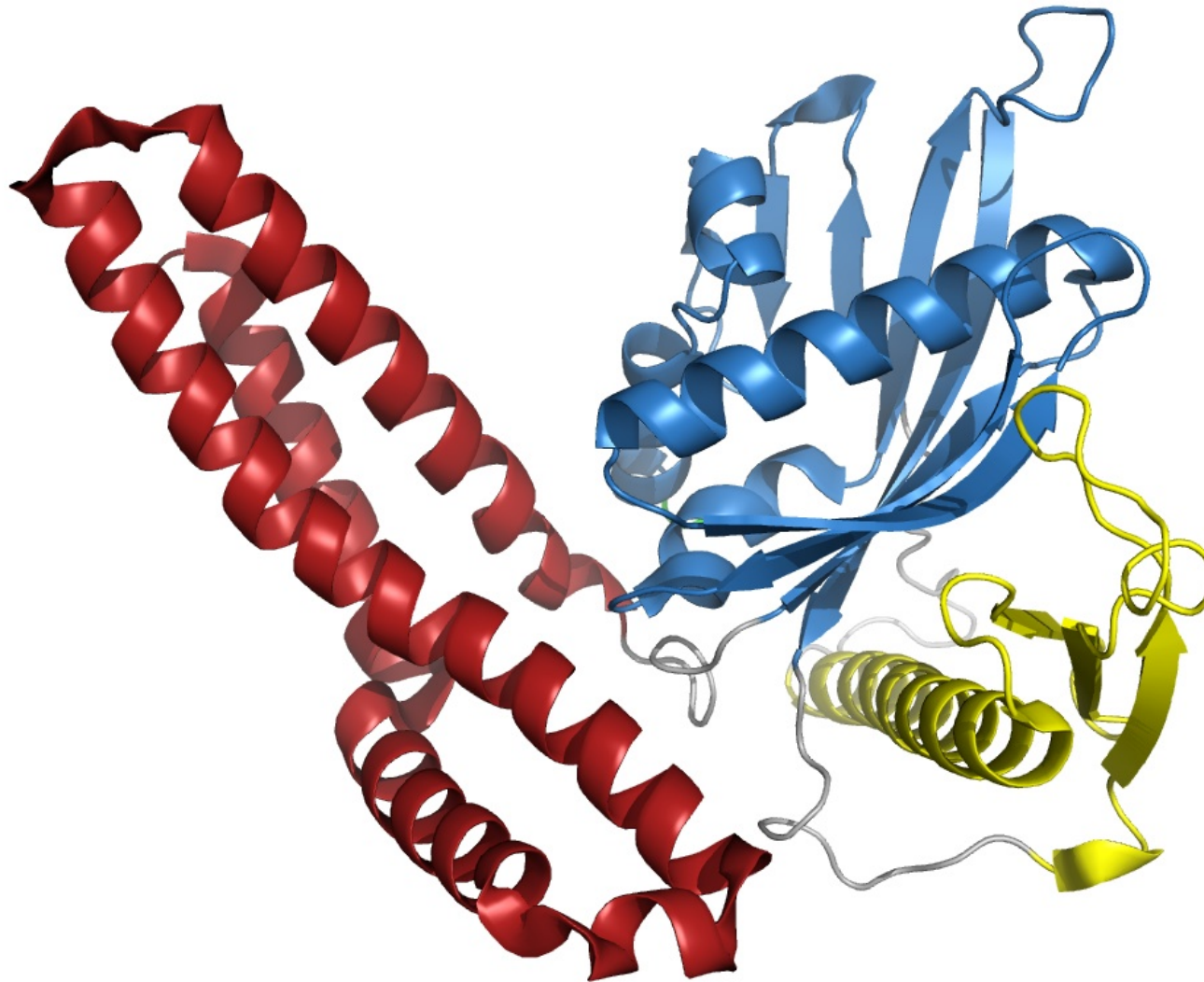


cartoon

Tertiärstruktur: Verlauf einer Polypeptidkette



Tertiärstruktur: Die Domäne als Einheit der Faltung



Polypeptide chain release factor 2 (RF2) – 1gqe

Überlagerung von Ubiquitin aus Mensch und Hefe

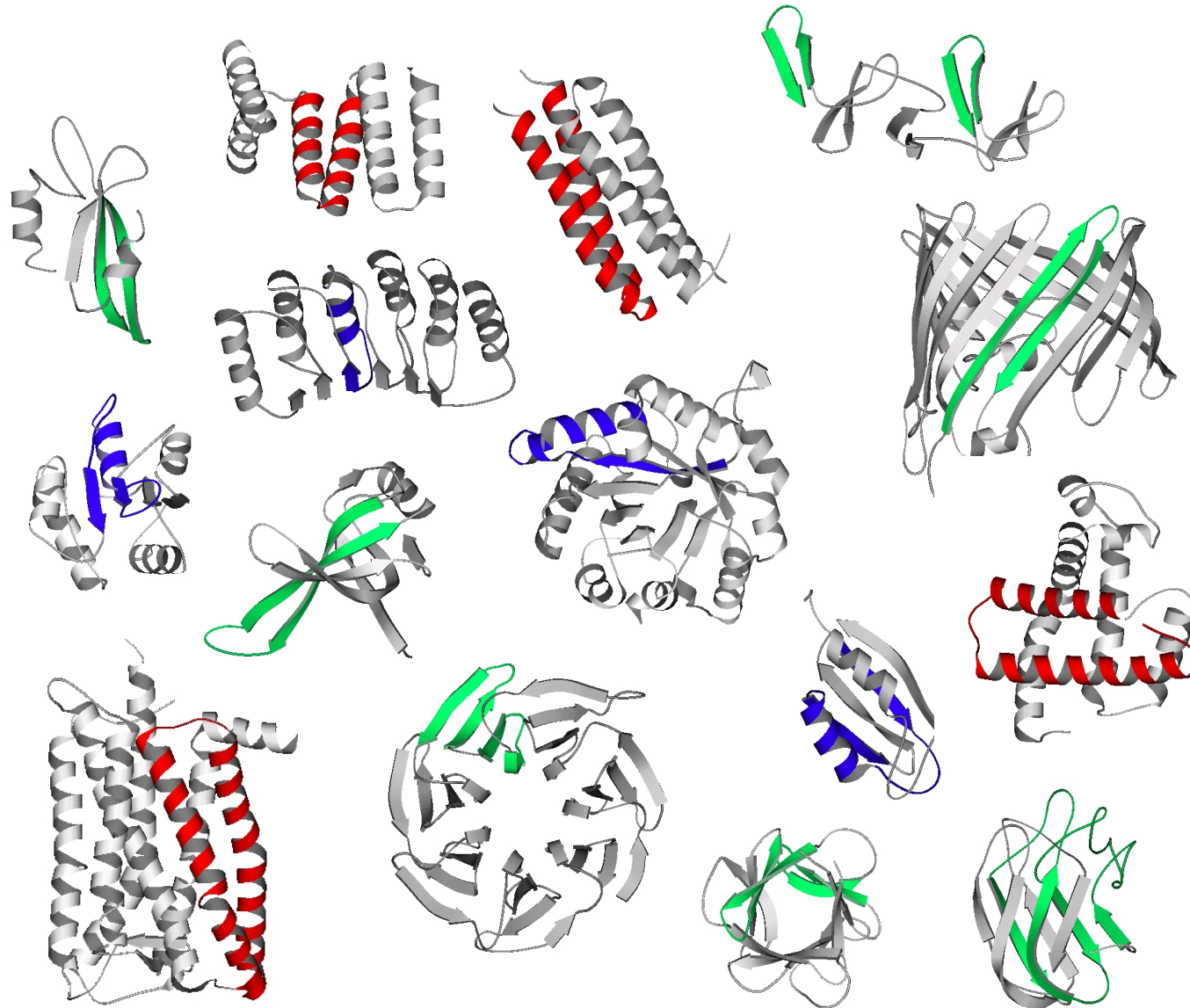
Überlagerung von SUMO und Ubiquitin aus Hefe

Das Proteinfaltungsproblem



- Proteine müssen falten um aktiv zu sein
- Proteinfaltung ist ein komplizierter, leicht zu störender Prozess
 - Zellen enthalten viele Faltungsfaktoren um Faltung zu erleichtern
 - Fehlgefaltete Proteine werden normalerweise zügig abgebaut
 - Trotzdem beruhen viele Krankheiten auf Fehlfaltung (Mukoviszidose, Morbus Alzheimer, Morbus Parkinson, BSE, Chorea Huntington)
- Wenn Proteine als zufällige Ketten von Aminosäuren synthetisiert würden, würde nur ein verschwindend geringer Anteil davon falten ($<1:10^{10}$)
- Die Natur entgeht dem Faltungsproblem bei der Evolution neuer Proteine, indem sie Teile von Proteinen benützt, die schon falten können (Domänen)

Die meisten Proteine bestehen aus Domänen, die auf etwa 5000 ancestrale Formen zurückgehen



