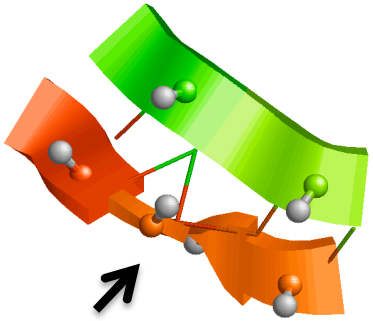
PRINCIPLE FOR DESIGNING IDEAL PROTEIN STRUCTURES

Nobuyasu & Rie Koga
University of Washington

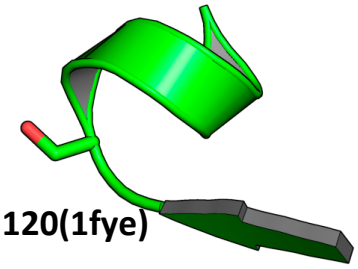
Naturally occurring protein structures are complicated

Functional site or junks of neutral drift ?

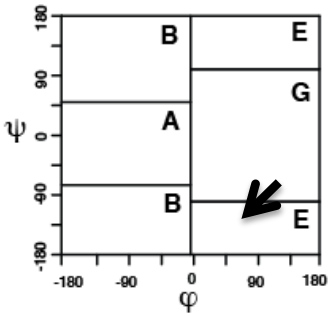
- Bulges of beta-strands



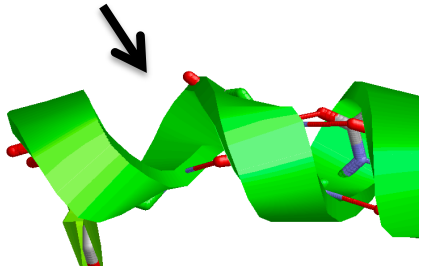
- Disallowed torsion angles in loop



SER 120(1fye)



- Kinks of helices



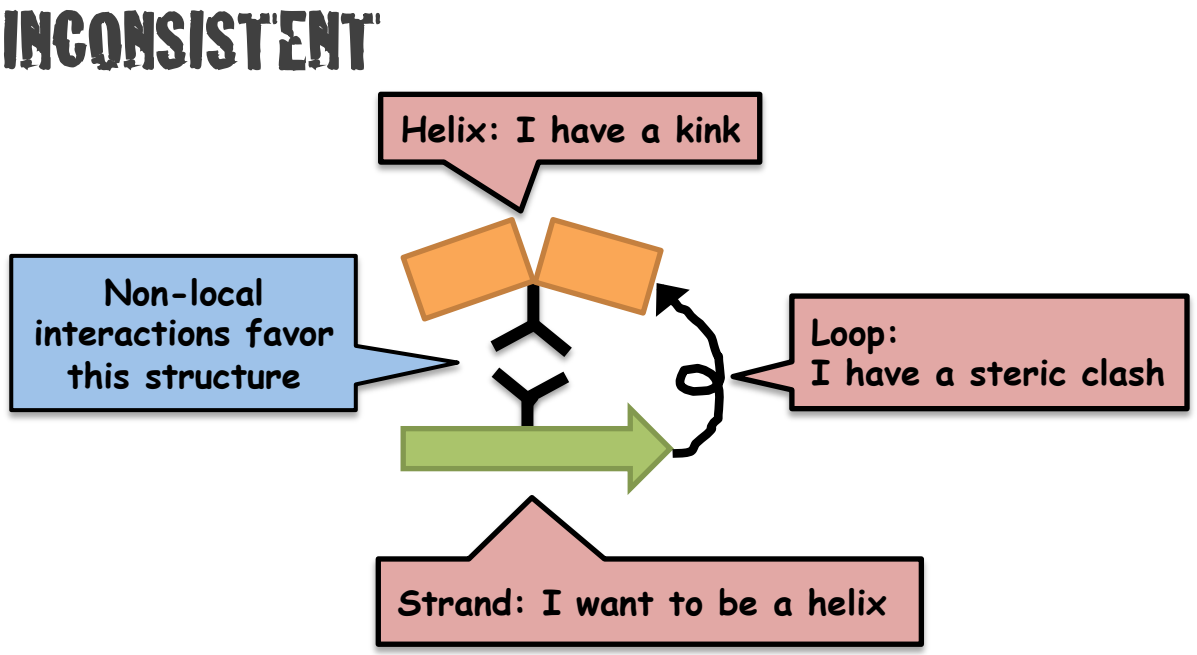
- Buried polar groups of sidechains and mainchains

Make difficult to understand the principle for protein folding.

=> Design simple and ideal protein structures

Consistency principle Nobuhiro Go 1983

Local and non-local interactions consistently stabilize native state



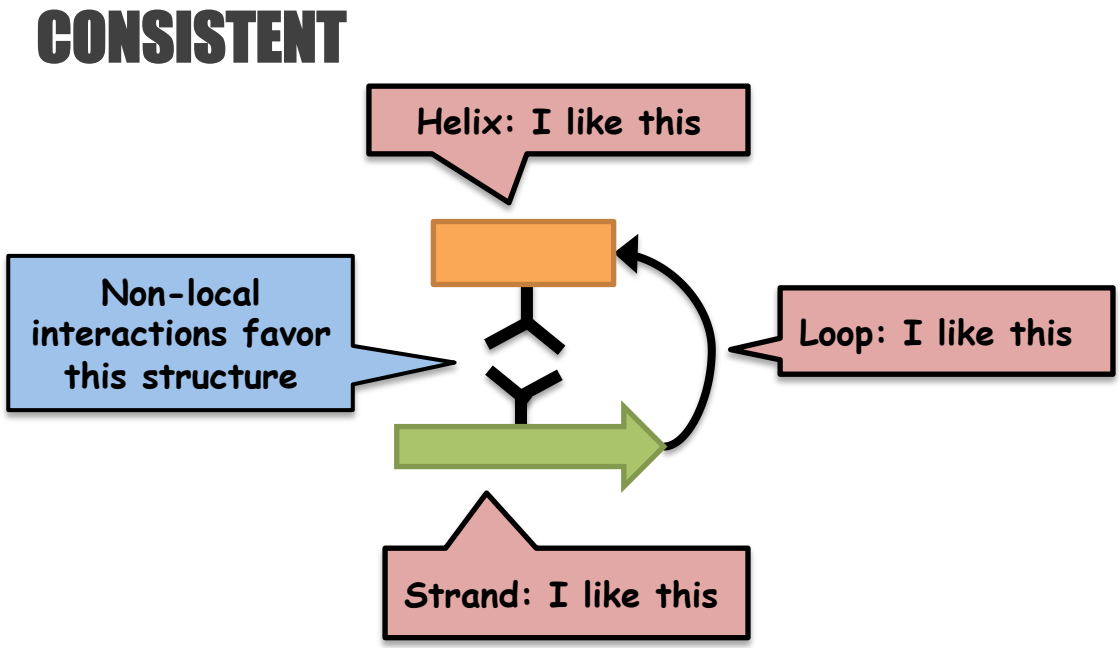
Natural (real) proteins have the consistency in the first approximation.

