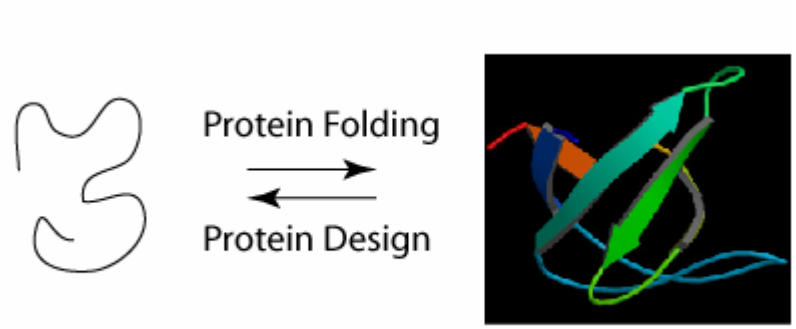
Protein design

- Branden & Tooze, Chapter 17
- Protein design is an extreme form of protein engineering, in which the entire polypeptide chain is built in its entirety rather than mutated piecemeal
  - –may or may not involve designing the backbone
    typically does not use an existing sequence
  - –sometimes called “de novo” protein design to stress the “ground-up” approach

- Can be characterized as “inverse protein folding” because one starts from structure and predicts the amino acid sequence



- Large number of possible amino acid sequences makes the problem *NP*-hard
  - –Pierce and Winfree, Protein Engineering 15, 779 (2001)

Knowledge-based design

Known structures contain information that may be used productively during a design study

- statistical data, e.g. distribution of amino acids on a helix
- modeling on a computer

Think small and dream big

- introduce elements to form and stabilize secondary structures
- hydrophobic patterns on a helix
- engineer connectivity by introducing turns, loops,

Drawbacks

- The problem can quickly get out of hand due to the large number of variables that must be evaluated
- “Knowledge” is very personal and difficult to objectify

Helix bundles

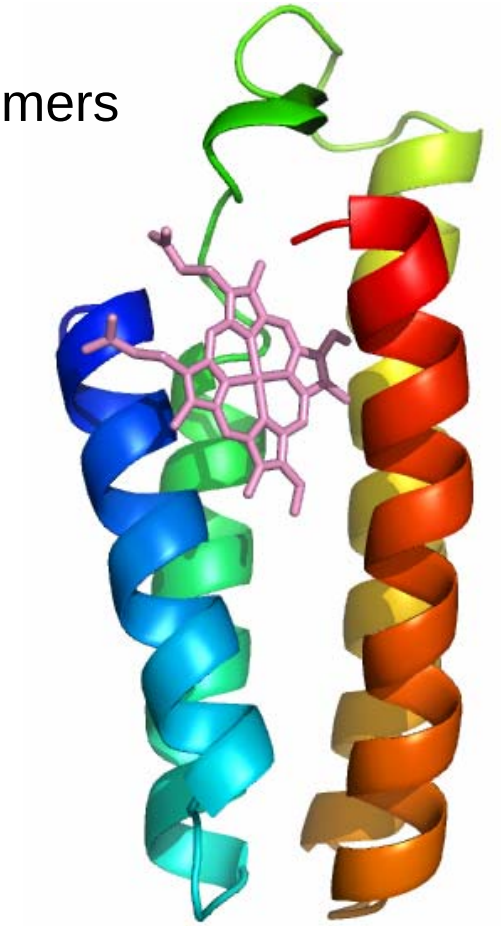
Helix bundles are biologically important and common in many structural proteins, transcription factors (coiled coil), and enzymes

Stable and soluble—less likely to create run-away oligomers

Helices are easier to design de novo because they are stabilized through local interactions, and the factors that contribute to their formation are better understood

Easy to characterize experimentally

- CD, fluorescence
- thermal and chemical denaturation

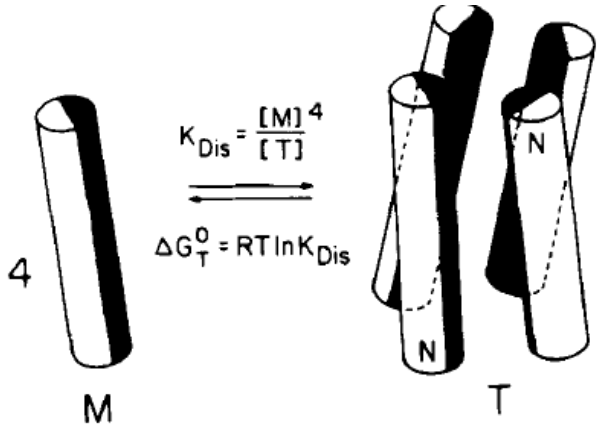


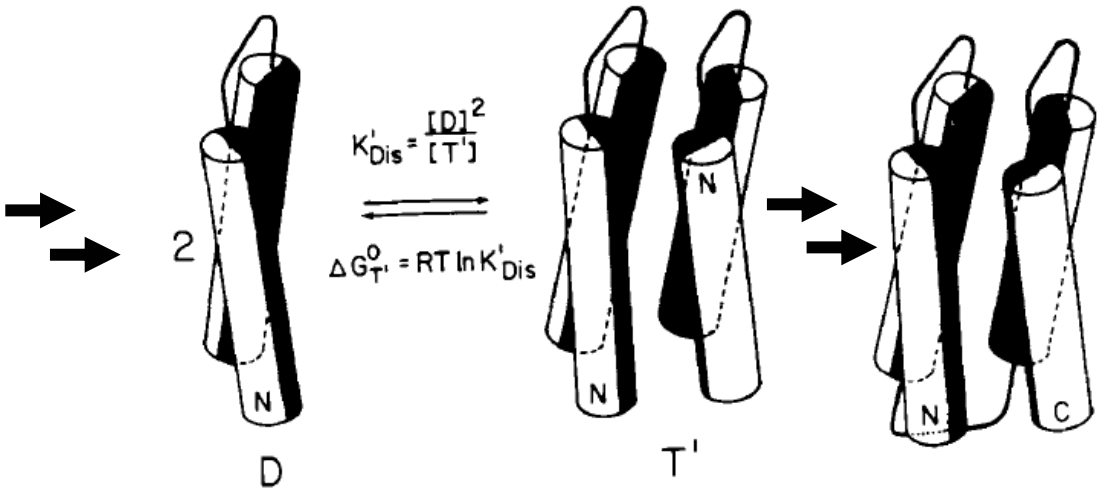
cytochrome b562

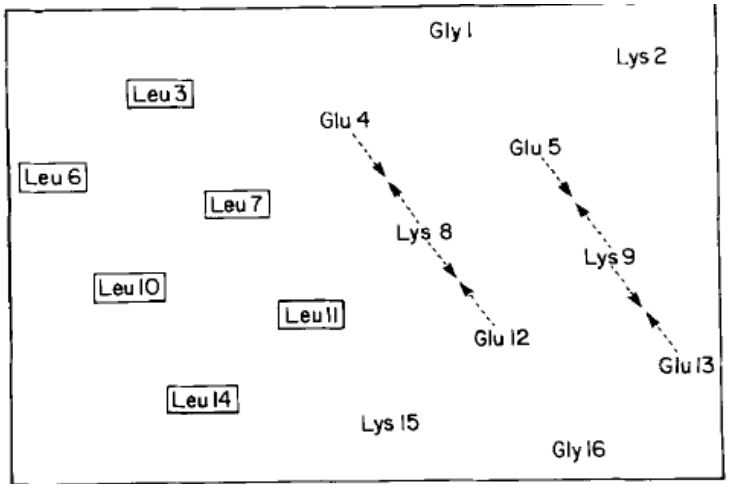
Designing helix bundles

Design 12 and 16 residue peptides that are likely to form helices and binary pattern them using Leu on one side, and Glu and Lys on the other

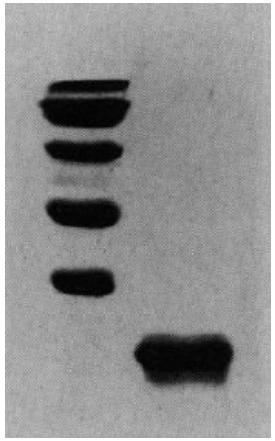
Hydrophobic surfaces should come together to avoid contact with solvent

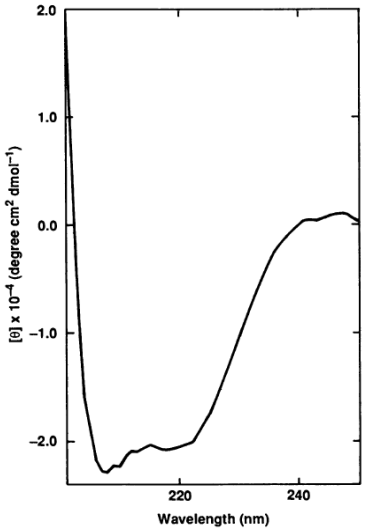






Eisenberg et al, Proteins 1,16 (1986)