Das Proteinfaltungsproblem



- Proteine müssen falten um aktiv zu sein
- Proteinfaltung ist ein komplizierter, leicht zu störender Prozess
 - Zellen enthalten viele Faltungsfaktoren um Faltung zu erleichtern
 - Fehlgefaltete Proteine werden normalerweise zügig abgebaut
 - Trotzdem beruhen viele Krankheiten auf Fehlfaltung (Mukoviszidose, Morbus Alzheimer, Morbus Parkinson, BSE, Chorea Huntington)
- Wenn Proteine als zufällige Ketten von Aminosäuren synthetisiert würden, würde nur ein verschwindend geringer Anteil davon falten ($<1:10^{10}$)
- Die Natur entgeht dem Faltungsproblem bei der Evolution neuer Proteine, indem sie Teile von Proteinen benützt, die schon falten können (Domänen)
- Domänen sind die Einheit der Proteinfaltung, sind jedoch zu komplex um de novo entstanden zu sein (Sequenzkomplexität im Bereich von 20^{100})
- Die Prozesse die zur Evolution von Domänen geführt haben sind substanziell unbekannt

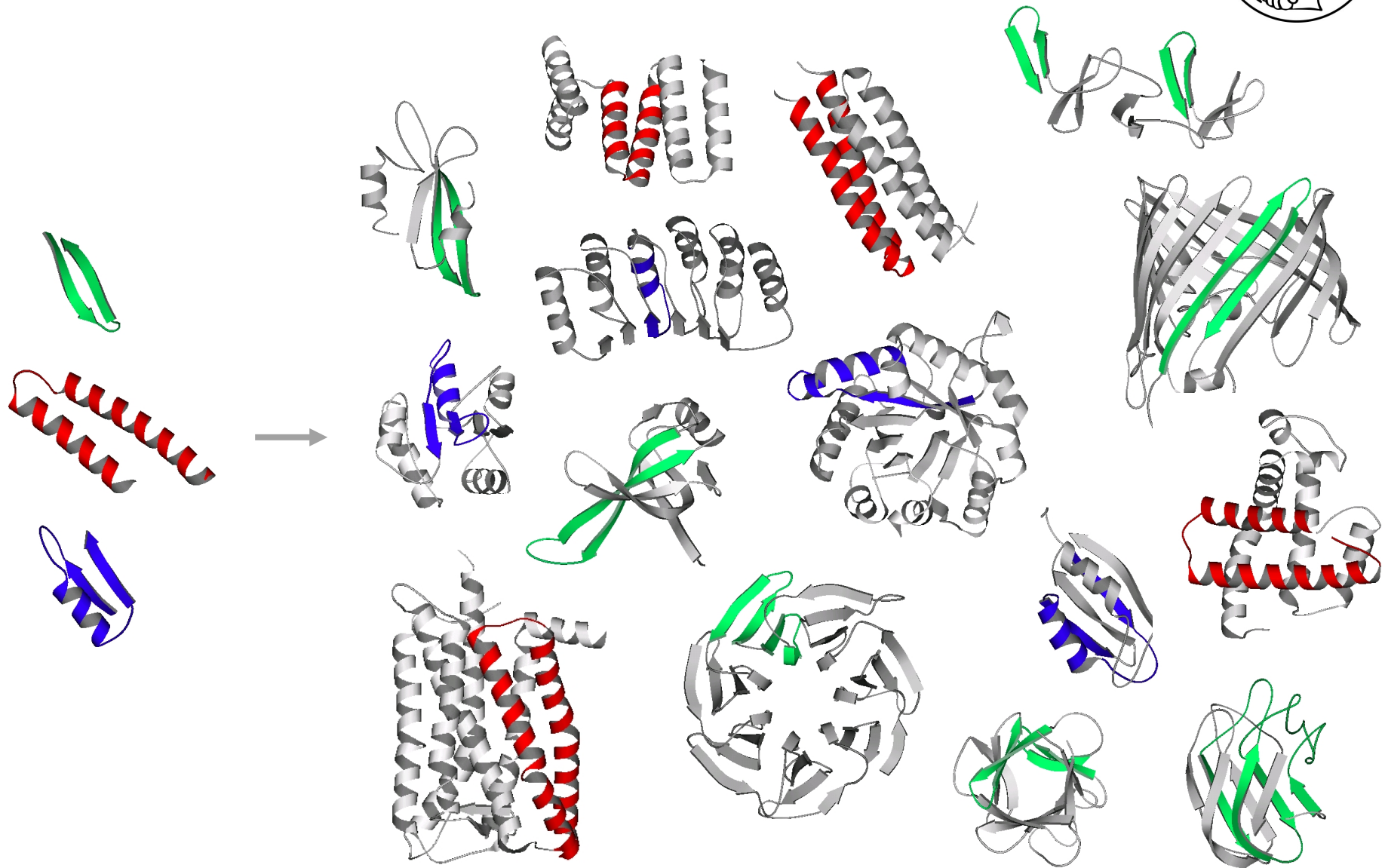
Das Proteinfaltungsproblem - Wie hat die Natur am Ursprung des Lebens stabile Faltungsformen entdeckt?



