ARTICLE

Principles for designing ideal protein structures

Nobuyasu Koga[1]*, Rie Tatsumi–Koga[1]*, Gaohua Liu[2,3]*, Rong Xiao[2,3], Thomas B. Acton[2,3], Gaetano T. Montelione[2,3] & David Baker[1]

Presenter: She Zhang

Introduction
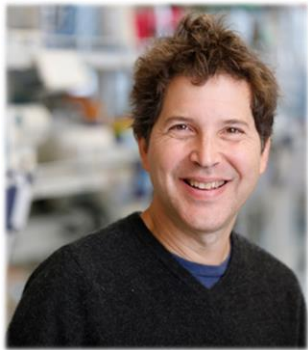
Dr. David Baker

COMMUNITY COMPUTING ALLOWS EVERYONE TO GET INVOLVED FROM HOME

Foldit is a computer game which enables you to contribute to cutting edge scientific research. Join this free online game and help us to design new proteins to cure diseases, create new materials, and develop new ways of capturing and storing energy.

Rosetta@home needs your help to determine the 3-dimensional shapes of proteins in research that may ultimately lead to finding cures for some major human diseases. By running the Rosetta program on your computer while you don't need it you will help us speed up and extend our research in ways we couldn't possibly attempt without your help. You will also be helping our efforts at designing new proteins to fight diseases such as HIV, Malaria, Cancer, and Alzheimer's (See our Disease Related Research for more information). Please join us in our efforts! *Rosetta@home is not for profit.*

Robetta: Full-chain Protein Structure Prediction

Introduction

Why design proteins *de novo*?

- It is not clear how non-covalent interactions favor one specific native structure over many other non-native structures.

- Protein design provides an opportunity to investigate the hypotheses and experimentally assessing them.

**What is the aim of this papers?**

Investigate the rules that enable us to design a funnel-shaped energy landscape for desired protein:

- Stabilizing the native state - **positive design**

- Destabilizing non-native state - **negative design**
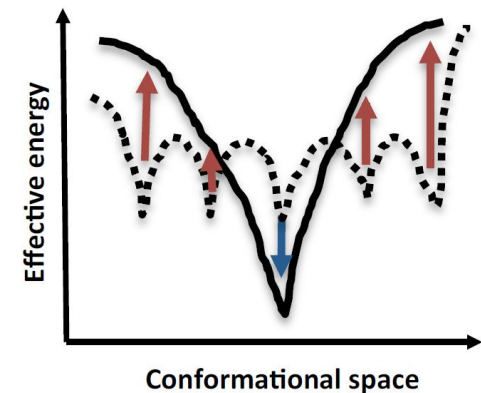
How the lengths of secondary structures – α-helix, β-strand, and random coils – contribute to the protein folding problem.

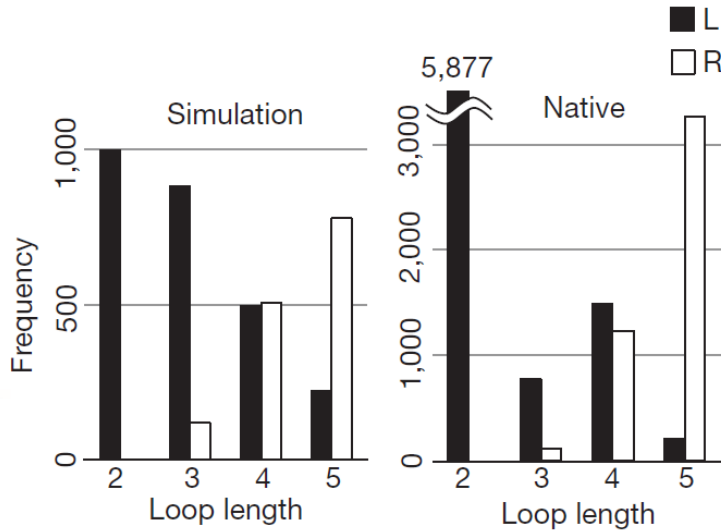# Secondary structure rules : definition of ββ-junction

L

2,3

R

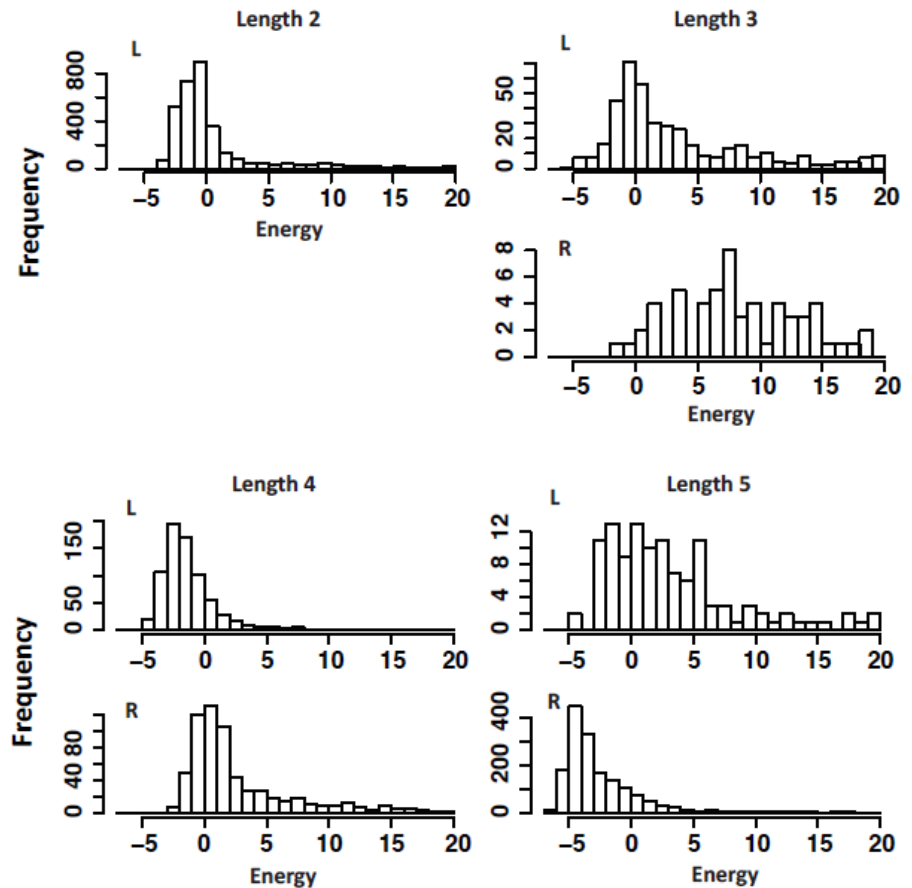5

$u \times v$

Cβ

Cα

v

u
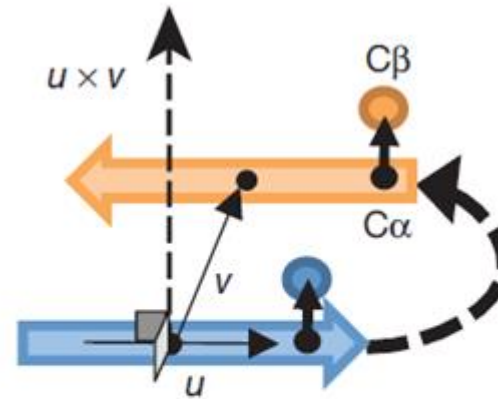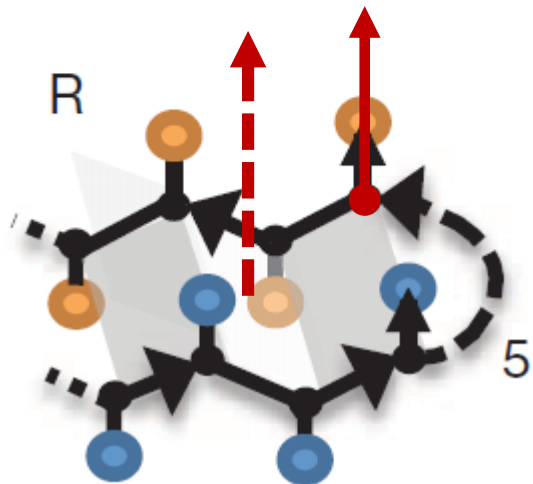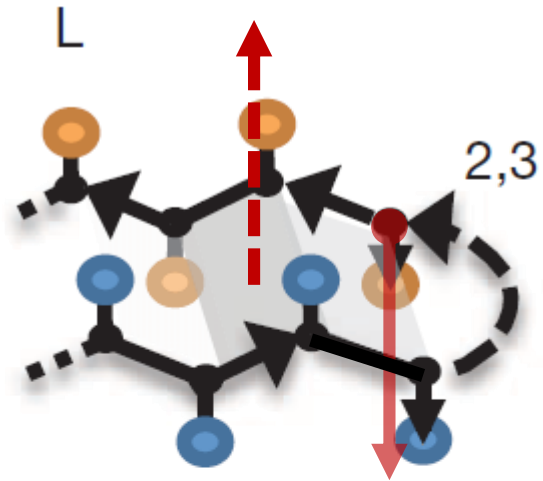
R if $(u \times v) \cdot \overrightarrow{C\alpha C\beta} > 0$