Design details

Helix: GELEELLKKLKELLKG, Loop: PRR

α4: Met-Helix-*loop*-Helix-*loop*-Helix-*loop*-Helix-COOH

Regan and DeGrado, Science 241, 976 (1988)

Protein	ΔG^{H_2O} (kcal mol^{-1})
α_4	−22.5
Lysozyme	−8.9
Myoglobin	−7.6
Ribonuclease A	−7.5
α-Lactalbumin	−4.2

Binary patterned helix can lead to unpredicted oligomeric states—a 12 residue fragment accidentally synthesized during the study (ELLKKLLEELKG) self-associates into a mix of hexamer and tetramer

Hill et al, Science 249, 543 (1990)

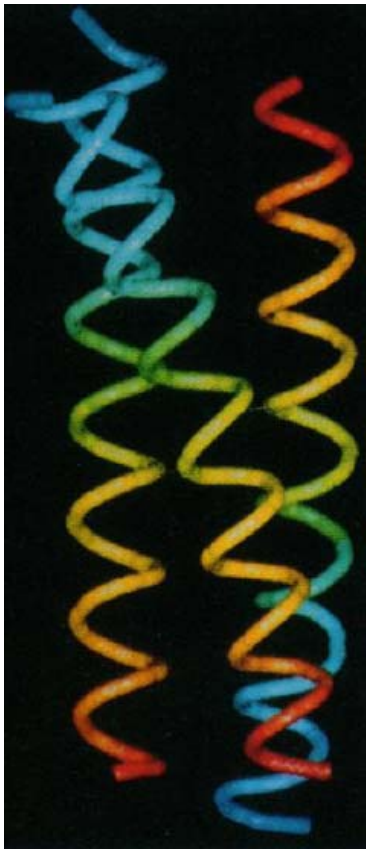
Not all amphiphilic helices associate into a four helical bundle
Side chain interdigitation is necessary—computer modeling

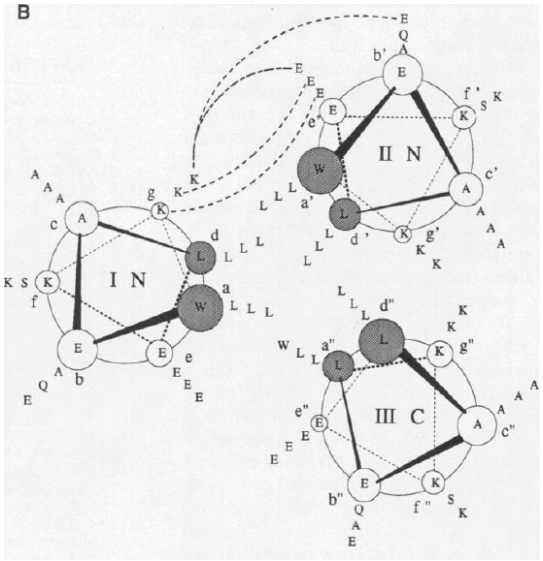
Coiled-coil design

"Coil-Ser" was designed to form a coiled-coil based on previous studies with the helical heptad repeat in mind (i.e. "**a**" and "**d**" hydrophobic = Leu)

studies show L is preferred at "d" whereas L, I, V, M are tolerated at "a"

Position	g	abcdefg	abcdefg	abcdefg	abcdefg	abcdef
Hodges	K	**L**EA**L**EGK	**L**EA**L**EGK	**L**EA**L**EGK	**L**EA**L**EGK	**L**EA**L**EG
GCN4-p1	R	**M**KQ**L**EDK	**V**EE**LL**SK	**N**YH**L**ENE	**V**AR**L**KKL	**V**GE**R**
coil-Ser	E	**W**EA**L**EKK	**L**AA**L**ESK	**L**QA**L**EKK	**L**EA**L**EHG	





Why a 3 helix bundle rather than a coiled coil?

- neutralization of macrodipole moment
- intermolecular salt bridges
- more favorable side chain diheral angles

Lovejoy et al, Science 259, 1288 (1993)

Determinants of coiled-coil topology

Coiled coil proteins with a heptad repeat sequence can form dimers, trimers and tetramers

Geometric properties of buried residues may influence the overall structure

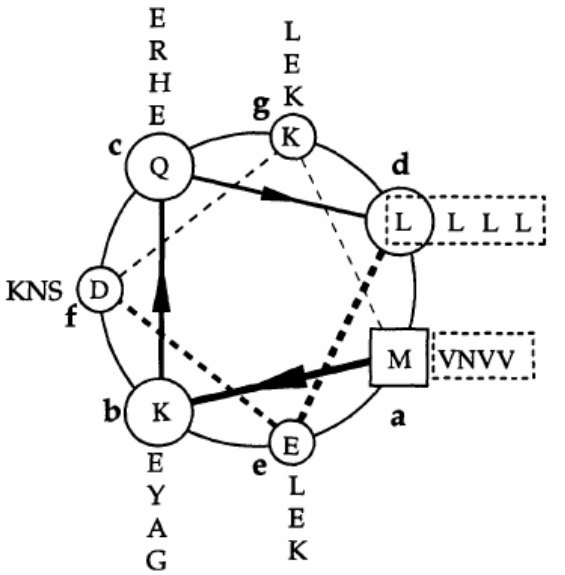


Table 1. Core mutants of GCN4-p1 form stable two-, three-, and four-helix structures.

Positions*		$-[\theta]_{222}$ (deg cm^2 dmol^{-1})	T_m (°C)	T_m^{GdmCl}† (°C)	No. of helices‡	
a	**d**				Unmodified	SS
GCN4-p1		33,300	53	<0	2	2
I	L	32,400	>100	77	2	2
I	I	32,400	>100	70	3	6
L	I	30,600	>100	94	4	4
V	I	22,500§	73	<0	–	(4,6)
L	V	30,600	81	<0	3	(2,–)
V	L	32,400	95	49	(2,3)	2
L	L	31,500	>100	76	3	(2,4,6,–)

*The residues inserted at four **a** and four **d** positions of GCN4-p1 (*41*). †T_m^{GdmCl} denotes the melting temperature in 3 M GdmCl. All scans and melts were performed at 10 μM peptide concentration. ‡The number of helices in the solution complex formed by unmodified peptides and by disulfide bonded peptides. The first four peptides were assigned on the basis of equilibrium analytical ultracentrifugation data and the last four on the basis of gel filtration data. Parentheses indicate that multiple species were present; a dash indicates the presence of a species that could not be assigned. §p-VI exhibits a $-[\theta]_{222}$ value of 31,500 deg cm^2 dmol^{-1} at 150 μM concentration.

Harbury et al, Science 262, 1401 (1993)

Designing new topology

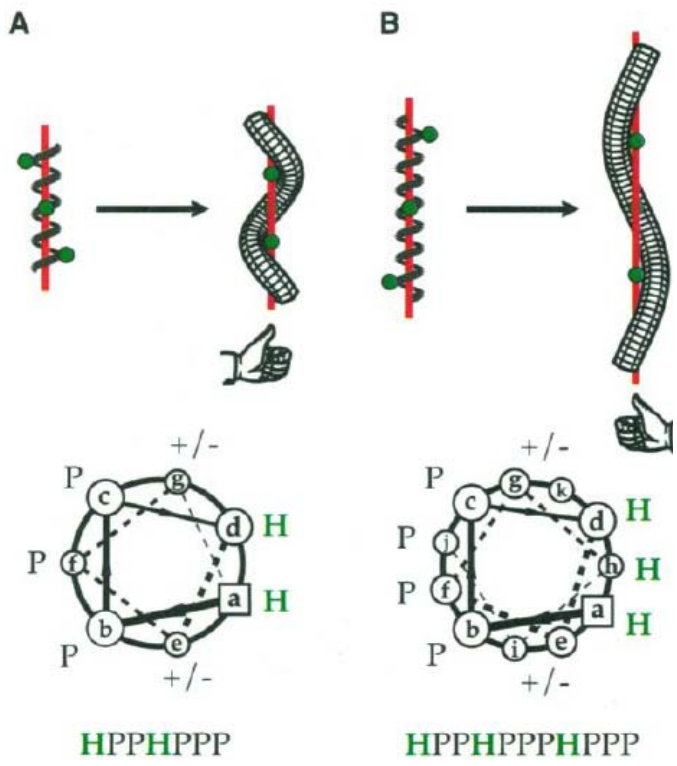
All coiled-coil proteins are left-handed

- 3.6 residues/turn ➔ 7 residues = 700° < 2 complete turns
- entire helix needs to coil around the superhelical axis by 20° every 7 residues to maintain hydrophobic contacts

Is it possible to design a right-handed coiled coil?