Hypothese:

- die ersten gefalteten Proteine entstanden durch Kombination und Verkettung eines begrenzten Satzes aus Urpeptiden
- diese waren durch abiotische Prozesse entstanden und dienten in einer auf Ribonukleinsäuren basierenden Vorform des Lebens als Kofaktoren für Replikation und Katalyse
- die Peptide falteten anfangs an Ribonukleinsäure-Molekülen wie an Gerüsten
- durch Aneinanderlagerung und später durch Verkettung gelang es ihnen selbständig zu falten
- schließlich verdrängten sie schrittweise die Ribonukleinsäuren durch ihre höhere Vielfältigkeit und Flexibilität bei der Katalyse

Supersekundärstrukturen als wiederkehrende Strukturelemente in verschiedenen Faltungsformen

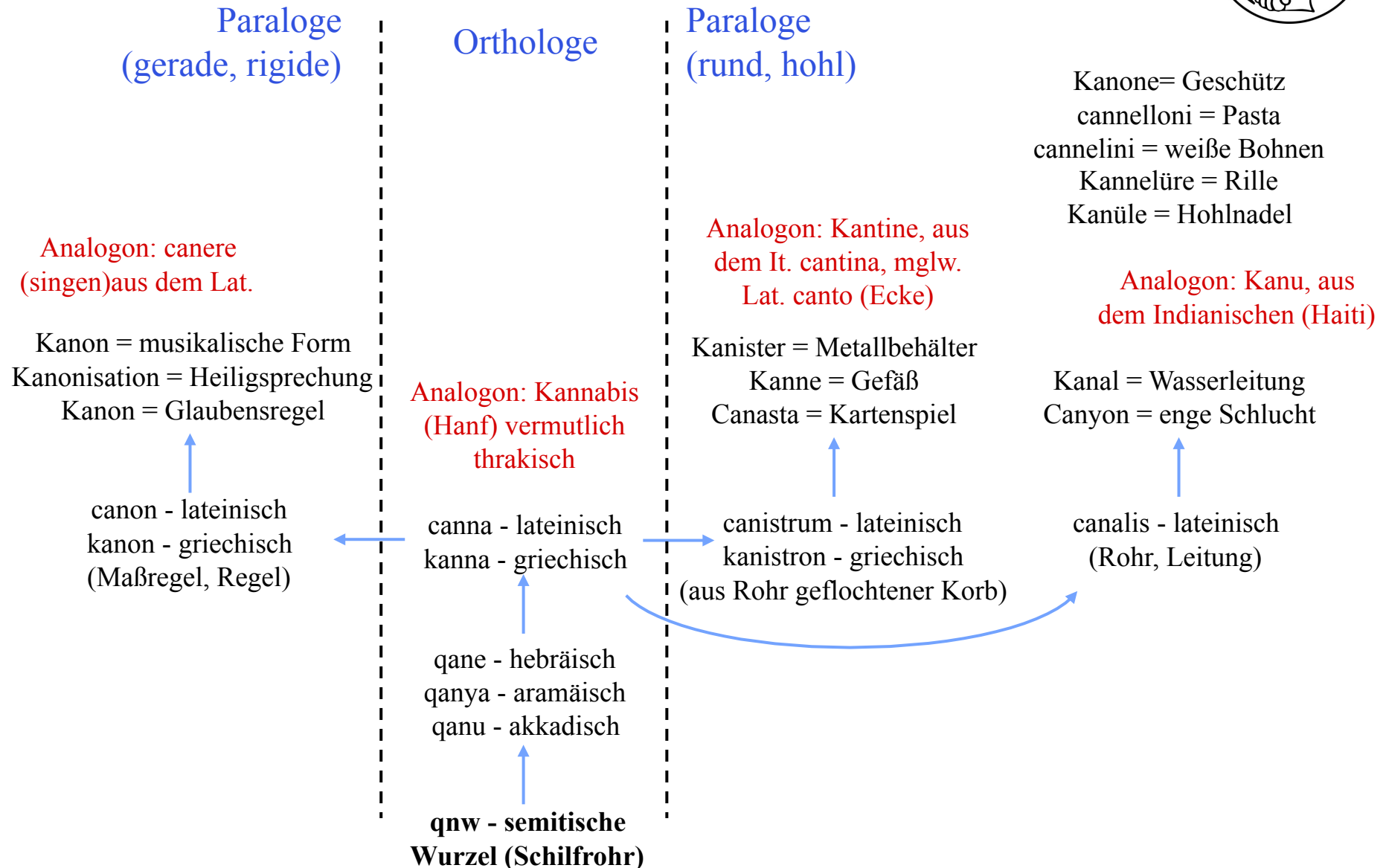


Urpeptide als ancestrale Supersekundärstrukturen



Sollte diese Hypothese zutreffen, müssten wir heute noch in den Sequenzen der Proteine die Spuren dieser Urpeptide finden, ebenso wie wir in europäischen Sprachen noch die Spuren ihres indoeuropäischen Ursprungs beobachten können.