L if $(u \times v) \cdot \overrightarrow{C\alpha C\beta} < 0$
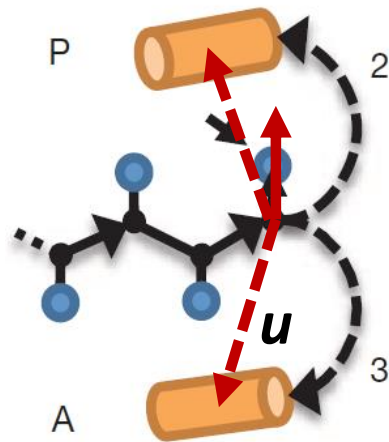
Secondary structure rules : ββ-rule

The simulated frequency and natural abundance of L- and R- conformers over different loop length.

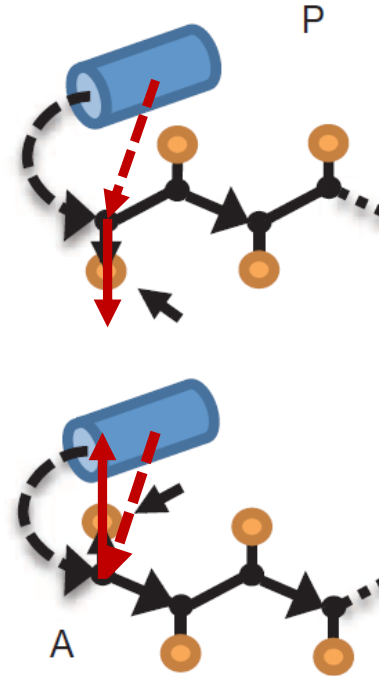The distribution of calculated torsion energies of connecting loops.

Secondary structure rules : definition of βα- and αβ- rule

βα- junction

αβ- junction

Define vector u to be from the last Cα in the β strand to the average coordinate of first 11 backbone atoms in α helix for βα, or the reverse for αβ.

P if $\langle u, \overrightarrow{C_\alpha C_\beta} \rangle \leq 80°$

A if $\langle u, \overrightarrow{C_\alpha C_\beta} \rangle \geq 100°$

Secondary structure rules : βα- and αβ- rule
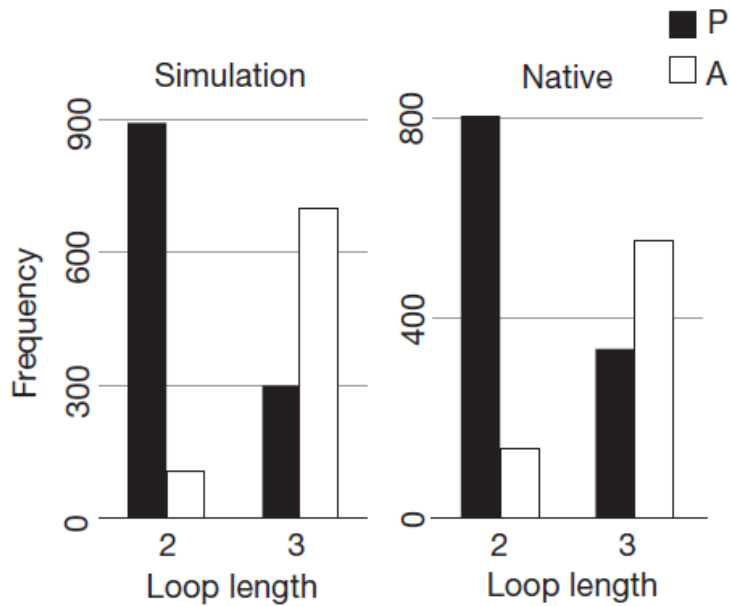
βα-junction

αβ-junction

The simulated frequency and natural abundance of L- and R- conformer.