- a sequence based on an 11 residue repeat
- 1100° > 3 complete turns (by 20°)

Incidentally, test the effect of burying hydrophobic residues (a, d, h, or 1, 4, 8) on defining a topology



Harbury et al, Science 282, 1462 (1998)

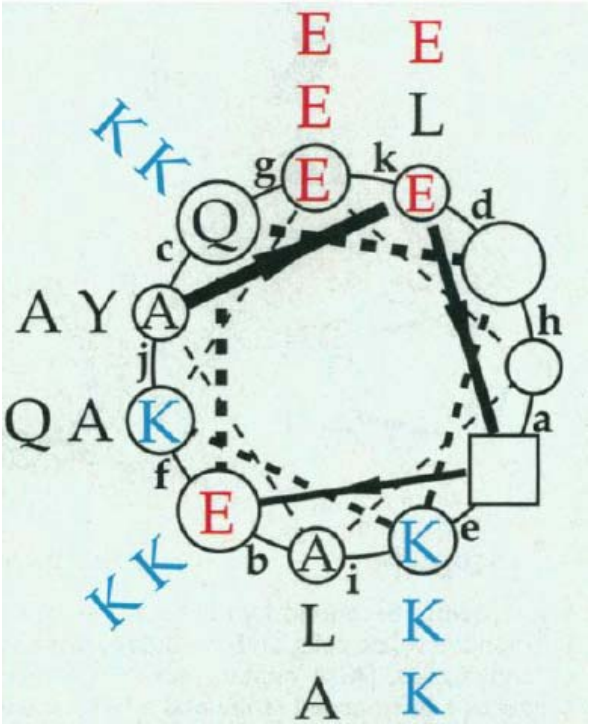
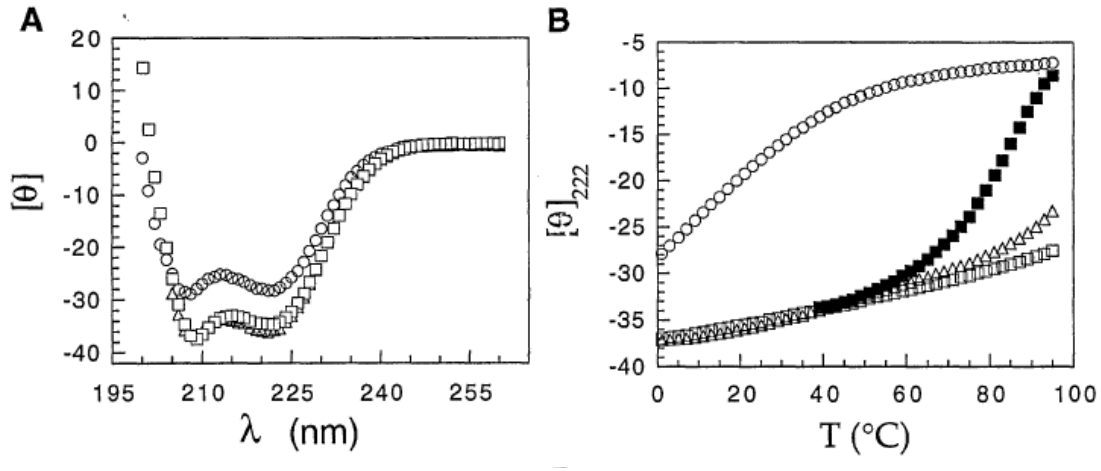
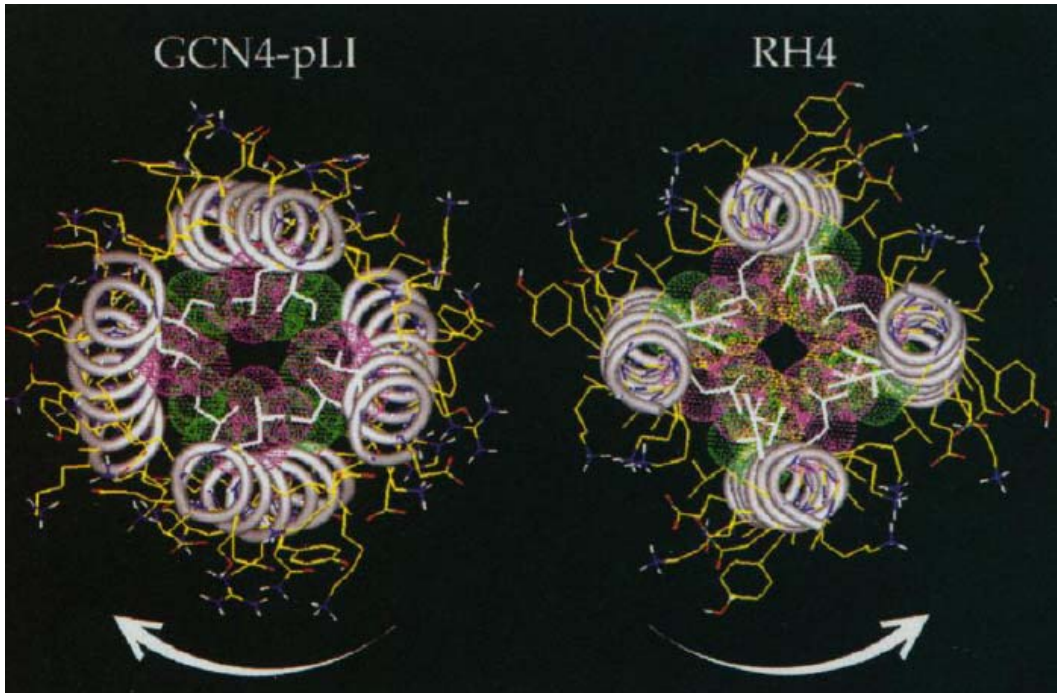


Table 1. Top sequence solutions for right-handed dimer, trimer, and tetramer coiled coils according to the packing calculation. The table consists of three lists, the first sorted according to dimer stability, the second according to trimer stability, and the third according to tetramer stability.

Sequence†			Stability‡	Specificity§	
a	d	h			
Dimer				Dimer-trimer	Dimer-tetramer
nV	L	nV	2.0	0.9	1.3
nV	nV	L	1.9	0.8	0.7
nV	L	L	1.9	0.7	0.9
I	L	nV*	1.8	1.8	1.0
Trimer				Trimer-dimer	Trimer-tetramer
nV	aI	L	2.6	0.3	−0.5
I	aI	L	2.4	0.5	0.2
I	aI	nV	2.4	0.6	0.6
nV	aI	nV	2.3	0.6	−0.4
L	aI	L	2.0	−0.1	−0.8
aI	aI	I*	2.0	1.9	2.0
Tetramer				Tetramer-dimer	Tetramer-trimer
nV	aI	I	3.1	2.2	1.7
nV	aI	L	2.6	0.6	0.5
L	aI	I*	2.5	1.9	2.0
L	aI	L	2.3	0.3	0.8

aI = alloisoleucine, nV = norvaline

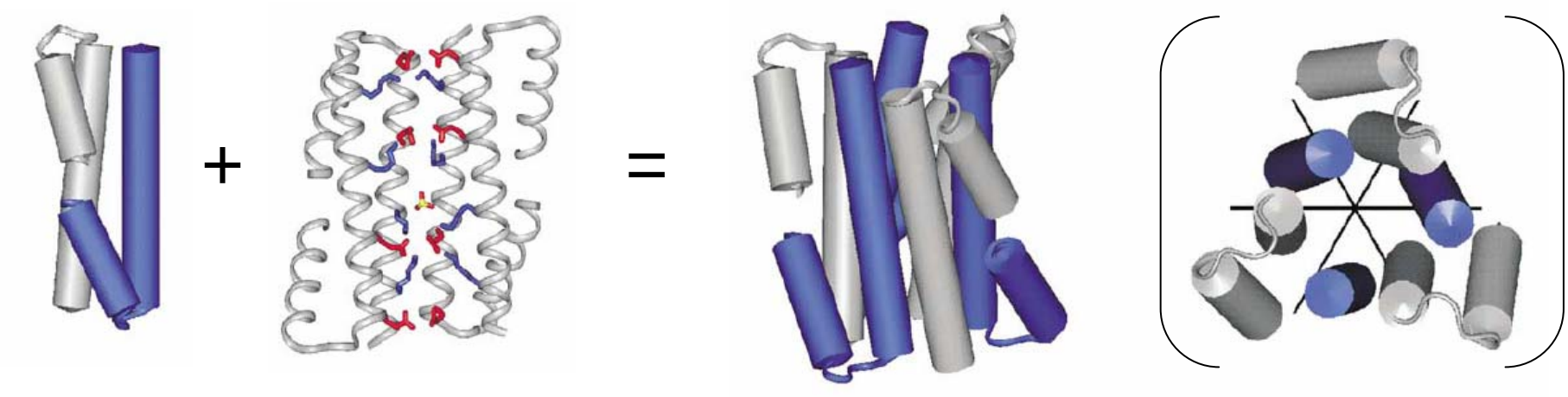

A
20
10
0
-10
-20
-30
-40
[θ]
195
210
225
240
255
λ (nm)
B
-5
-10
-15
-20
-25
-30
-35
-40
[θ]222
0
20
40
60
80
100
T (°C)