Homologie von Proteinen - Sprache als Simile

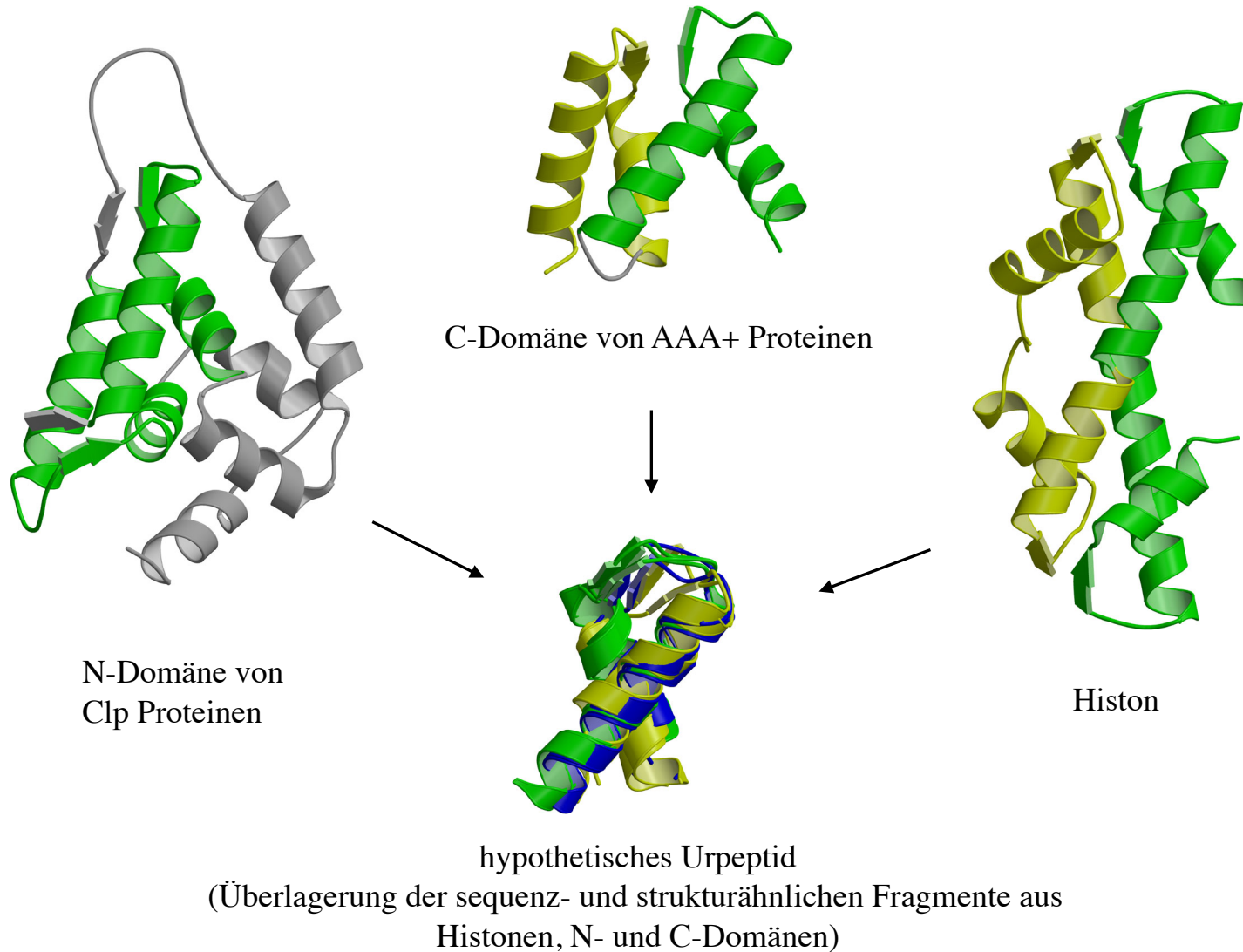


Auf der Suche nach ancestralen Supersekundärstrukturen: Sequenzvergleiche

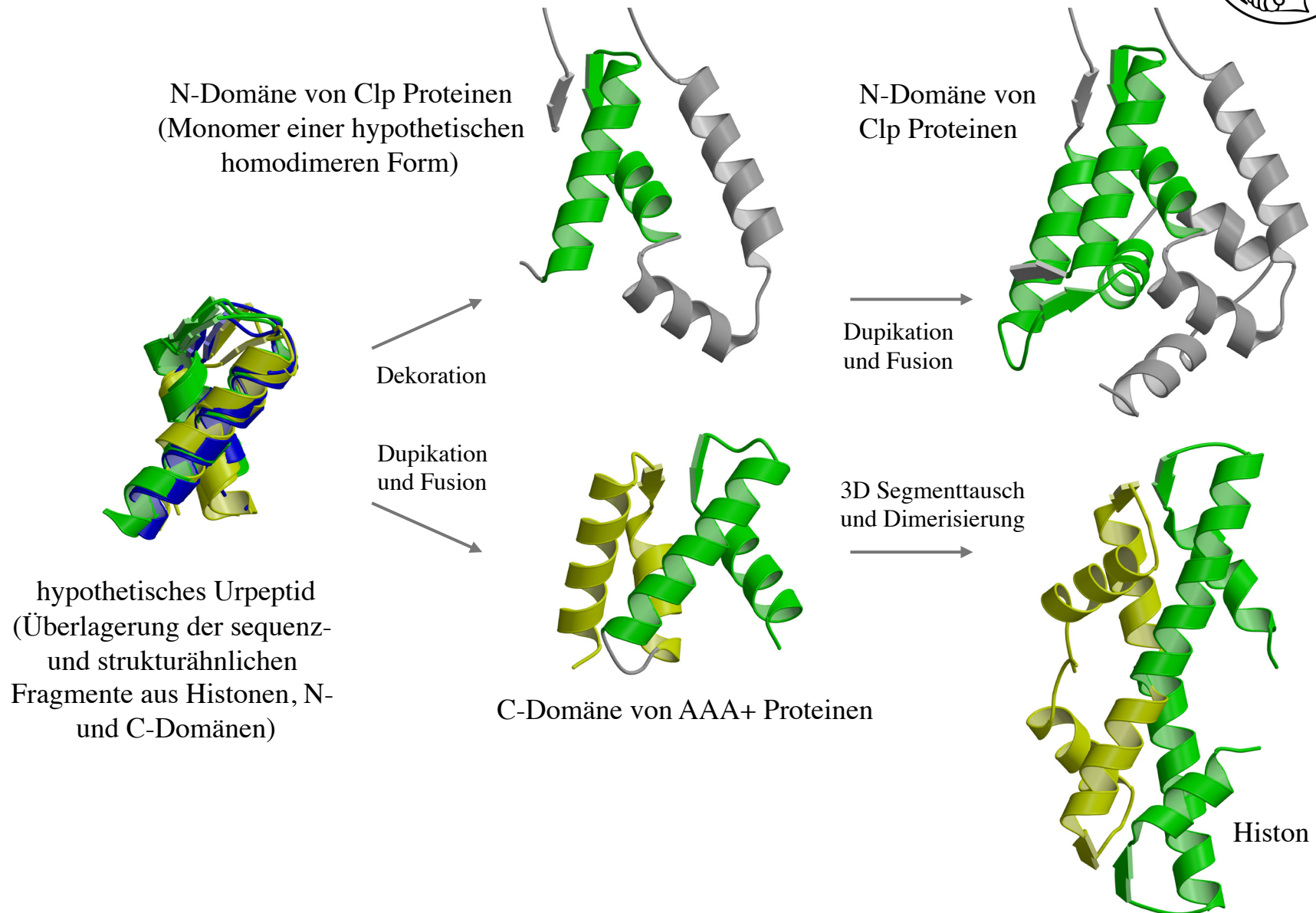


Histone fold	$\alpha 1$	$\beta 1$	$\alpha 2$ ----- $\alpha 2$	$\beta 2$	$\alpha 3$		
1b67_A (2)	-gelpIAPIGRIIKNAGae-----	RVS-DDARIALAKVL-----	EEMGEEIASEAVKLAKHAGR-	KTIK---AEDIELARKMf			
1taf_A (19)	--pkdAQVIMSILKELNVq-----	EYE-PRVVNQLLEFT-----	FRYVTSILDDAKVYANHARK-	KTIID---LDDVRLATEVt			
1taf_B (3)	gssisAESMKVIAESIGVg-----	SLS-DDAAKELAEDV-----	SIKLKRIVQDAAKFMNHAKR-	QKLS---VRDIDMSLKV-			
1tzy_A (21)	glqfpVGRVHRLLLRKGNyae-----	RVG-AGAPVYLA AVL-----	EYLTAEILELAGNAARDNKK-	TRI I---PRHLQLAIRNd			
1tzy_B (33)	rkesySIYVYKVLKQVhpd-----	GIS-SKAMGIMNSFV-----	NDIFERIAGEASRLAHYNKR-	STIT---SREIQTAVRLl			