Secondary structure rules : emergent rules

Emergent rules

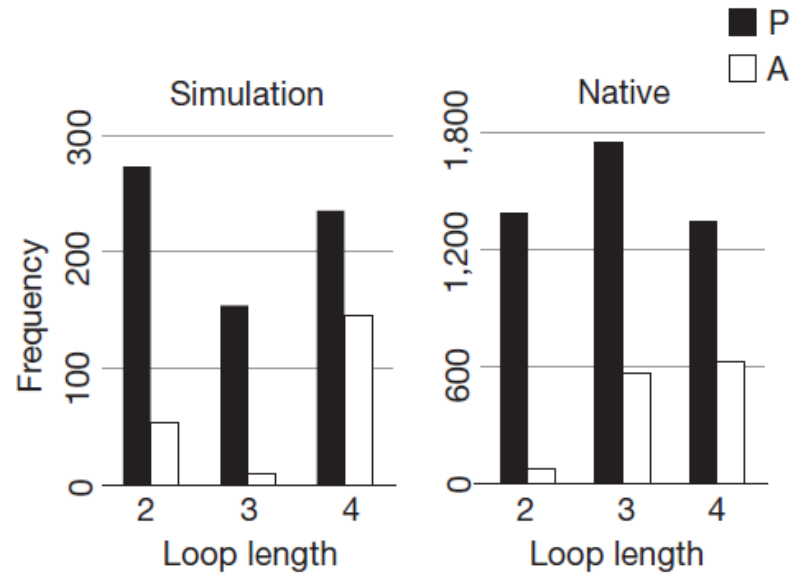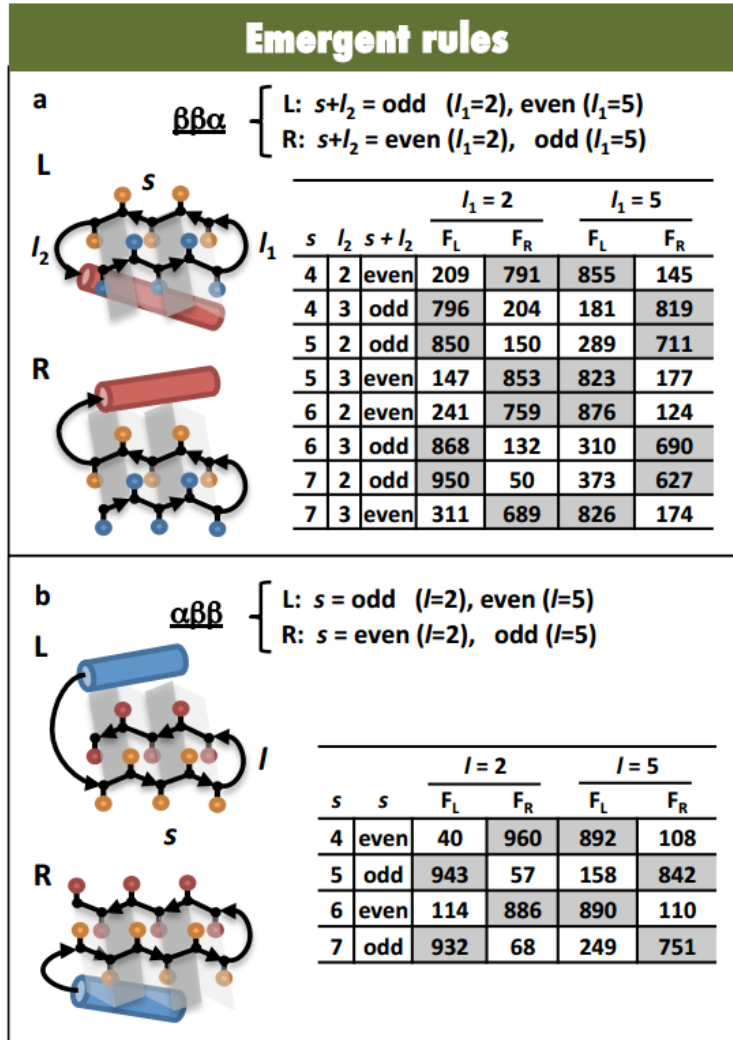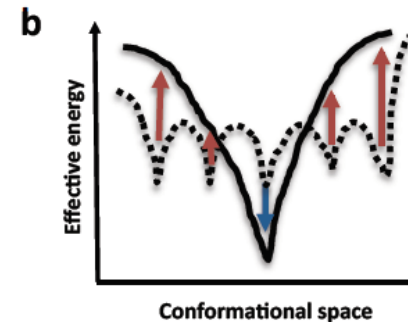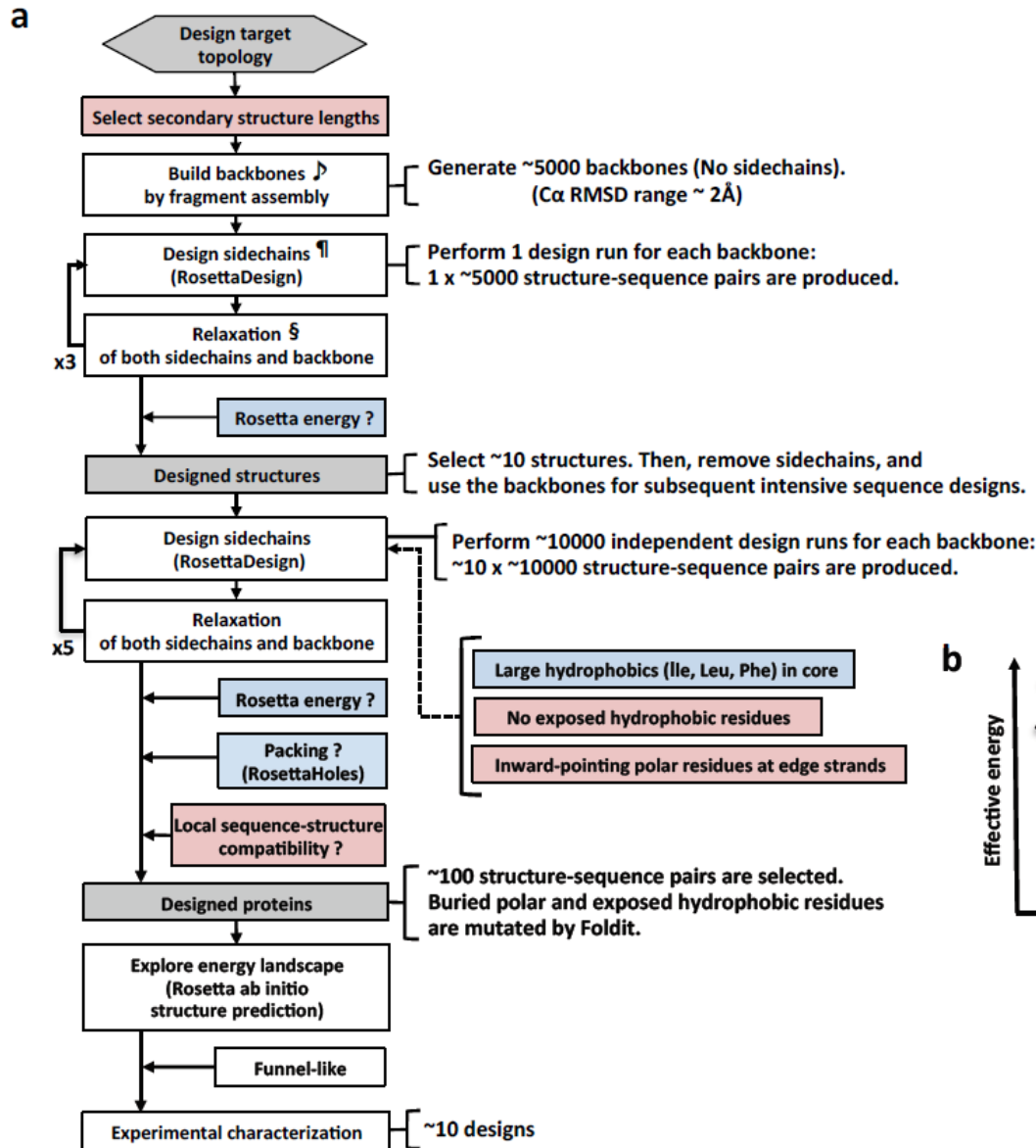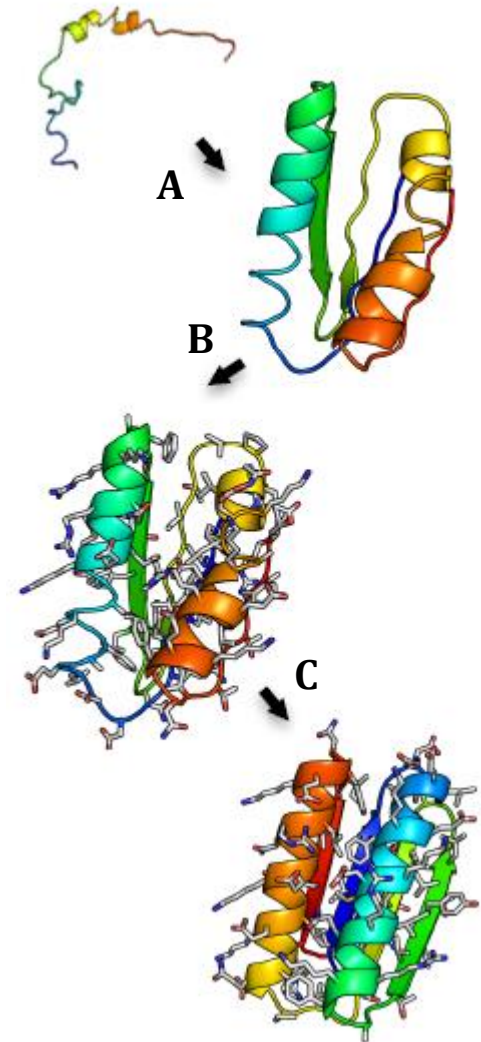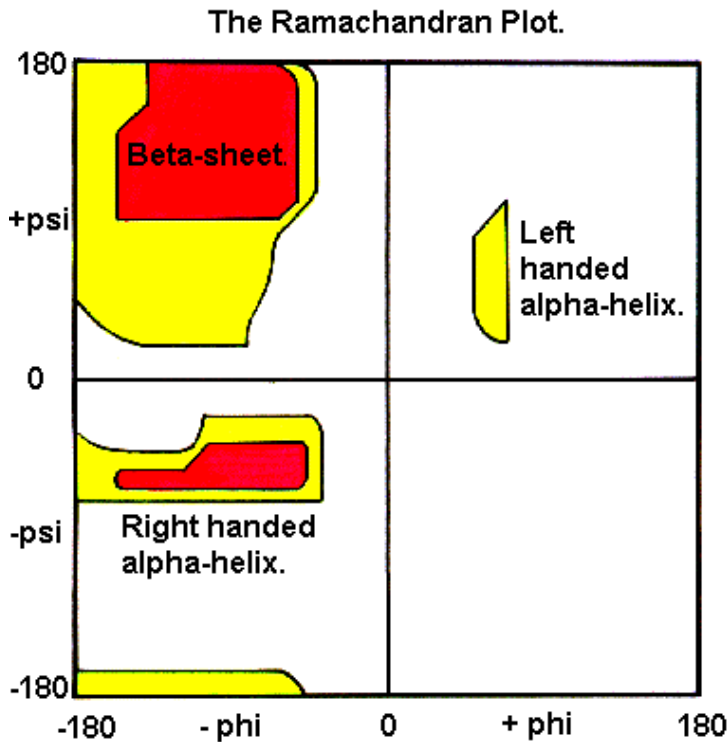**a** $\underline{\beta\beta\alpha}$ $\left\{\begin{array}{l} L: s+l_2 = odd\ (l_1=2),\ even\ (l_1=5) \\ R: s+l_2 = even\ (l_1=2),\ odd\ (l_1=5) \end{array}\right.$