GCN4-pLI
RH4

Hexameric helix bundle

Apply symmetry to simplify the design task and engineer a bigger protein complex

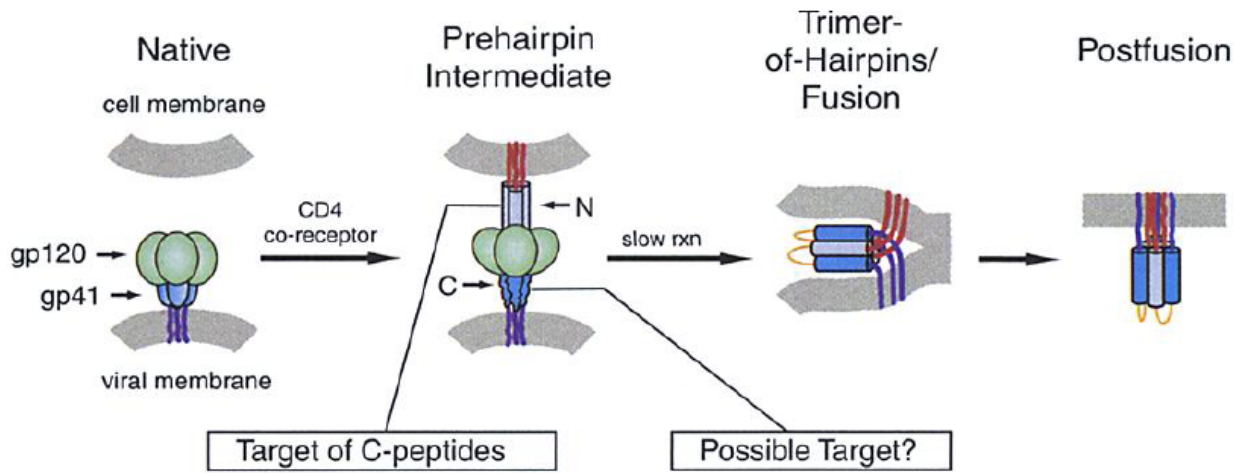
Introduce salt bridges at the interface to stabilize interaction

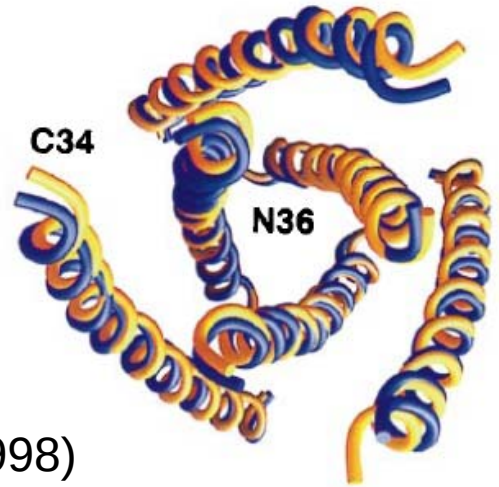


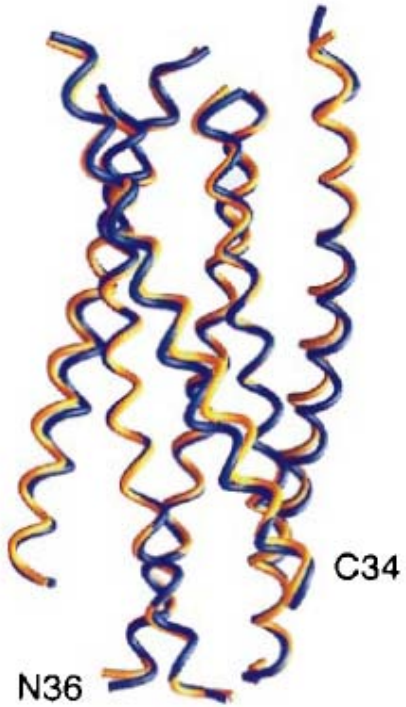
Ghirlanda et al, JMB 319, 243 (2002)

Helix bundle as HIV-1 inhibitor

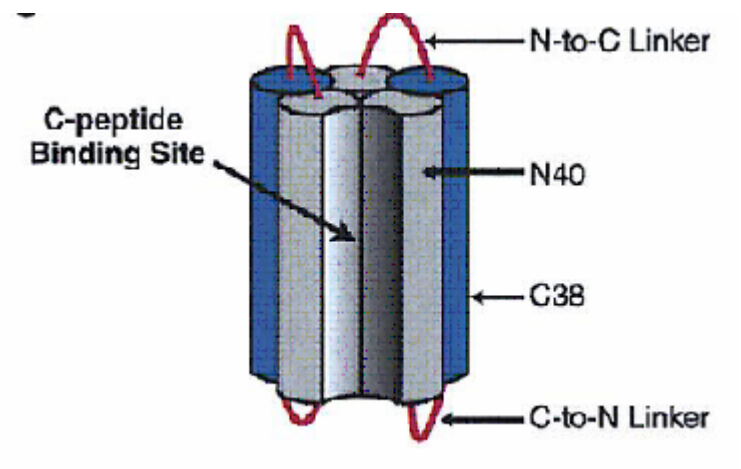
Helix bundle may be engineered for medical applications







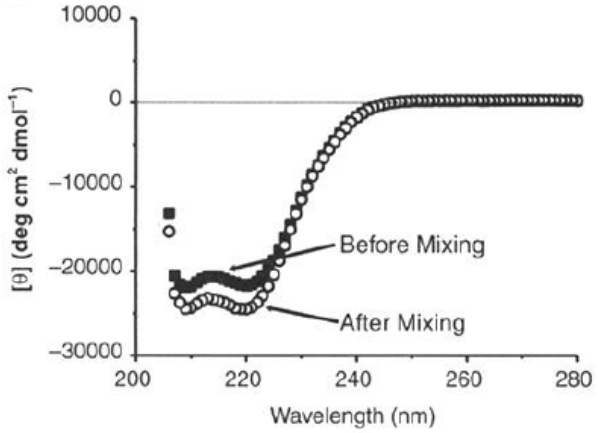
Malashkevich et al, PNAS 95, 9134 (1998)

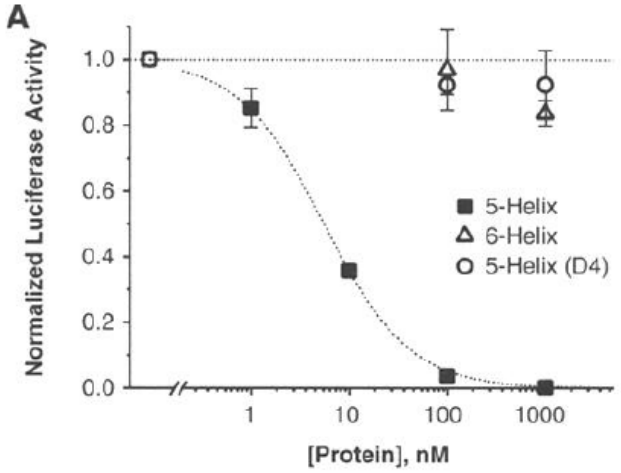


GGSGG

GSSGG

infectivity assay





Root et al, Science 291, 884 (2001)