1tzy_C (62)	rklpfQRLVREIAQDFKtdl-----	RFQ-SSAVMALQEAS-----	EAYLVGLFEDTNLCAIHAKR-	VTIM---PKDIQLARRIr			
1tzy_D (27)	-qgitKPAIRRLARRGv-----	RIS-GLIYEETRGLV-----	KVFLENVIRD AVTYTEHAKR-	KTVT---AMDVVYALKRq			
C-domain of AAA+	$\alpha 1$	$\beta 1$	$\alpha 2$	hinge	$\alpha 3$	$\beta 2$	$\alpha 4$
lin4_A (181)	tVKELKEIKRAASLMDV-----	EIE-DAAAEMI AKR-srgt-----	PRIAIRLTKRVRDMLTVVKA-	DRIN---TDIVLKTMEVl			
1r6b_A (350)	sIETTVQIINGLKPkyeahhdv---	RYT-AKAVRAAVEL-avkyindrhl	PDKAIDVIDEAGARARLMpvsk	rktvnVADIESVVARi			
1lv7_A (326)	dVRGREQILKVHM--rrv-----	PLApdidAAIIARG-tpgfs----	GADLANLVNEAALFAARGNK-	RVVS---MVEFEKAKDKi			
1ny5_A (314)	rKEDIIPLANHFLKKFs rkyakeveg	FT-KSAQELLLSy pwygn-----	VRELKNVIERAVLFSegk---	FID---RGELSCLV---			
1g8p_A (220)	dVETRVEVIRRRDTYD V-----	EAP-NTALYDCAAL-cialgs--dg	LRGELTLLRSARALAALEGA-	TAVG---RDHLKRVAT--			
Clp-N motif		[$\beta 2$]	$\alpha 1$	$\beta 1$	$\alpha 2$		