			$l_1 = 2$		$l_1 = 5$	
s	l_2	$s+l_2$	F_L	F_R	F_L	F_R
4	2	even	209	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

b $\underline{\alpha\beta\beta}$ $\left\{\begin{array}{l} L: s = odd\ (l=2),\ even\ (l=5) \\ R: s = even\ (l=2),\ odd\ (l=5) \end{array}\right.$

		$l = 2$		$l = 5$	
s	s	F_L	F_R	F_L	F_R
4	even	40	960	892	108
5	odd	943	57	158	842
6	even	114	886	890	110
7	odd	932	68	249	751

c $A_{\beta\alpha} + P_{\alpha\beta}$ $\underline{\beta\alpha\beta}$ $\left\{\ \beta\alpha + \alpha\beta\right.$

$P_{\beta\alpha} + A_{\alpha\beta}$

$S = 5$

- ■ $A_{\beta\alpha} + P_{\alpha\beta}$
- □ $P_{\beta\alpha} + A_{\alpha\beta}$

Frequency — Loop length (l)

d $\vec{V}_1 \times \vec{V}_{21}$ SS_3 $\vec{V}_{31}$ SS_2 $\vec{V}_{21}$

R if $(\vec{V}_1 \times \vec{V}_{21}) \bullet \vec{V}_{31} > 0$
L if $(\vec{V}_1 \times \vec{V}_{21}) \bullet \vec{V}_{31} < 0$

SS_1 $\vec{V}_1$ N C

Protein design pipeline

a

Design target topology

↓

Select secondary structure lengths

↓

Build backbones ♪ by fragment assembly — Generate ~5000 backbones (No sidechains). (Cα RMSD range ~ 2Å)

↓

Design sidechains ¶ (RosettaDesign) — Perform 1 design run for each backbone: 1 x ~5000 structure-sequence pairs are produced.

↓

Relaxation § of both sidechains and backbone

x3

↓

Rosetta energy ?

↓

Designed structures — Select ~10 structures. Then, remove sidechains, and use the backbones for subsequent intensive sequence designs.

↓

Design sidechains (RosettaDesign) — Perform ~10000 independent design runs for each backbone: ~10 x ~10000 structure-sequence pairs are produced.

↓

Relaxation of both sidechains and backbone

x5

↓

Rosetta energy ?

Large hydrophobics (Ile, Leu, Phe) in core

No exposed hydrophobic residues

↓

Packing ? (RosettaHoles)

Inward-pointing polar residues at edge strands

↓

Local sequence-structure compatibility ?

↓

Designed proteins — ~100 structure-sequence pairs are selected. Buried polar and exposed hydrophobic residues are mutated by Foldit.

↓

Explore energy landscape (Rosetta ab initio structure prediction)

↓

Funnel-like

↓

Experimental characterization — ~10 designs

b

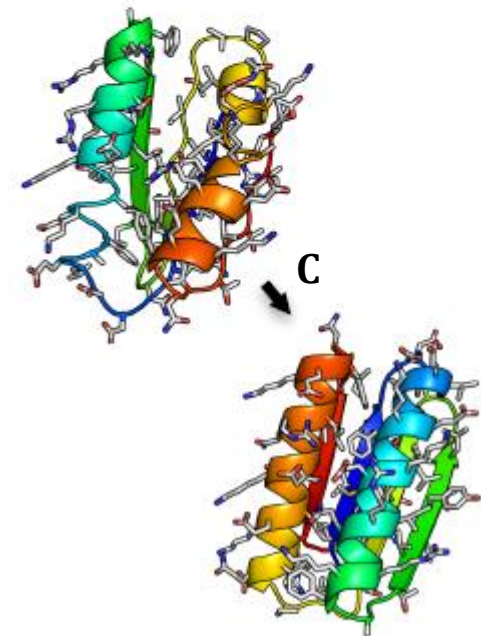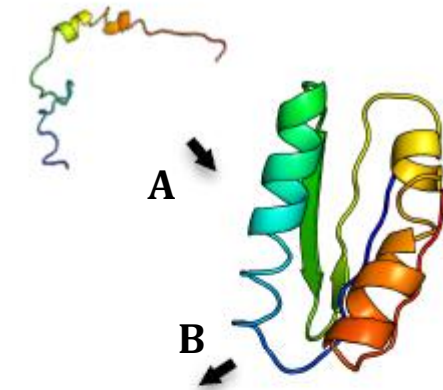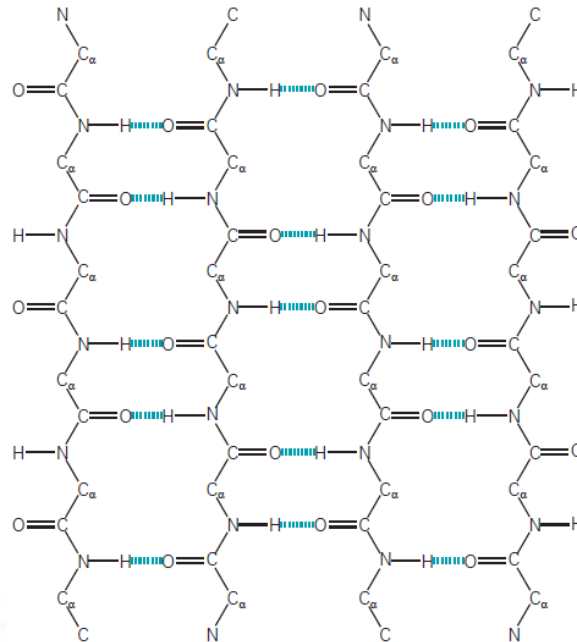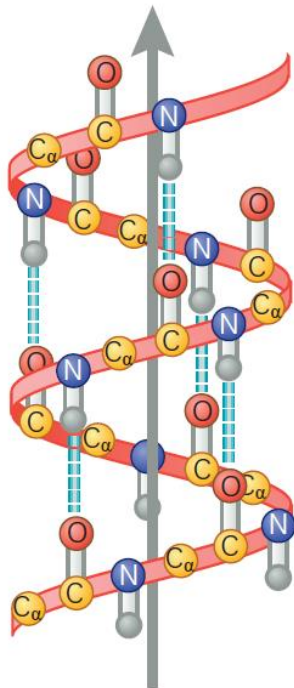Effective energy

Conformational space
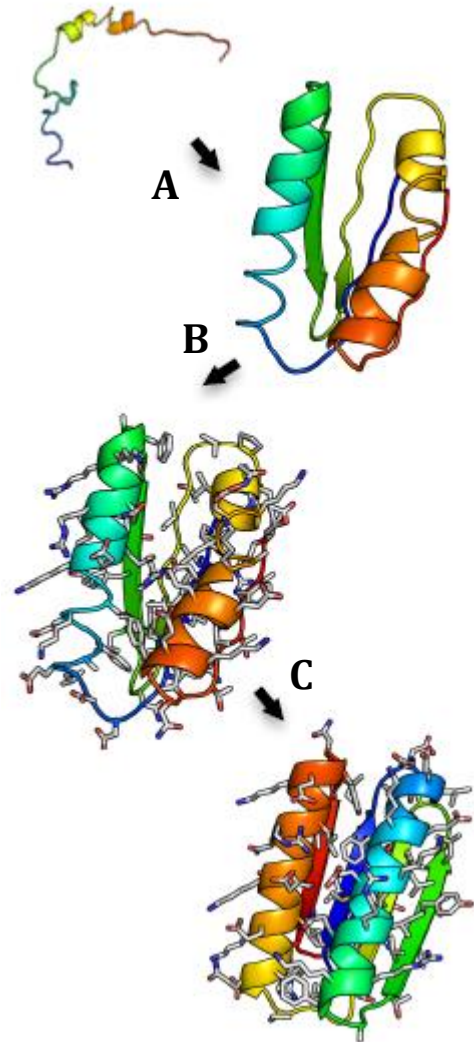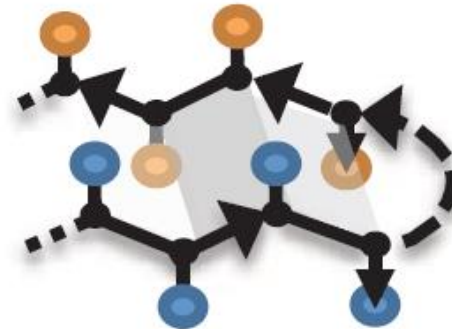
Protein design pipeline

Build the sequence-independent backbone model:
- Assign the secondary structures to backbone;
- Select the lengths of secondary structures and connecting loops based on rules.