Anticancer helix

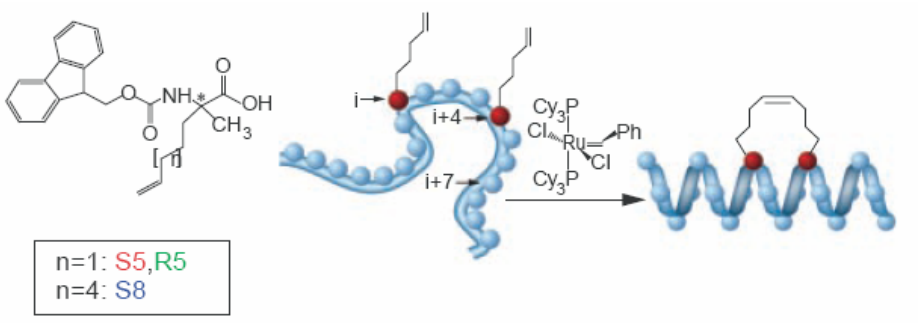
Bcl-2 regulates apoptosis in B-cells through conserved alpha helical domain (BH domain or “death” domain)

anti-apoptotic: Bcl-2, Bcl-X_L

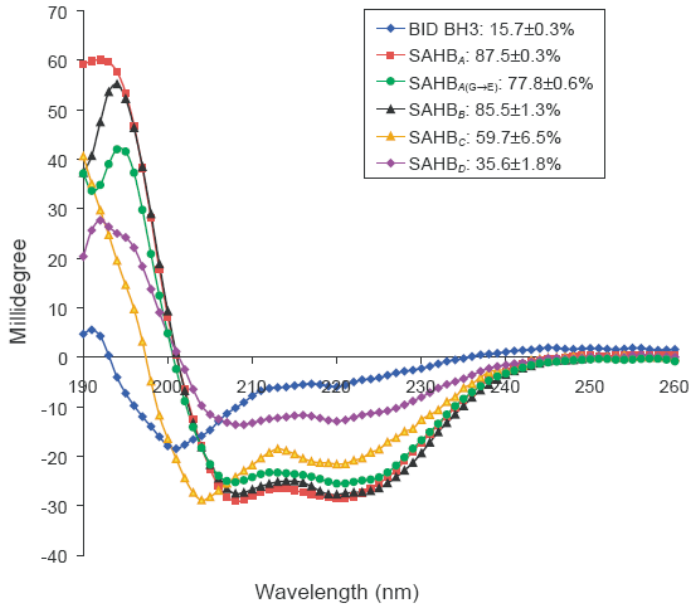
pro-apoptotic: BAX, BAK, BID, BAD

BH helix from BID can sequester anti-apoptotic Bcl-2 proteins, but is unstable as isolated helix

Hydrocarbon “staple” across the helix turn can stabilize the helix

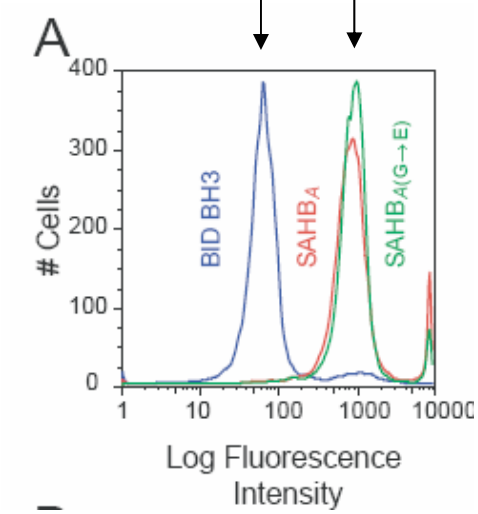


Walensky et al, Science 305, 1466 (2004)

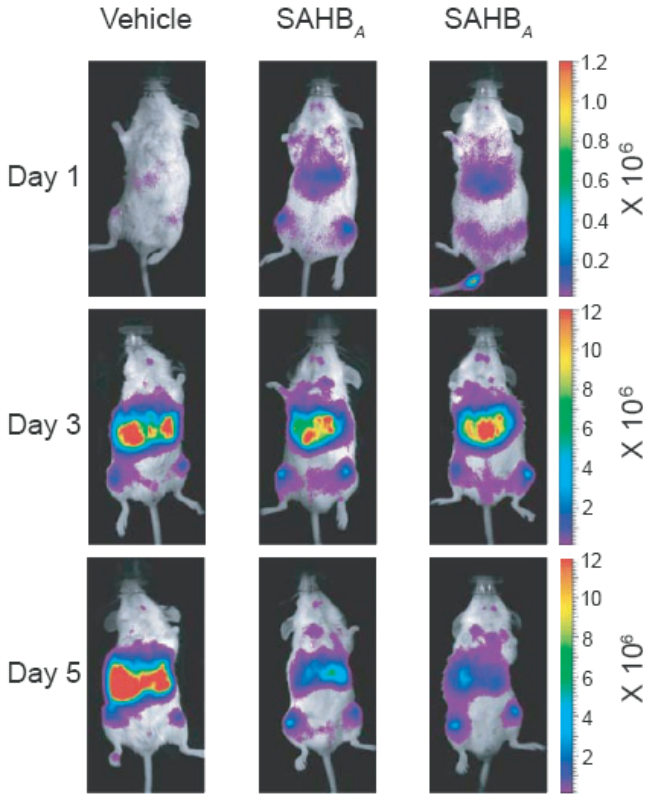


live cells

dead cells



in vitro assay



red = leukemic cells

Beta sheet design

The factors that modulate the formation and stability of beta sheet protein are not well understood

- both local and global effects are important
- stabilizing factors become apparent only in large structures
- long distance interactions within a beta sheet are difficult to engineer systematically
- contrasts with alpha helix

Practical concerns

- designed beta hairpins and beta sheets tend to have low solubility and easily aggregate—in part due to exposed main chain polar atoms
- amino acids of high beta sheet propensity are in general hydrophobic, e.g. Val, Ile, Phe, Tyr

Hierarchic design of beta sheet protein

Can an iterative hierarchic approach be applied to design a soluble monomeric beta sheet protein so that it folds without a metal or a disulfide bond?

- small proteins containing disulfide bonds are classified separately in SCOP

Information available that may be used during hierarchical design

- experimental data on beta hairpin stability
- amino acid beta sheet propensities
- statistical preferences for interstrand residue pairs
  - Wouters and Curmi, Proteins 22, 119 (1995)
- side chain rotamer modeling

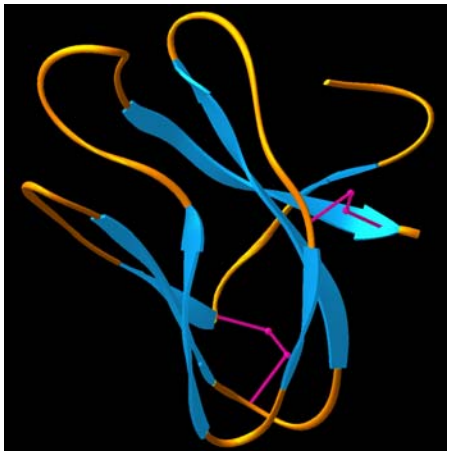
TABLE IV. High Scoring Pairs*

H-bonded		Non-H-bonded	
Cys-Cys	4.9	Cys-Cys	9.9
Glu-Lys	3.4	Glu-Lys	3.2
Glu-Arg	3.4	Asp-His	3.0
Gln-Arg	2.5	Ser-Asn	2.1
Phe-Phe	2.4	Thr-Thr	2.0
Ser-Ser	2.2		
Asp-Lys	2.1		
Gln-Lys	2.1		
Thr-Asn	2.0		

Beta hairpin peptide

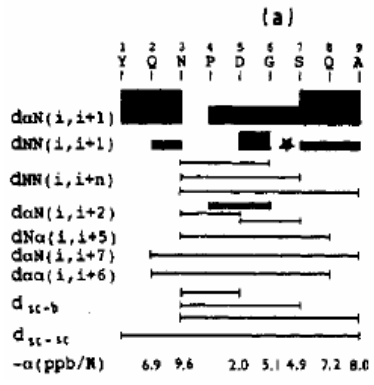
Beta turns form early during protein folding and stabilize chain reversal

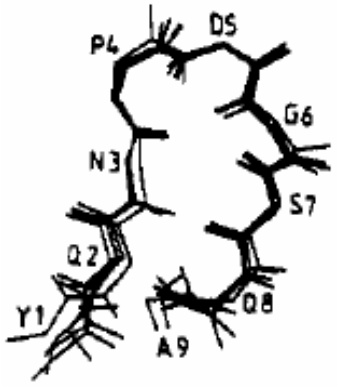
Early designs of non helical structures included beta hairpins containing 8-16 residues



tendamistat

YQSWRYSQA (wt)
YQNPDGSQA (designed)
GHNPDGHG (neg control)