Ideal proteins have perfect consistency of local and non-local interactions.

Design the **ideal proteins** in reality

Consistency principle Nobuhiro Go 1983

Local and non-local interactions consistently stabilize native state

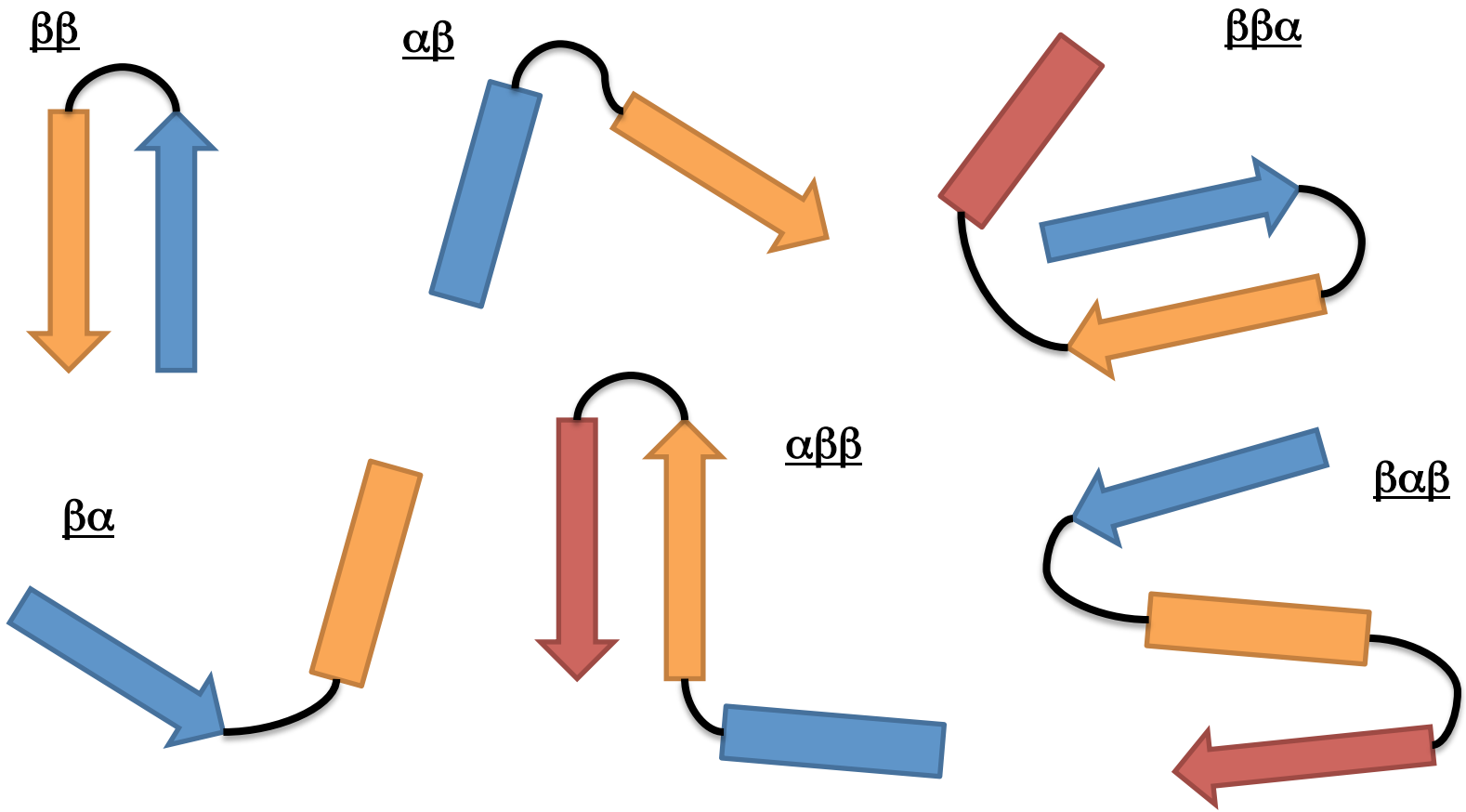


Natural (real) proteins have the consistency in the first approximation.

Ideal proteins have perfect consistency of local and non-local interactions.

Design the **ideal proteins** in reality

Mapping from Secondary structure patterns (local) to Favorable tertiary structures (non-local)

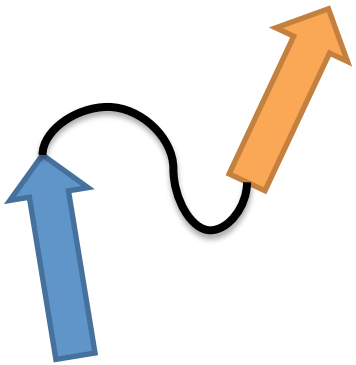


Simple rules describe

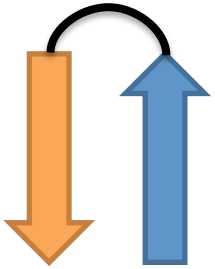
"favorable tertiary structure, depending on SS lengths"

Folding simulations with sequence-independent backbone model

Rosetta centroid



Fragment assembly with 1-mer



Secondary structure lengths

=> Conformational preference ?

Score terms:

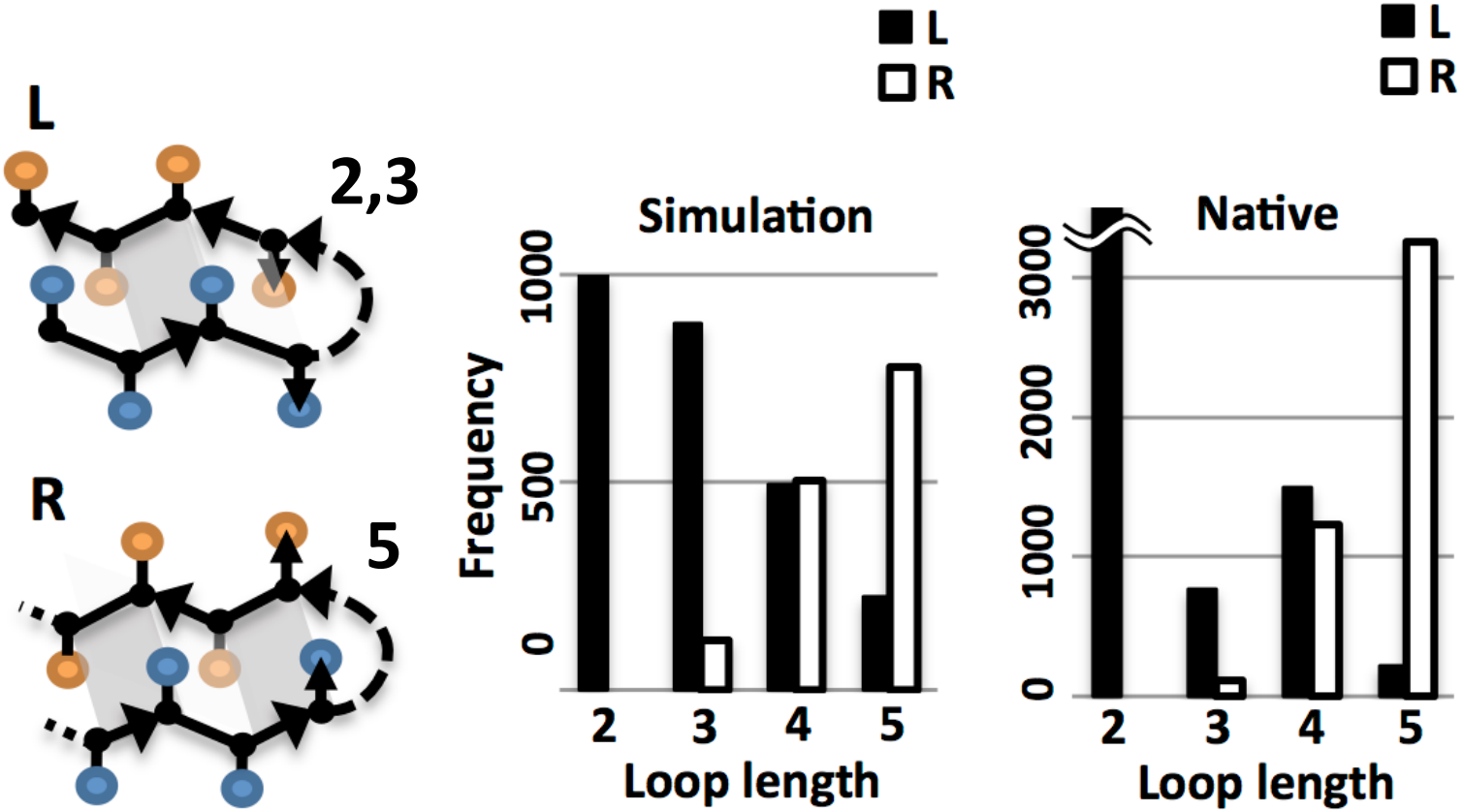
rg, vdw, sspair, rsigma, hspair

hbond_sr_bb, hbond_lr_bb

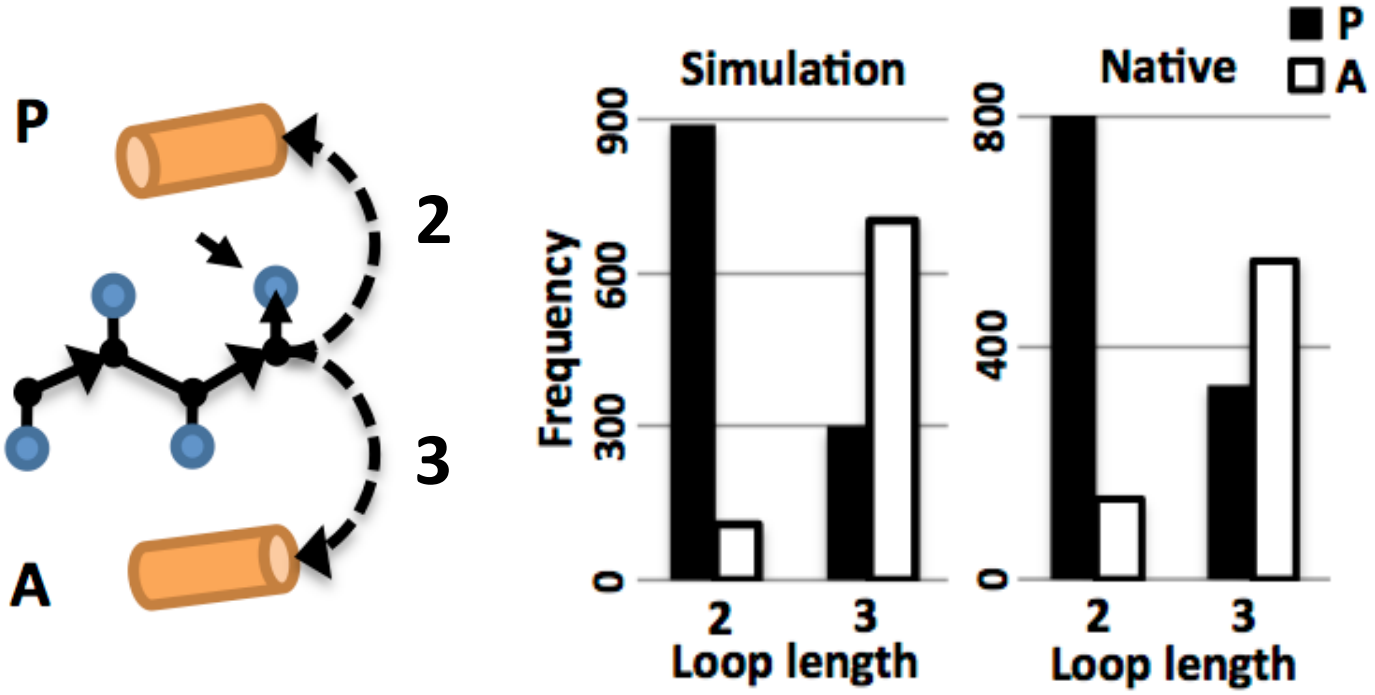
※ No amino-acid-type dependent terms

Fundamental rules: ββ, βα, αβ

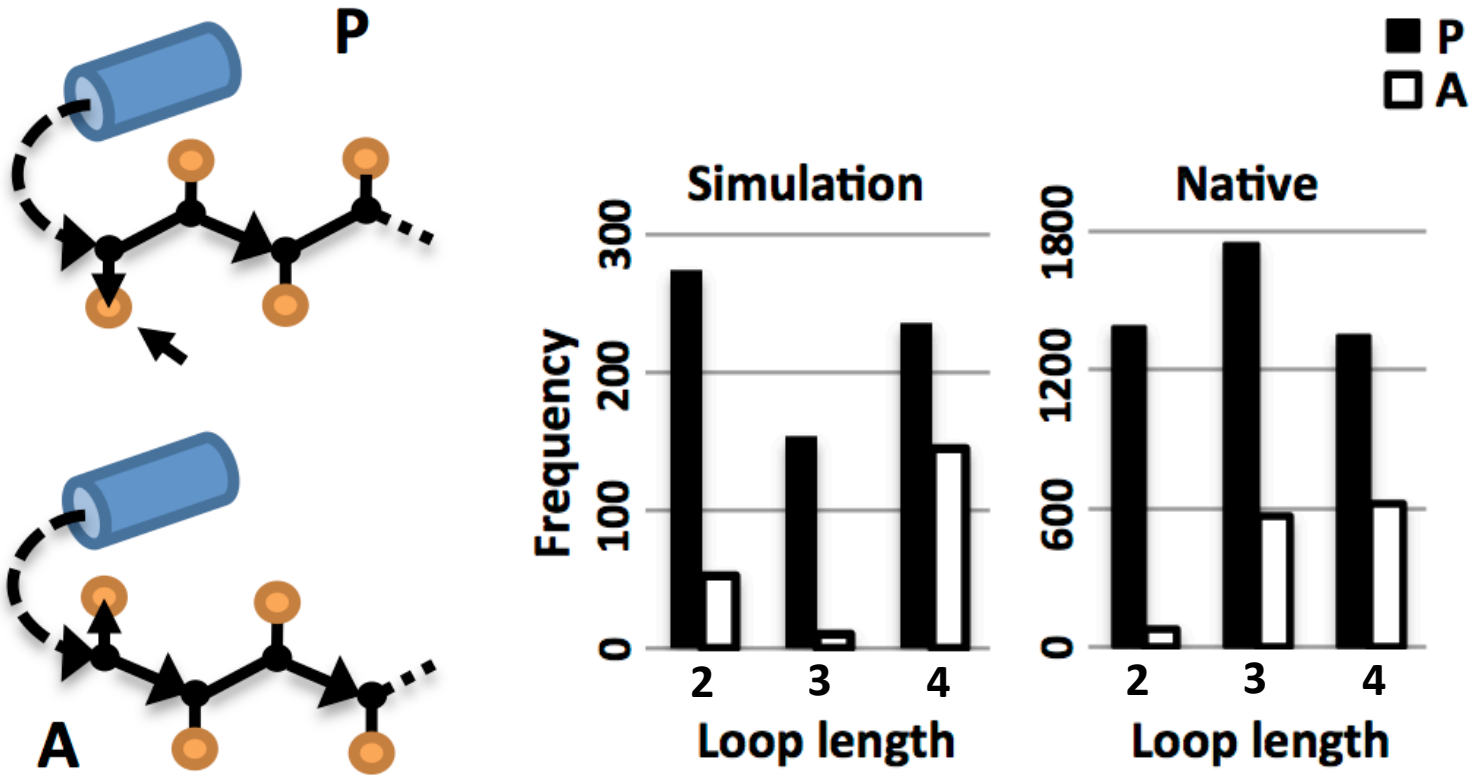
ββ-rule: Chirality of β-hairpin is determined by loop length



βα-rule: Helix direction is determined by loop length & pleat direction of last strand residue



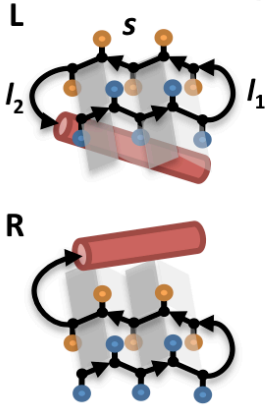
αβ-rule: Pleat of the first strand residue points away from the helix



Emergent rules