The Ramachandran Plot.

A

B

C

Protein design pipeline
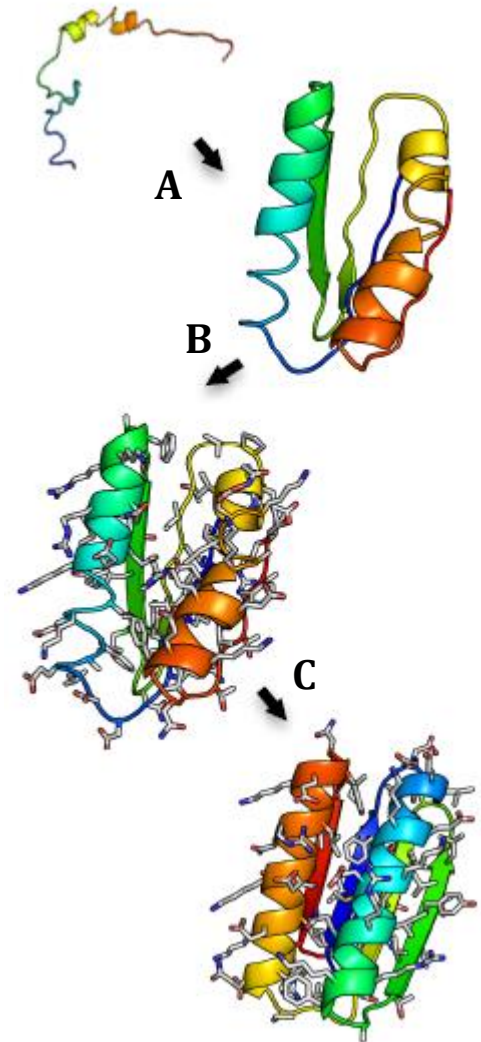
Build the sequence-independent backbone model:
- Assign the secondary structures to backbone;
- Select the lengths of secondary structures and connecting loops based on rules.

A. Run Monte Carlo simulations that minimize a potential function:

Potential = hydrogen bonds + repulsive force + compaction

Protein design pipeline

Build the sequence-independent backbone model:
- Assign the secondary structures to backbone;
- Select the lengths of secondary structures and connecting loops based on rules.

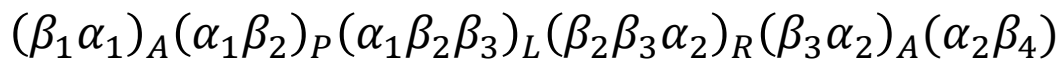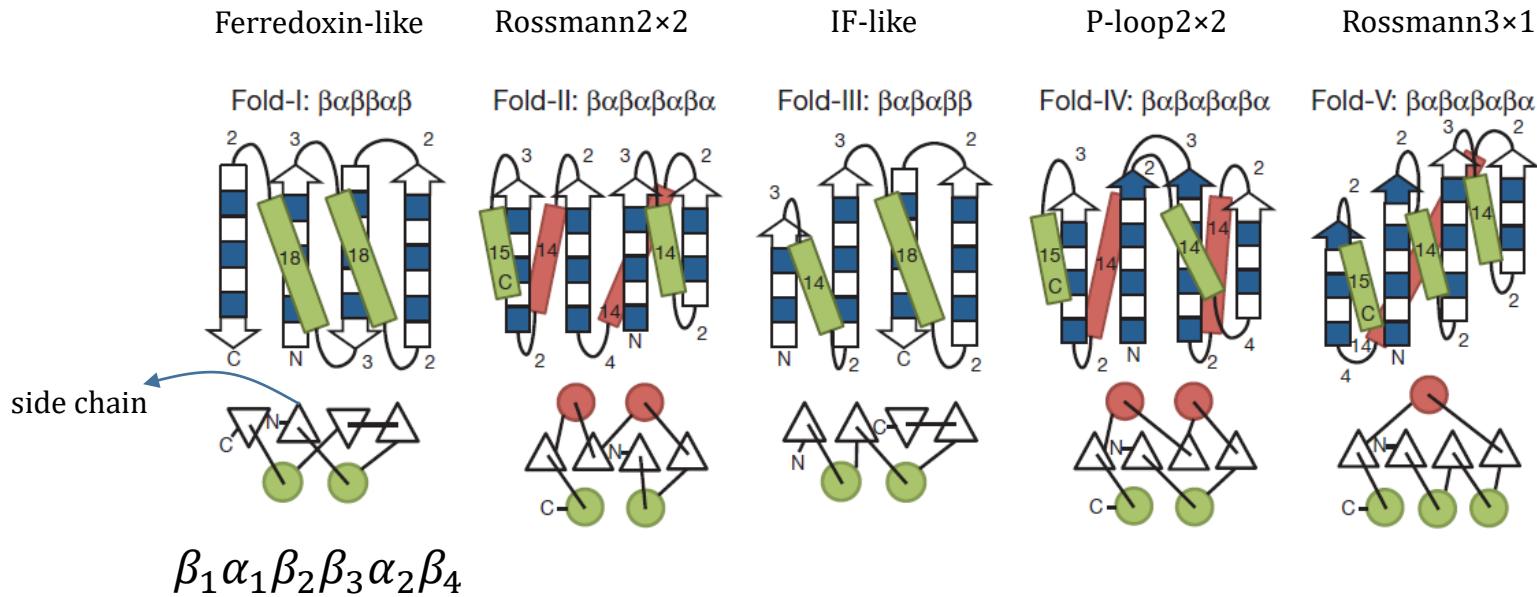**A.** Run Monte Carlo simulations that minimize a potential function:

Potential = hydrogen bonds + repulsive force + compaction

B. Design side chains that favor/stabilize the secondary structures and the tertiary structure.

Protein design pipeline

Build the sequence-independent backbone model:
- Assign the secondary structures to backbone;
- Select the lengths of secondary structures and connecting loops based on rules.

A. Run Monte Carlo simulations that minimize a potential function:

Potential = hydrogen bonds + repulsive force + compaction

B. Design side chains that favor/stabilize the secondary structures and the tertiary structure.
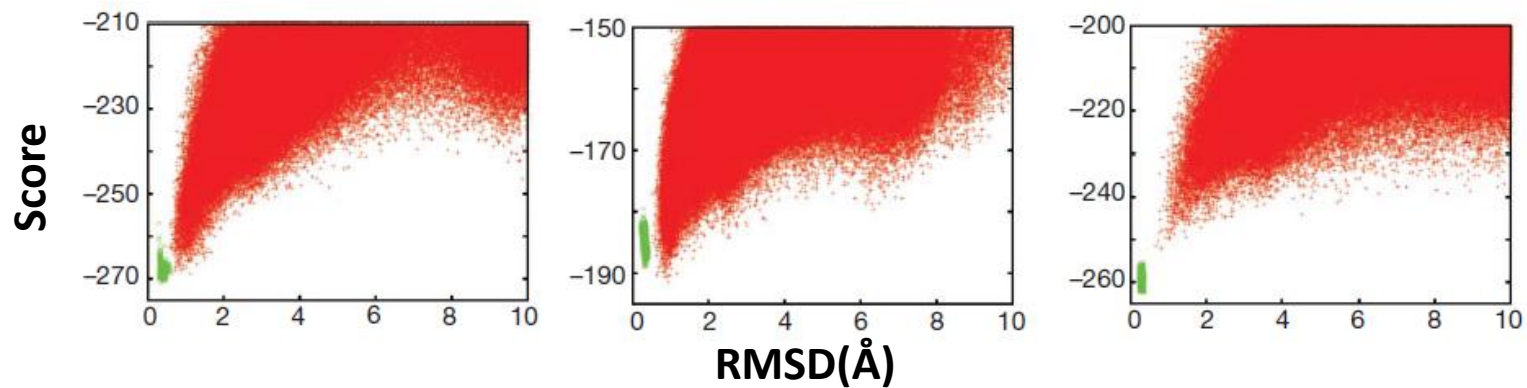
C. Relax the backbone and side chains all together.