1k6k_A (4)	-----	[qp t]	-----QELELSLNMAFARAREHRH-	EFMT---VEHLLLALLSn			
1k6k_A (82)	-----	[ml n]	-----LSFQRVLQRAVFHVQSSGR-	NEVT---GANVLVAIFSe			
1khy_A (8)	-----	[qp s]	-----NKFQLALADAQSLALGHDN-	QFIE---PLHLMSALLNq			
1khy_A (85)	-----	[rl t]	-----QDLVRVLNLCDKLAQKRGD-	NFIS---SELFVLAALes			

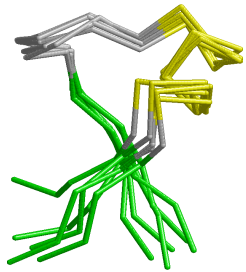
Auf der Suche nach ancestralen Supersekundärstrukturen: Strukturvergleiche



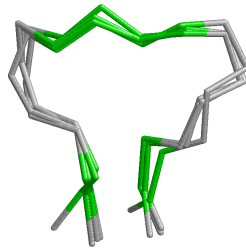
Auf der Suche nach ancestralen Supersekundärstrukturen: evolutionäre Szenarien



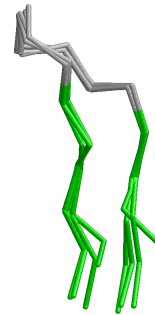
Ancestral Supersekundärstrukturen sind durch den Vergleich moderner Proteine heute noch erkennbar



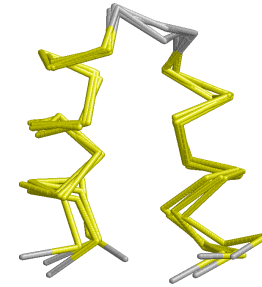
GD-box βαβ motif (present in 6 folds / 8 superfamilies)



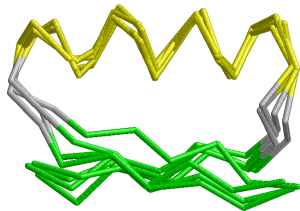
β-hammerhead motif (4/8)