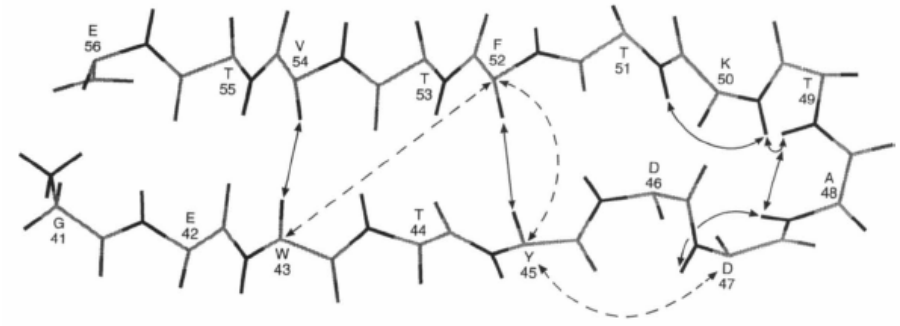




long distance NOE provide the most conclusive evidence

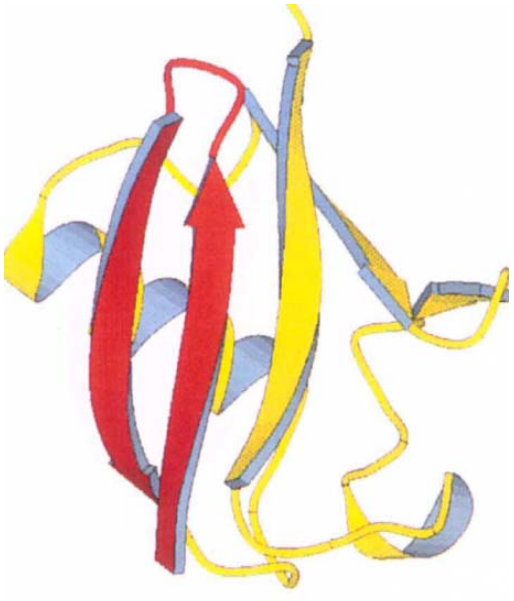
Blanco et al, JACS 115, 5887 (1993)

Some peptides isolated from natural proteins containing beta hairpin form beta hairpins in vitro, while others do not



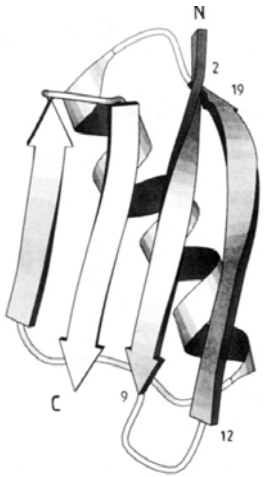
Protein G, B1 domain 41-56

Blanco et al, NSB 1, 584 (1994)



ubiquitin residues 1-17 require optimization of the turn sequence TLTGK ➔ **NPDG**

Searle et al, NSB 2, 999 (1995)



But not the first hairpin of GB1

Blanco et al, Biochem 33, 6004 (1994)

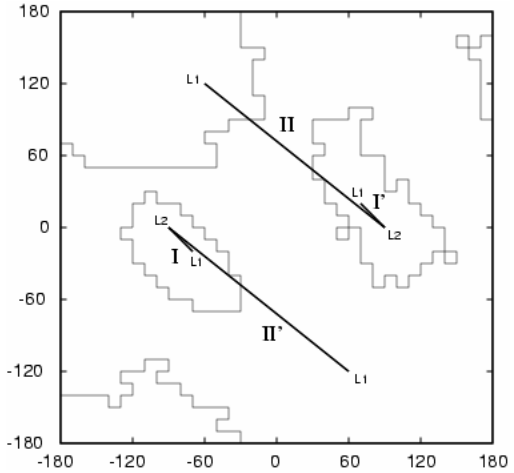
Loop residue chirality

D-amino acids intrinsically prefer type I' and II' turns (why would this be?)

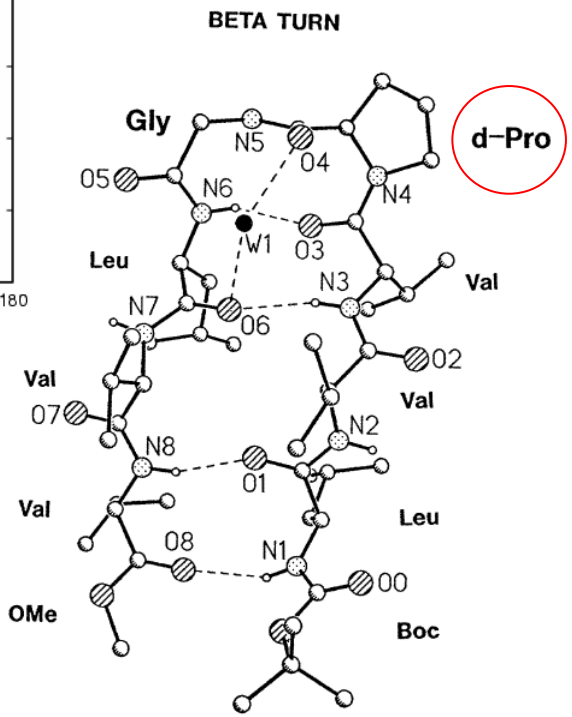
MQIFVKSXXKTITLVKV

- **L-P-L-A (x)**
- **D-P-D-A (√)**
- **D-P-L-A (√)**
- **D-P-G (√)**
- **L-P-G (x)**

Haque & Gellman , JACS 119, 2303 (1997)



type II'



RYVEVXGOKILQ

- **D-P (√√) > L-N (√)**
- **L-P (x)**
- O = ornithine

Karle et al, PNAS 93, 8189 (1996)

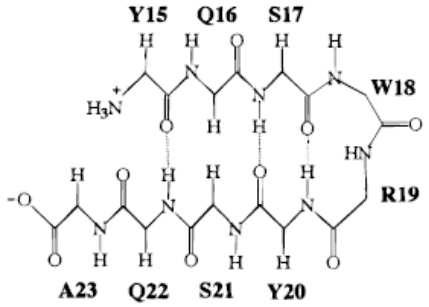
Stanger & Gellman , JACS 120, 4236 (1998)

Turn residues are key

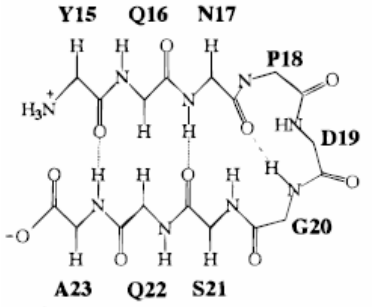
Turn conformation depends on the sequence of the loop

Interactions involving side chain of residues before and after the turn (inter-strand interactions) can further influence the turn stability and geometry

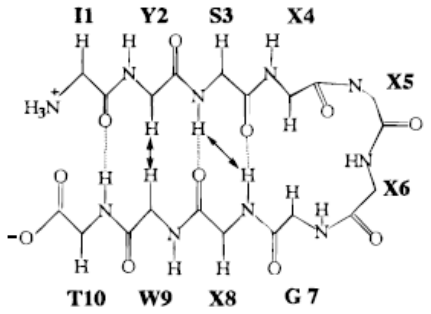
Peptide	X4	X5	X6		X8
2	N	P	D	G	T
3	N	S	D	G	T
4	N	S	D	G	V
5	A	P	D	G	T
6	A	K	A	G	T



a) β-hairpin 2:2, type I β-turn



b) β-hairpin 3:5



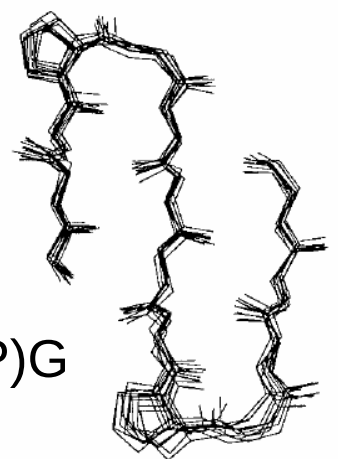
c) β-hairpin 4:4

de Alba et al, JACS 119, 175 (1997)

Three stranded sheet

(Slightly longer) peptides with two turn sequences

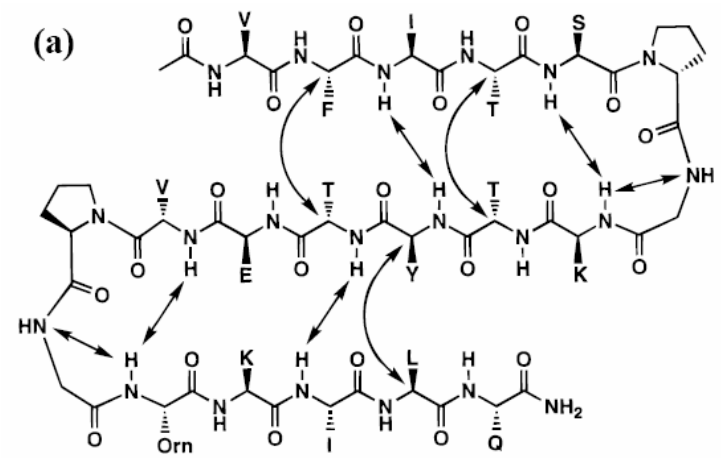
- optimization of folding and solubility (often requires non-aqueous solvent)
- not as easy as it seems
- works better with D-amino acids