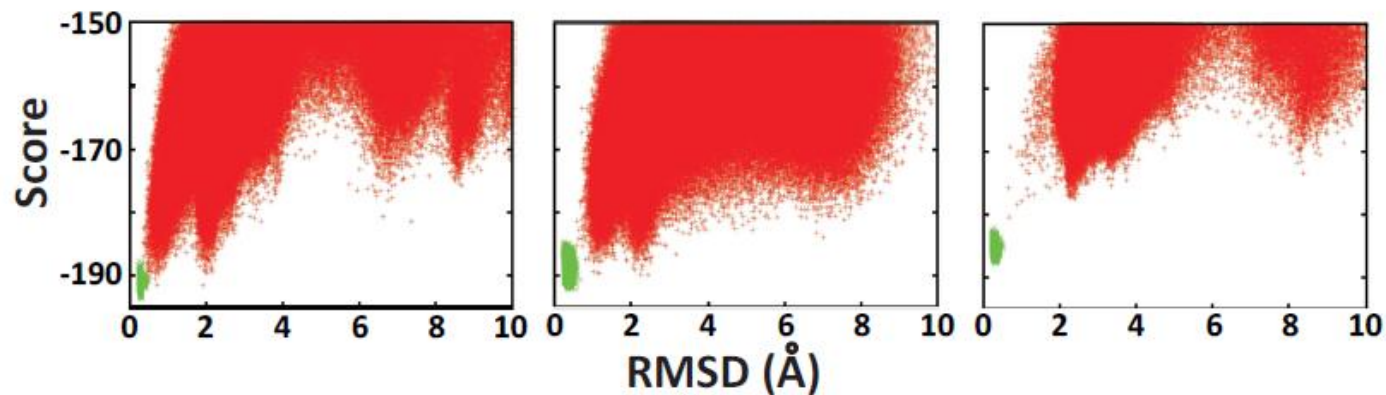# *De novo* design: 5 representative structures

Ferredoxin-like	Rossmann2×2	IF-like	P-loop2×2	Rossmann3×1
Fold-I: $\beta\alpha\beta\beta\alpha\beta$	Fold-II: $\beta\alpha\beta\alpha\beta\alpha$	Fold-III: $\beta\alpha\beta\alpha\beta\beta$	Fold-IV: $\beta\alpha\beta\alpha\beta\alpha\beta\alpha$	Fold-V: $\beta\alpha\beta\alpha\beta\alpha\beta\alpha$

side chain

$$\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$$

$$(\beta_1\alpha_1)_A(\alpha_1\beta_2)_P(\alpha_1\beta_2\beta_3)_L(\beta_2\beta_3\alpha_2)_R(\beta_3\alpha_2)_A(\alpha_2\beta_4)$$

BLAST E-value < 0.02 against the NCBI nr database

Simulated energy landscapes

Funneled energies

Non-funneled energies

Experimental characterization: Circular Dichroism (CD)

The signal of CD is the ellipticity of the circularly polarized light.

- The intensity of light will be reduced by propagating through a sample.
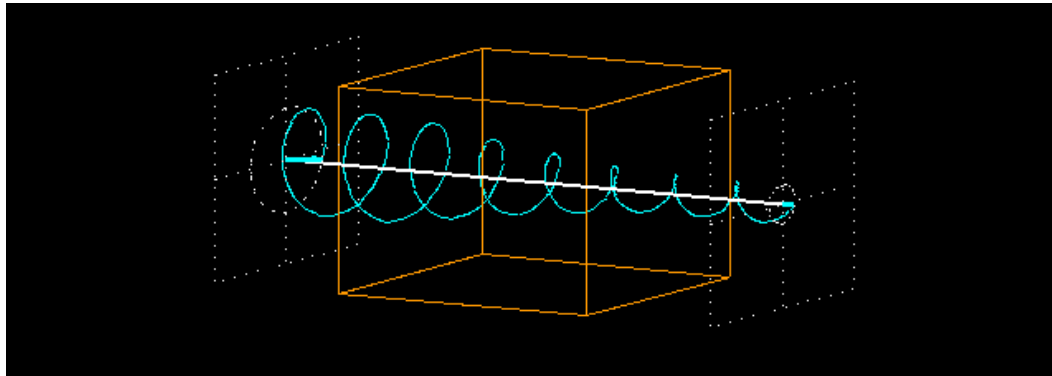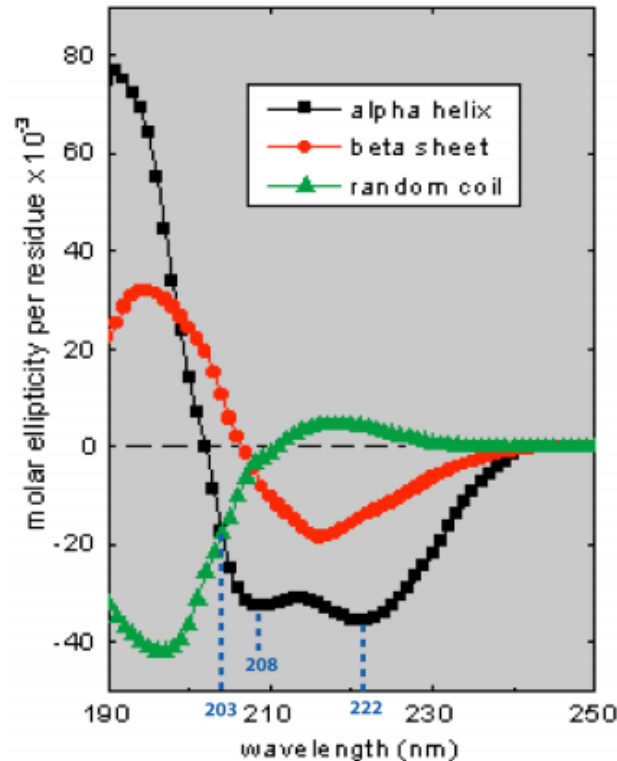