ββα-rule

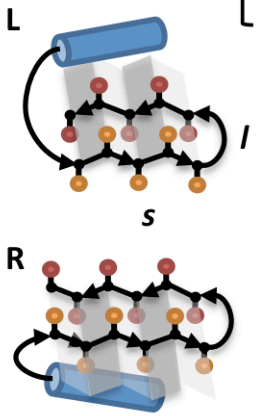
L: $s+l_2$ = odd (l_1=2), even (l_1=5)
R: $s+l_2$ = even (l_1=2), odd (l_1=5)



			l_1 = 2		l_1 = 5	
s	l_2	$s + l_2$	F_L	F_R	F_L	F_R
4	2	even	200	791	855	145
4	3	odd	796	204	181	819
5	2	odd	850	150	289	711
5	3	even	147	853	823	177
6	2	even	241	759	876	124
6	3	odd	868	132	310	690
7	2	odd	950	50	373	627
7	3	even	311	689	826	174

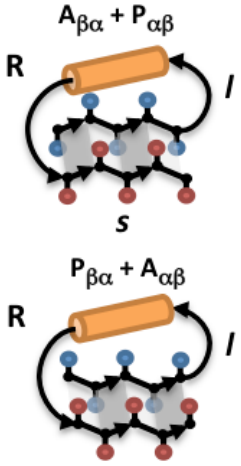
αββ-rule

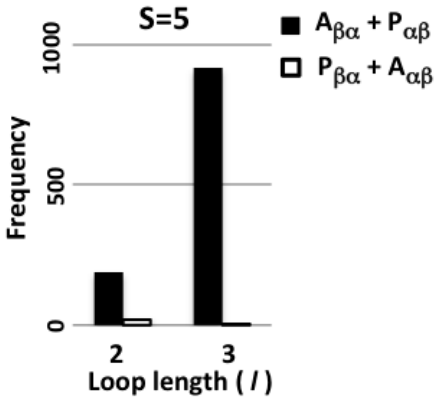
L: s = odd (l=2), even (l=5)
R: s = even (l=2), odd (l=5)