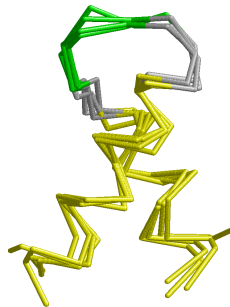
CBS domain β-hairpin (3/3)



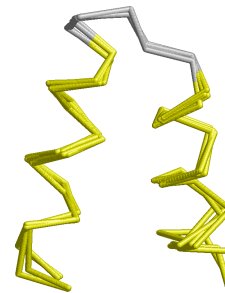
EF-Tu-bdg α-hairpin motif (2/2)



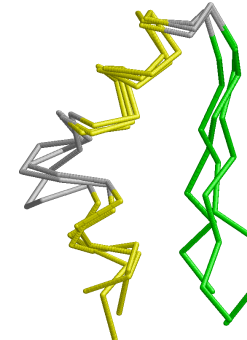
Nucleotide-bdg βαβ motif (10/10)



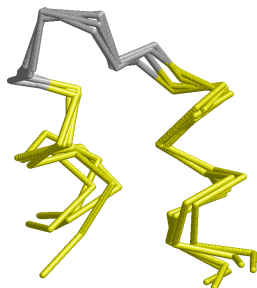
histone HtH motif (3/3)



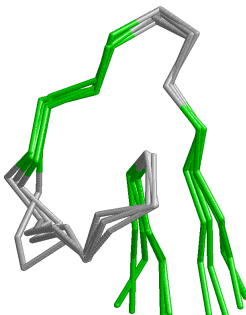
SKP1-dimerization motif (2/2)



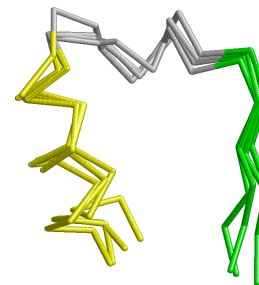
RNA-bdg KH-motif (2/2)



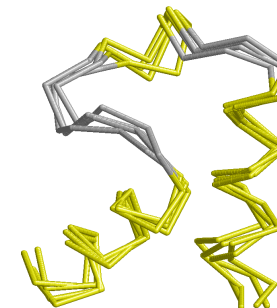
DNA-bdg HhH motif (18/24)



Alpha-L RNA-bdg motif (3/4)



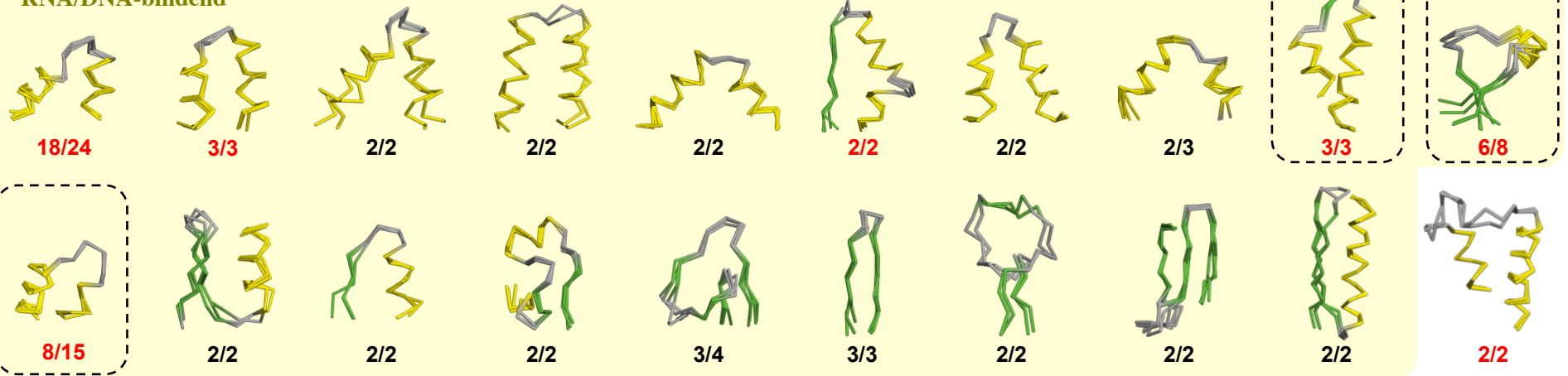
actin αβ motif (2/2)



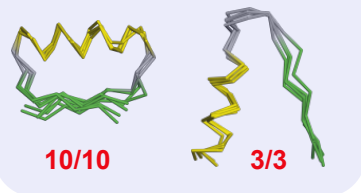
DNA-binding HttH motif (2/2)

Das Vokabular anzestraler Supersekundärstrukturen

RNA/DNA-bindend



Nukleotid-bindend



Metall-bindend