- If the absorptions of left and right circularly polarized light are different, then the line becomes an ellipse.
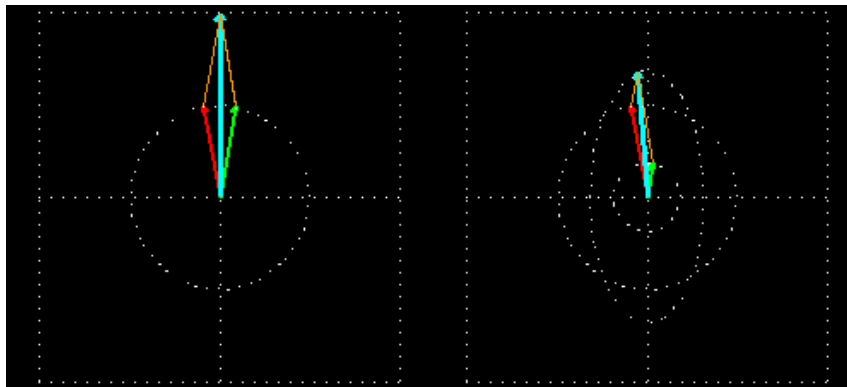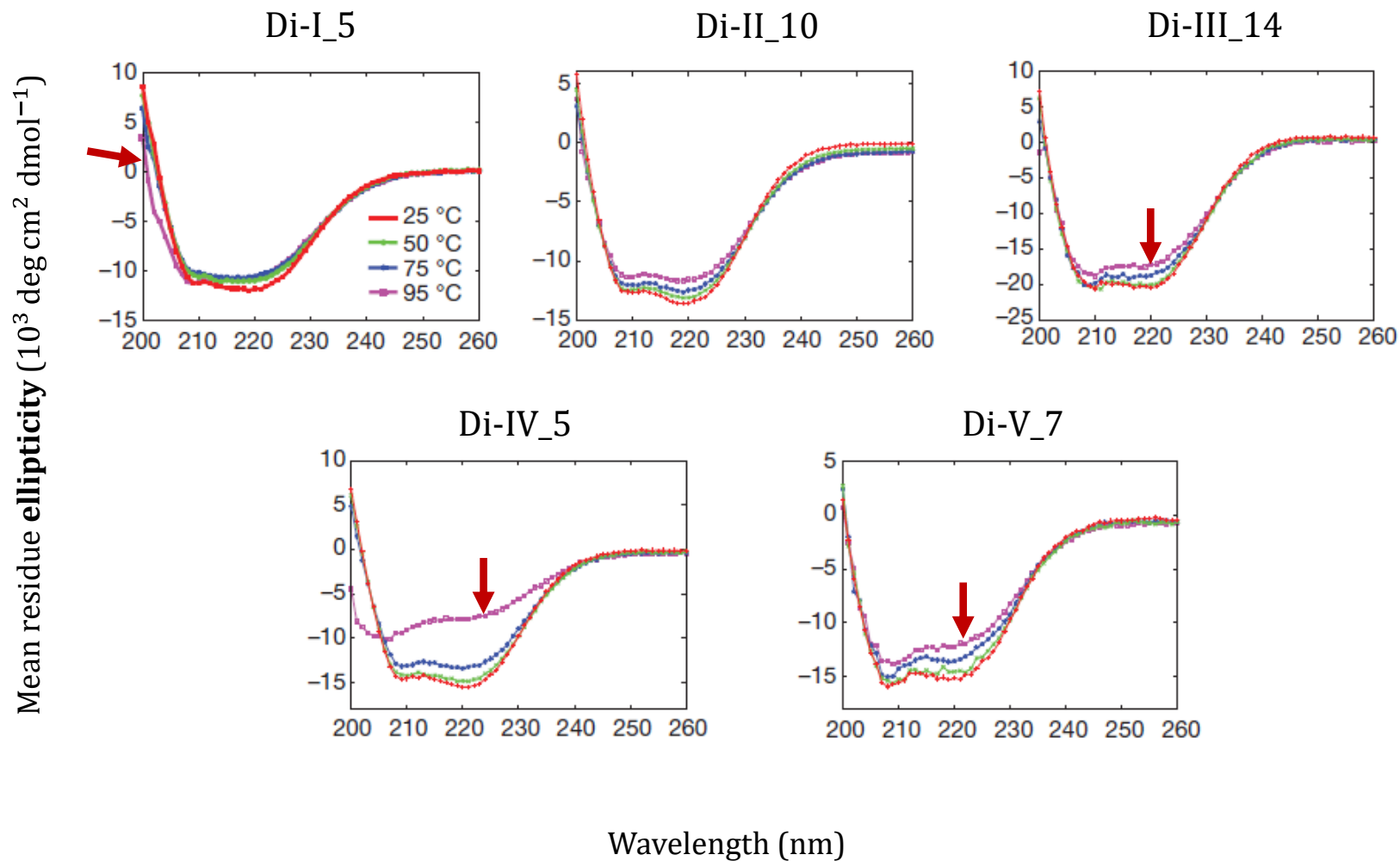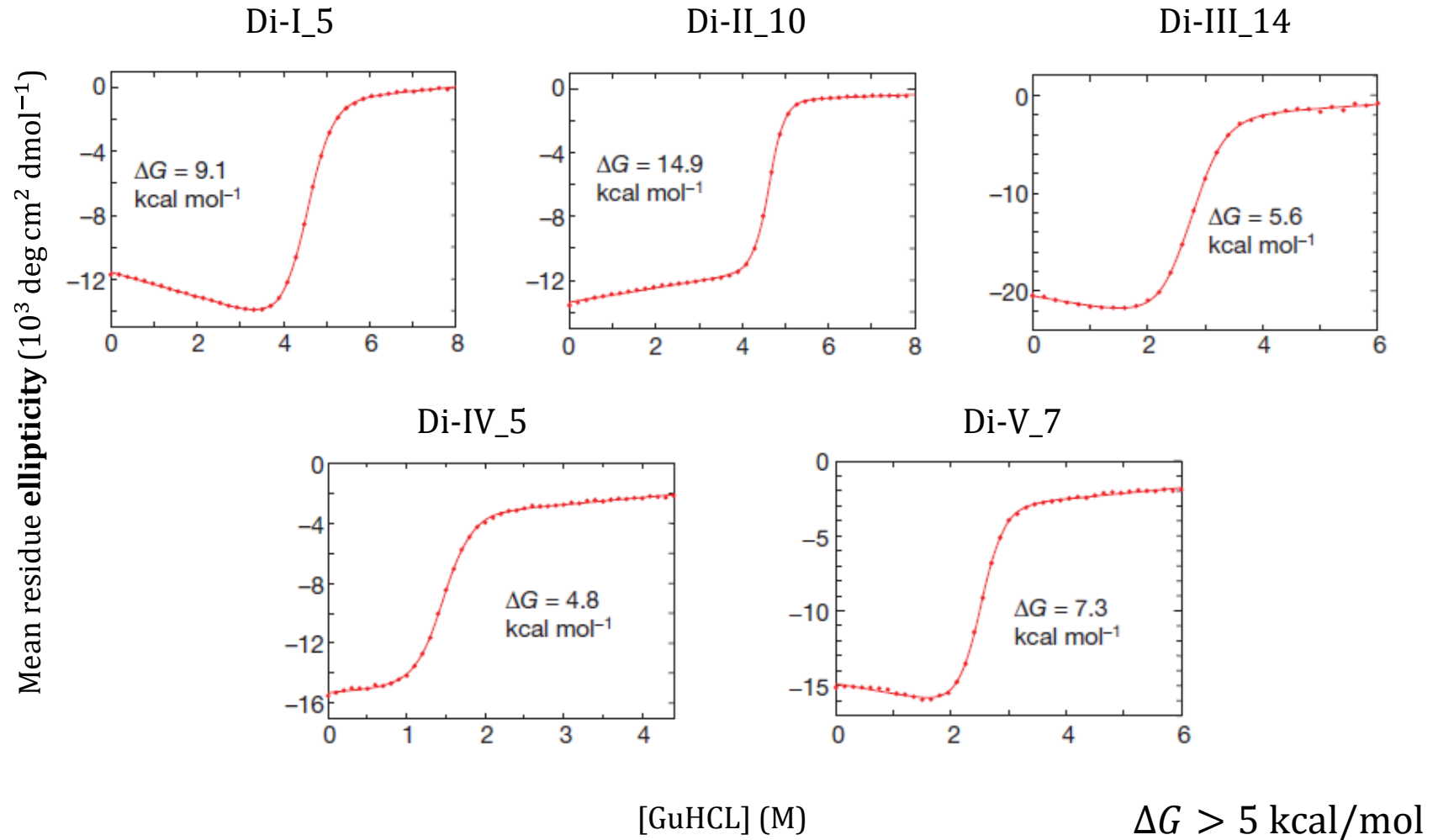
Experimental characterization: Circular Dichroism (CD)

- **Alpha helix** has negative bands at 222nm and 208nm and a positive one at 190nm.

- **Beta sheet** shows a negative band at 218 nm and a positive one at 196 nm.

- **Random coil** has a positive band at 212 nm and a negative one around 195 nm.