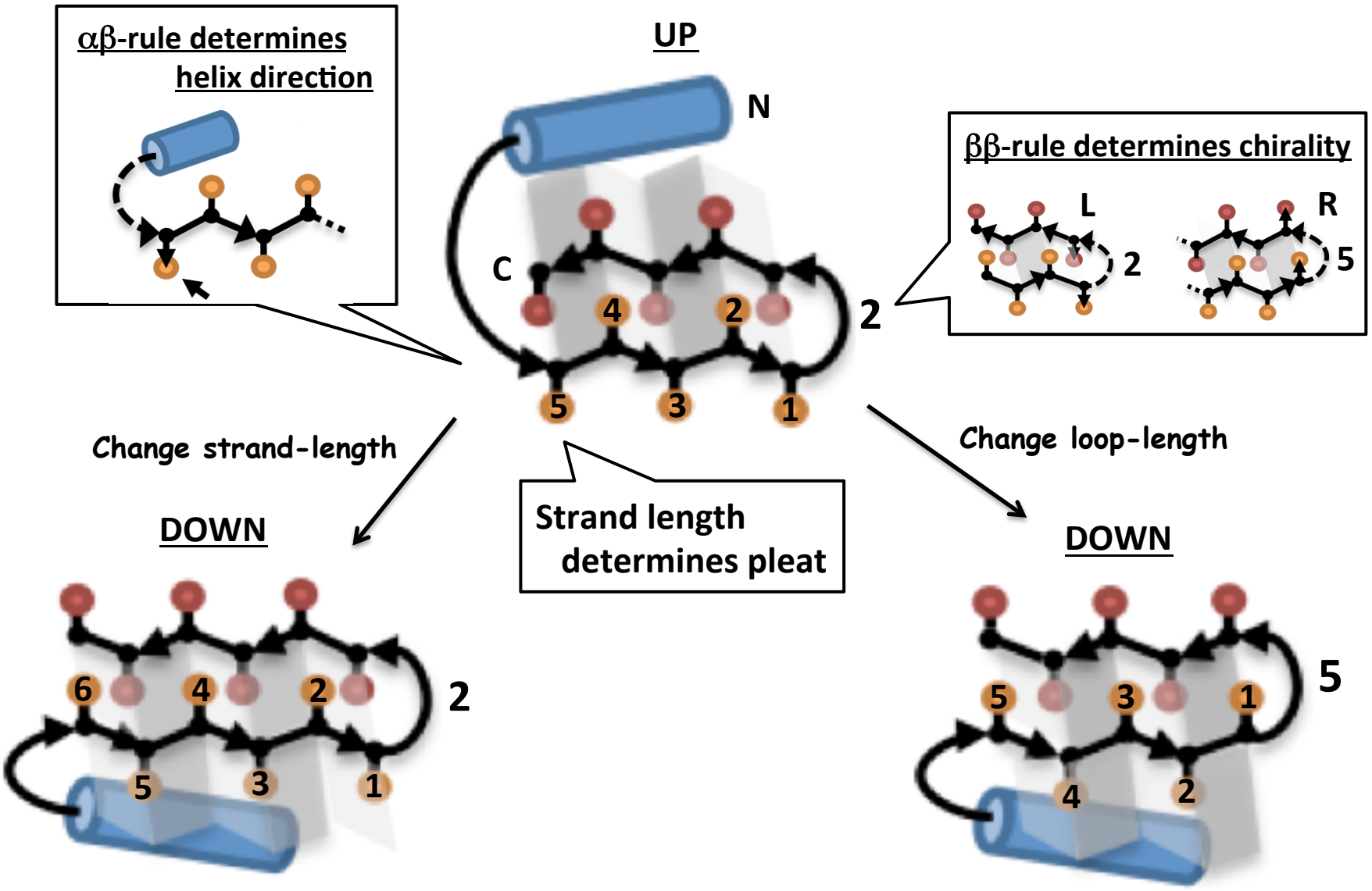
		l = 2		l = 5	
s	s	F_L	F_R	F_L	F_R
4	even	40	960	898	108
5	odd	943	57	158	842
6	even	114	886	890	110
7	odd	932	68	249	751

βαβ-rule



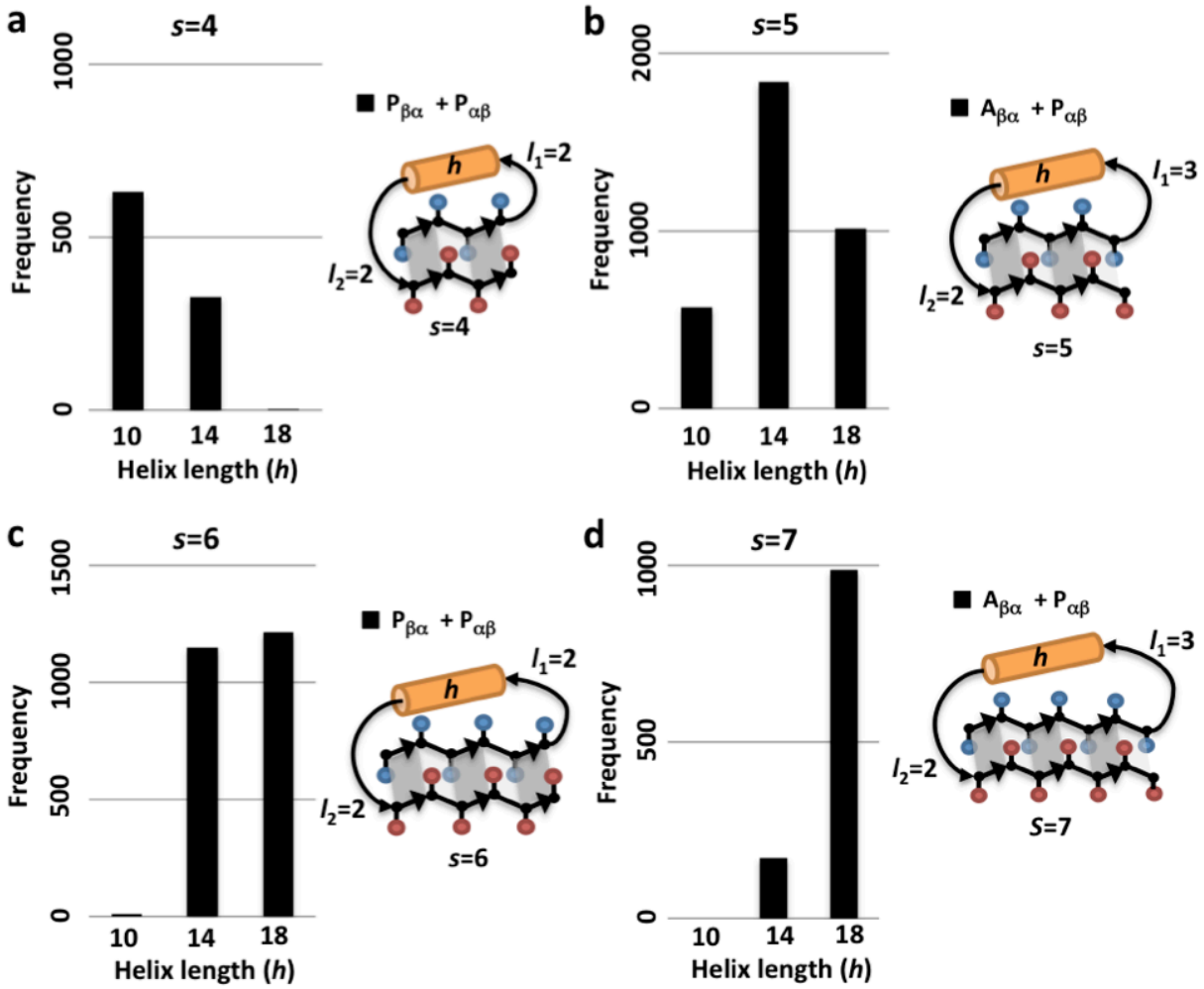


αββ-rule


αβ-rule determines helix direction
UP
N
C
4
2
5
3
1
2
ββ-rule determines chirality
L
R
2
5
Change strand-length
DOWN
6
4
2
5
3
1
2
Strand length determines pleat
Change loop-length
DOWN
5
3
1
4
2
5