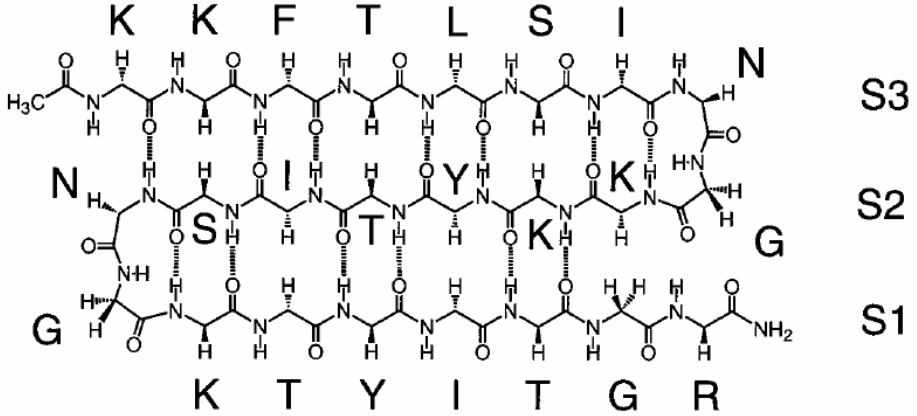
LFV(^{D}P)GLVLA(^{D}P)G
FVL
in chloroform/DMSO

Das et al, JACS 120, 5812 (1998)

Ac-VFITS(^{D}P)GKTYTEV(^{D}P)GOKILQ-NH2



Schenck and Gellman, JACS 120, 4869 (1998)



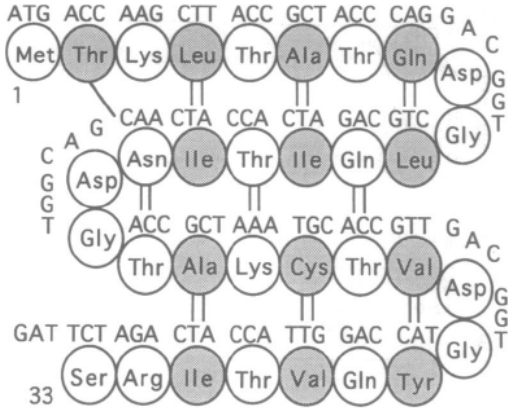
Aqueous methanol
Sharman and Searle, JACS 120, 5291 (1998)

Betadoublet

Beta sandwich topology is common in nature and include Ig, Con A, etc

Design and characterize a beta sandwich protein de novo

- Construct backbone using poly-ala (phi = -139°, psi = 139°), twist the sheet, and further modify phi/psi angles of residues for interstrand hydrogen bond
- Type I' turns throughout—XDGX
- Sheets oriented at 30° of each other
- Disulfide bond between cys
- "Negative design" to disfavor Greek key—short turn sequence



Quinn et al, PNAS91, 8747 (1994)

Analysis by size exclusion chromatography

improved solubility (> 10 mg/ml), good disperson of NMR peaks, cooperative thermal melting, reasonable CD, minimal ANS binding

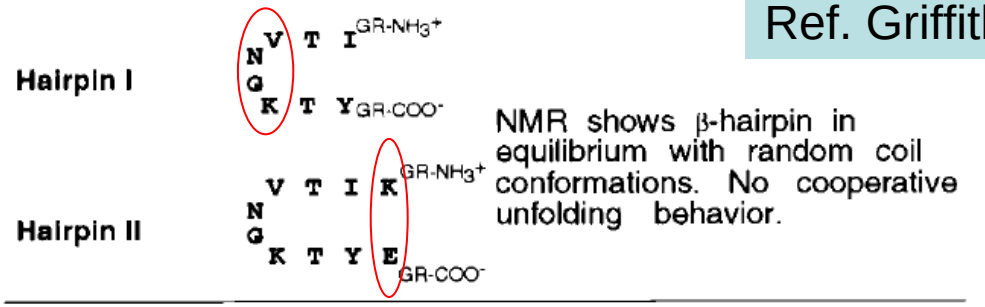
Also read about "Betabellin" by Yan & Erickson, Protein Sci 3, 1069 (1994)

Betanova

Start from a known hairpin structure and extend it long enough to constitute antiparallel beta strands

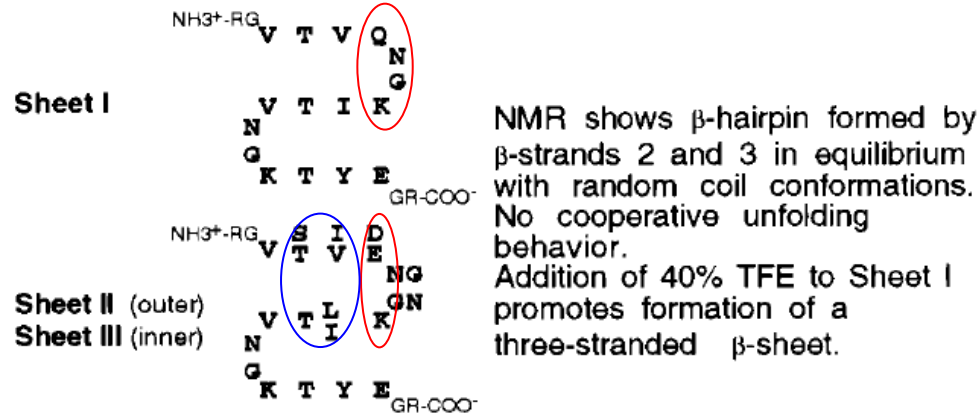
Ref. Griffiths-Jones, JMB

fold structure
35% to 45%



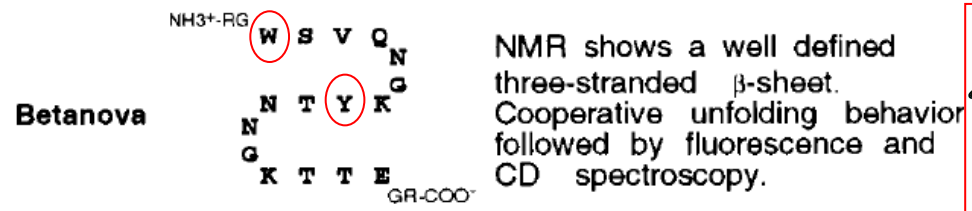
turn 1—ideal type I’

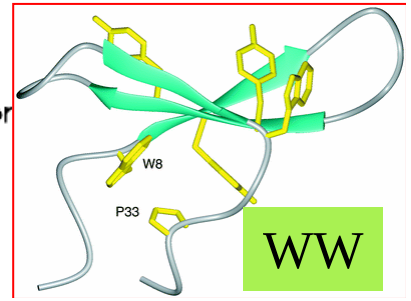
•interstrand packing
•ionic pair



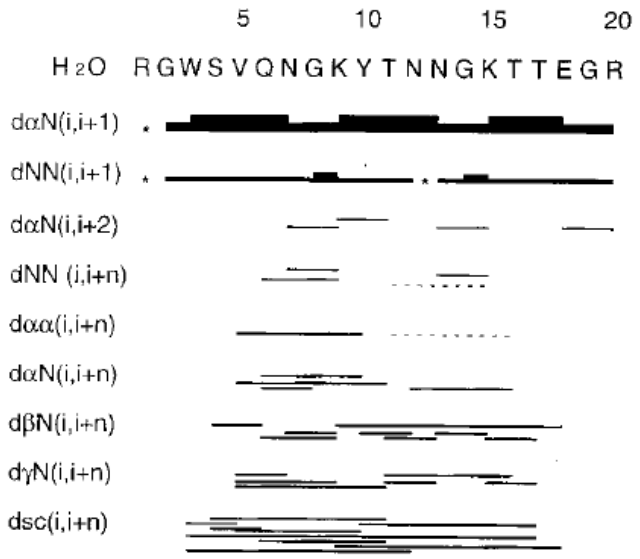
not much improvement

more hydrophobic residues in strands 1 and 2



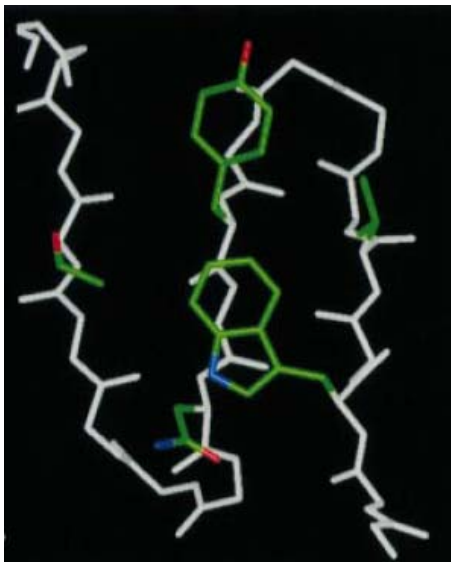


Kortemme et al, Science 281, 253 (1998)