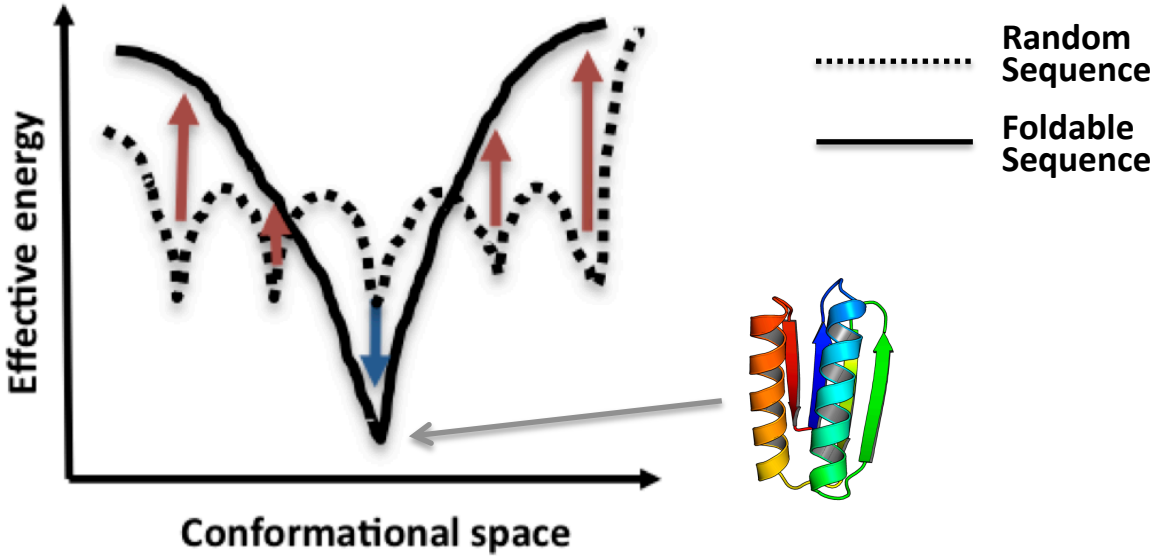
βαβ-rule

Strand and Helix lengths are codependent.



Design principle:

Consistent local & non-local interactions lead to funnel-shaped energy landscape



Non-local:

Stabilize structures by packing, hydrophobic effect, …

Local:

Choose secondary structure lengths based on the rules

Stabilize the secondary structures

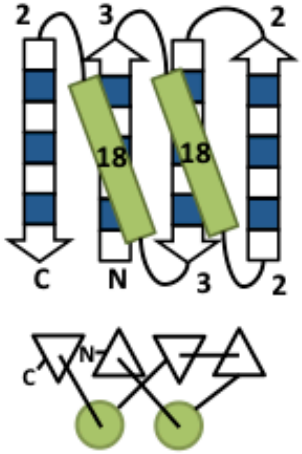
=> Eliminate non-native conformations by local backbone preference.

Let's design protein structures

Design 5 target folds

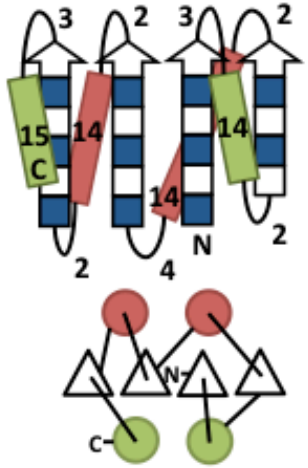
Fold-I

Ferredoxin-like



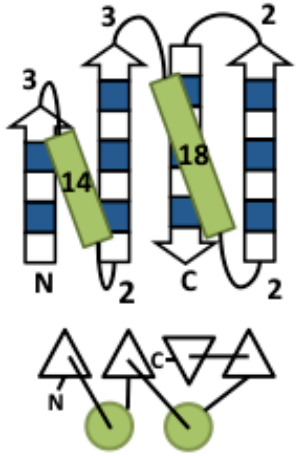
Fold-II

Rossmann2x2



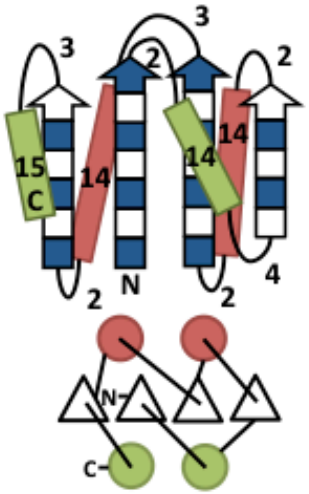
Fold-III

IF3-like



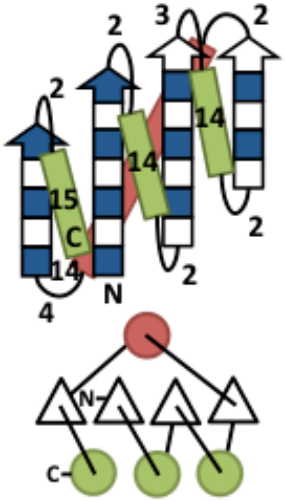
Fold-IV

Ploop2x2



Fold-V

Rossmann3x1



Design Protocol

Design Topology

Select secondary structure (SS) lengths based on the rules

Build backbones by SS dependent fragment assembly

Design Sidechains (Amino acid sequence is introduced)

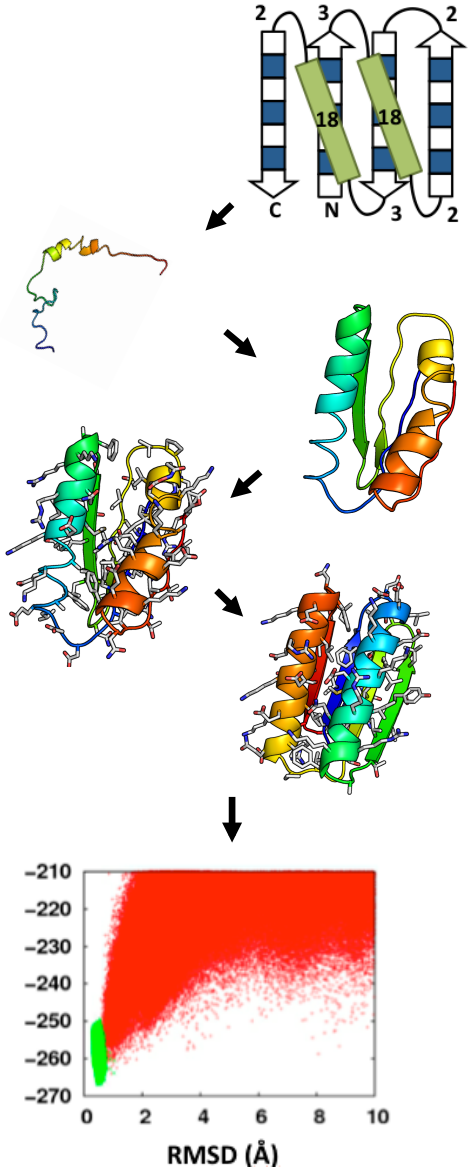
Relax both sidechains & backbone

Rosetta Energy ?
Packing (RosettaHoles) ?
Local sequence-structure compatibility ?