NOE cross peak intensities

20 NMR structures

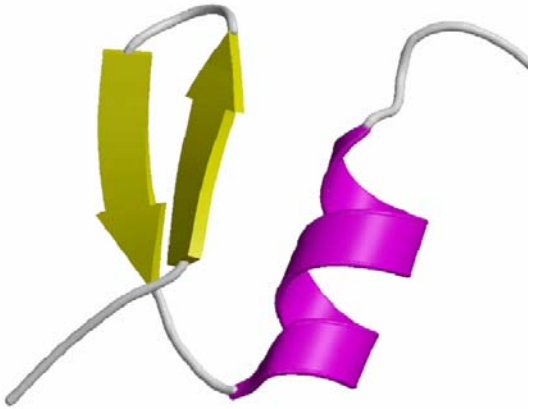


energy minimized structure

First example of a soluble three strand beta sheet protein made of natural amino acids stabilized purely through noncovalent bonds

Protein of mixed topology

Design soluble small ββα motif domain



β Hairpin → Loop → α Helix

	1							7							12											25
Zif268 1	-P-	Y-	A-	C-	P-	V-	E-	S-	C-	D-	R-	R-	F-	S-	R-	S-	D-	E-	L-	T-	R-	H-	I-	R-	I-	H-	T-	G-	Q-
Consensus	-X-	F-	X-	C-	X-	X-	X-	X-	C-	X-	X-	X-	F-	X-	X-	X-	X-	X-	L-	X-	X-	H-	X-	X-	X-	H-	X-		

- Redesign of β hairpin to incorporate type II turn, shortening loop by 4 residues
- Manipulation of ligand sphere to include reporter functionality ($Cys_2His_2 \rightarrow His_2Fen$)

β Hairpin → Loop → α Helix

	1							8				12									22	
First generation 1	Ac-	F-	A-	H-	P-	D-S	Z-	A-	F-	S-	Q-	S-	D-	E-	L-	T-	K-	A-	I-	R-	T-	H-	T-NH_2

CD: metal-dependent structure

- Stabilization of helix and sheet secondary structure: Ala→Thr in β hairpin: Thr, Ile→Ala, Leu in helix
- Incorporation of C-terminal helix cap

Second generation 2	Ac-	F-	T-	H-	P-	D-S	Z-	T-	F-	S-	R-	S-	D-	E-	L-	A-	K-	L-	L-	R-	L-	H-	A-	G-NH_2

CD: greater preorganization of secondary structure, metal induces a higher degree of structure

- Replace type II with type II′ turn: promote β-hairpin formation Pro ^{4}DSer^5→DPro ^{4}Ser^5

Third generation 3	Ac-	F-	T-	H-	D-P	S-	Z-	T-	F-	S-	R-	S-	D-	E-	L-	A-	K-	L-	L-	R-	L-	H-	A-	G-NH_2

CD: secondary-structure metal independent
NMR: 1:1 DPro^4 *cis:trans* amide isomer ratio precluding structural analysis

- His^3→Val^3:minimize *cis* DPro^4 amide isomer
- Phe^1→Tyr^1

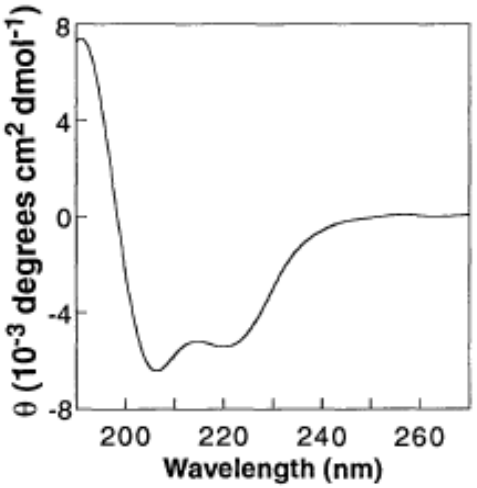
Fourth generation BBA1	Ac-	Y-	T-	V-	D-P	S-	Z-	T-	F-	S-	R-	S-	D-	E-	L-	A-	K-	L-	L-	R-	L-	H-	A-	G-NH_2

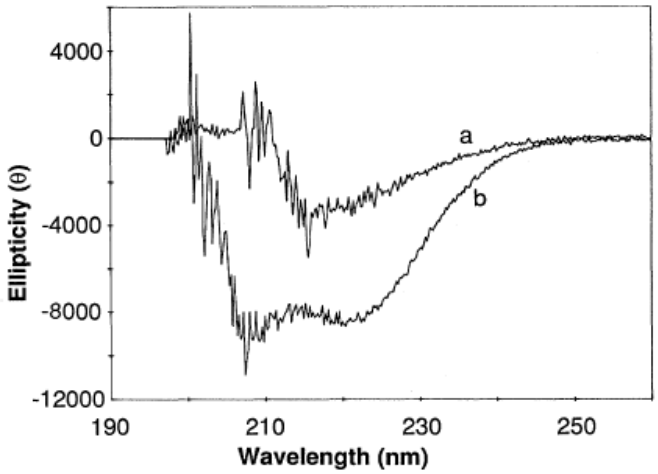
CD:metal-independent structure
NMR: cis isomer <15%; complete structure analysis; well-defined secondary and tertiary structure

Type II turn Control	Ac-	Y-	T-	V-	P-	D-S	Z-	T-	F-	S-	R-	S-	D-	E-	L-	A-	K-	L-	L-	R-	L-	H-	A-	G-NH_2

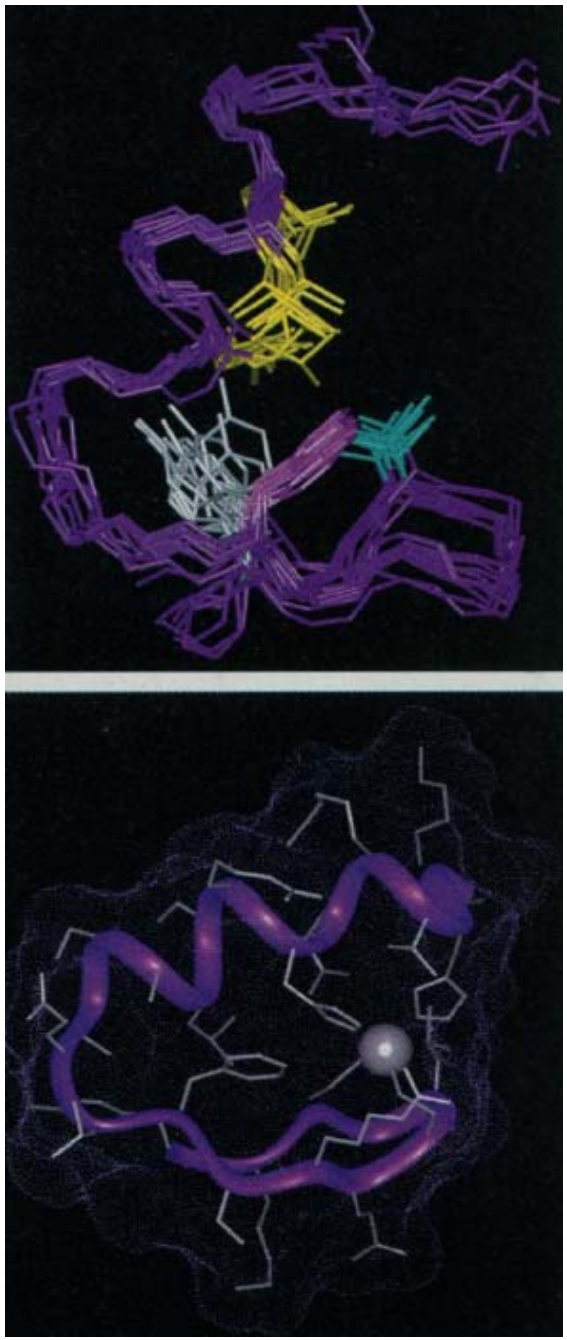
NMR: well-developed helix; complete loss of β-sheet and tertiary structure

Struthers et al, Science 271, 342 (1996)





known zinc finger CD spectra



BBA1

Zif268

State of knowledge based design

Helices and simple helical proteins (coiled coils and helical bundles) can be reliably designed

Even simple designs may have profound biomedical implications (HIV inhibitor and cancer therapeutics)

Ability to design beta sheet proteins lags about 10 years

Intuition plays a critical role in all design problems