Explore energy landscape by structure prediction

Funnel-like ?

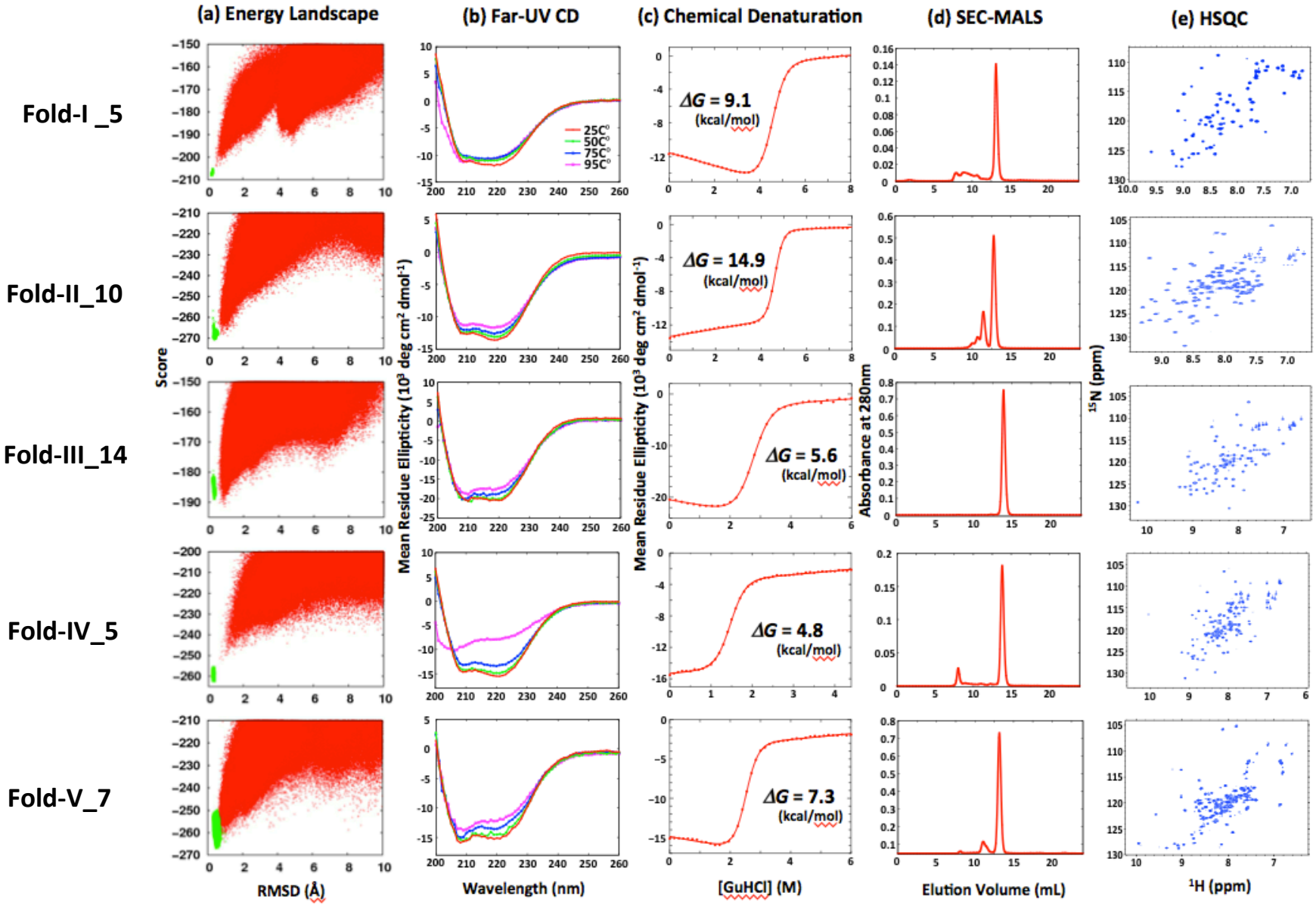
Experimental characterization (CD, SEC-MALS, NMR)



Characterization of designed proteins

	#designs tested	Expressed	Soluble	Expected CD spectrum	Stable (*Tm*>=95 C^O)	Monomeric	Well resolved NMR
Fold-I	11	9	8	6	3	2	3
Fold-II	12	12	12	10	10	4	4
Fold-III	14	13	11	9	7	6	5
Fold-IV	5	4	4	4	2	4	3
Fold-V	12	11	10	3	3	1	1

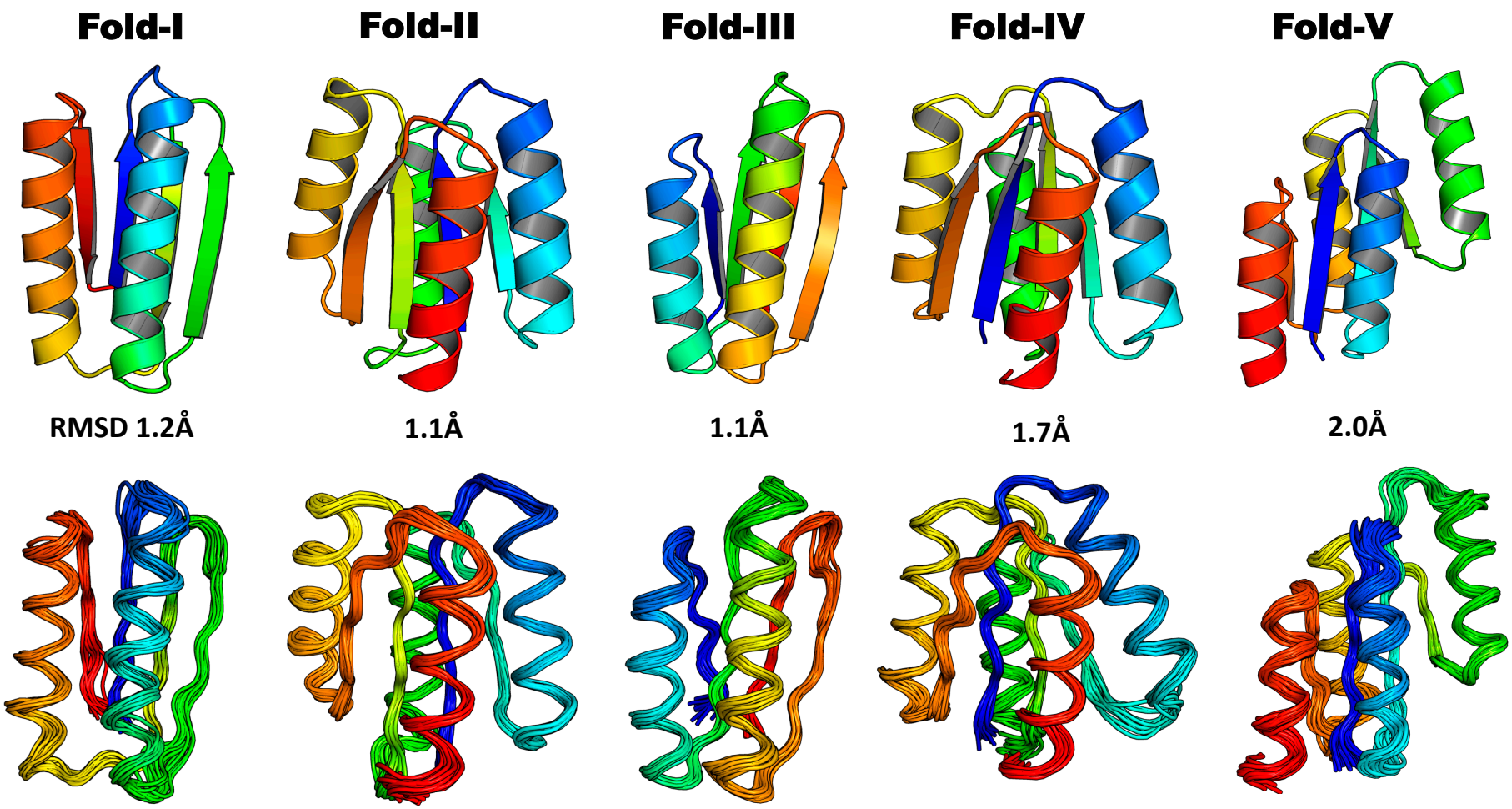
Characterization of designed proteins



Design model and NMR structures

Upper: Design model

Lower: NMR



NMR was solved by Gaohua Liu & Guy Montelione (Rutgers Univ.)

Summary of experimental results

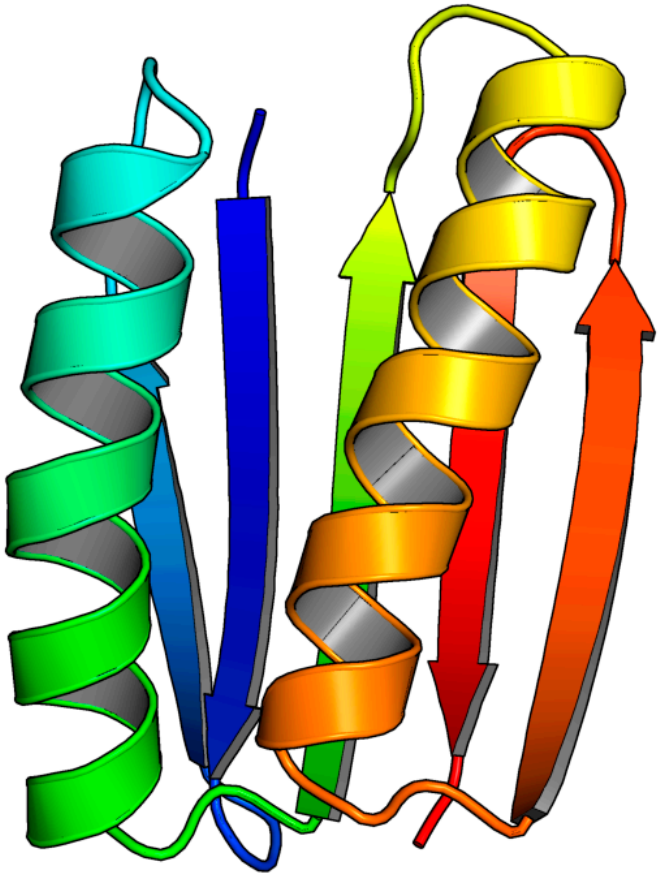
	#designs tested	Success (rate %)
Fold-I	11	1 (9)
Fold-II	12	4 (33)
Fold-III	14	3 (21)
Fold-IV	5	2 (40)
Fold-V	12	1 (8)

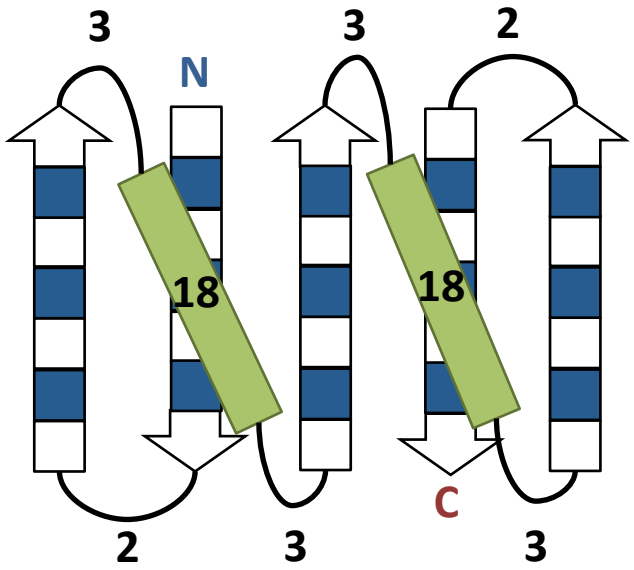
Success criteria:

- **Expressed & Soluble**
- **Expected CD spectrum**
- **Stable (Tm >=95C°)**
- **Monomeric**
- **Well resolved NMR**

Top7 is also ideal protein

Brian Kuhlman et al., Science (2003)





Secondary structure lengths
follow the rules !

Conclusion

1. We found **Rules** relating secondary structure lengths (local) to tertiary motifs (non-local)

2. The rules enabled us to design **ideal** protein structures
   stabilized by completely consistent local & non-local interactions

3. Consistent interactions readily lead to funnel-shaped energy landscapes
   Non-native conformations are disfavored by local backbone preferences

4. We designed ideal protein structures of **5** different folds based on the rules

5. The designed protein structures were monomeric, very stable (>95C), and
   adopt structure nearly identical to the computational models.

6. Natural proteins might use the rules for making funnel-shaped energy landscape

Thanks !

Structure determination

Gaohua Liu

Rong Xiao

Thomas Acton

Guy Montelione

Fold suggestion

Nick Grishin

1D-NMR for Fold-I & Fold-II

Ponni Rajagopal

Mass spec

Justin Siegel

Computational help

Possu Huang

Yih-En Andrew Ban

De novo designer

Javier Castellanos

Yu-Ru Lin

Amanda Clouser

Our neighbor

T

J505

Boss

David Baker