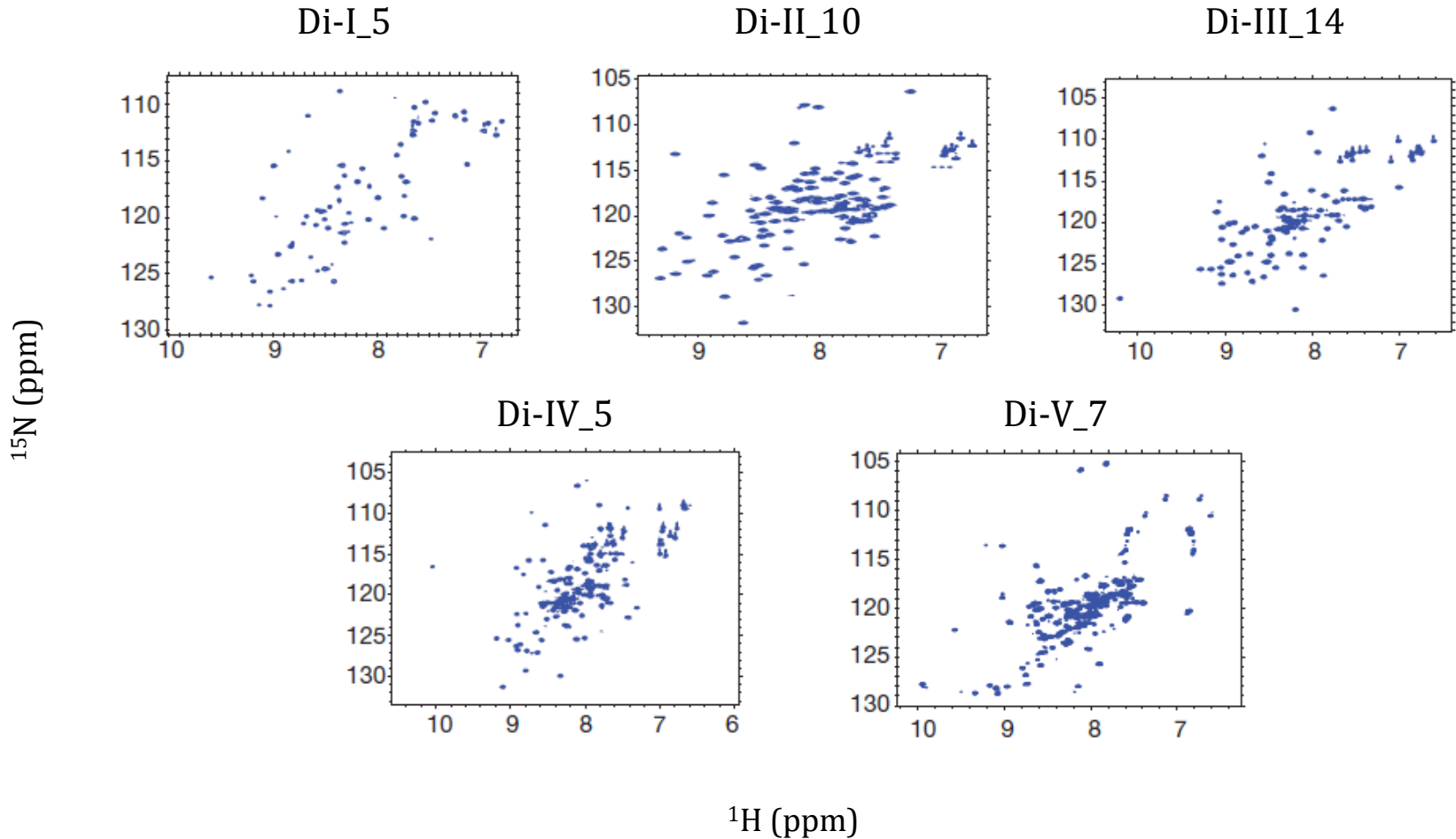# Experimental characterization: Circular Dichroism (CD)

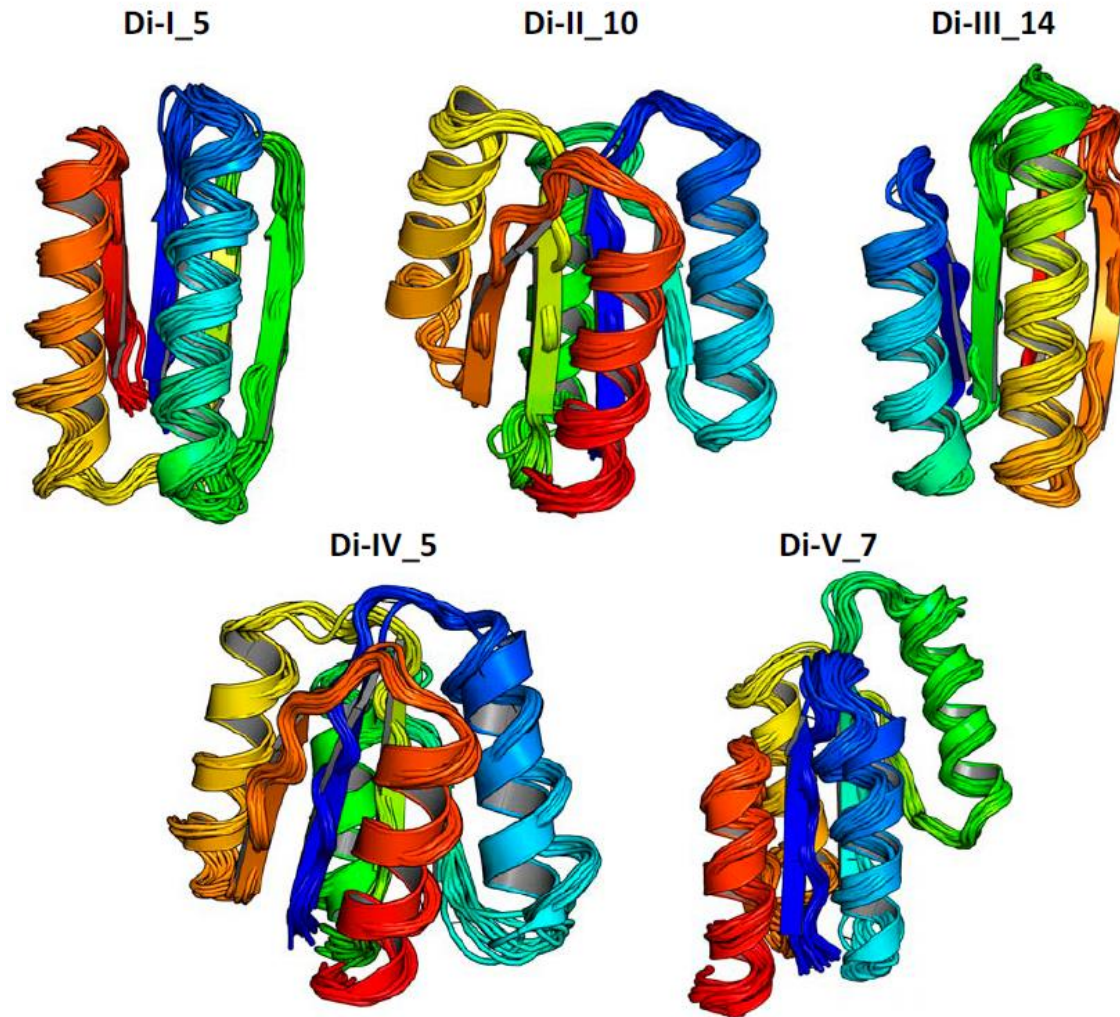# Experimental characterization: Chemical denaturation

CD at 220 nm

$\Delta G > 5$ kcal/mol

Experimental characterization: 2D-NMR (HSQC)

"The two-dimensional ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) spectra show the expected number of well-dispersed sharp peaks."
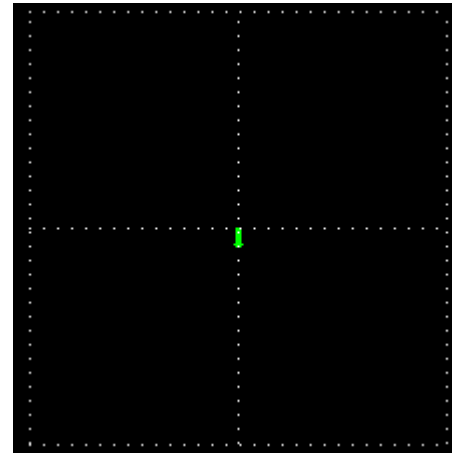
Experimental-Simulation comparison

Di-I_5　　　Di-II_10　　　Di-III_14
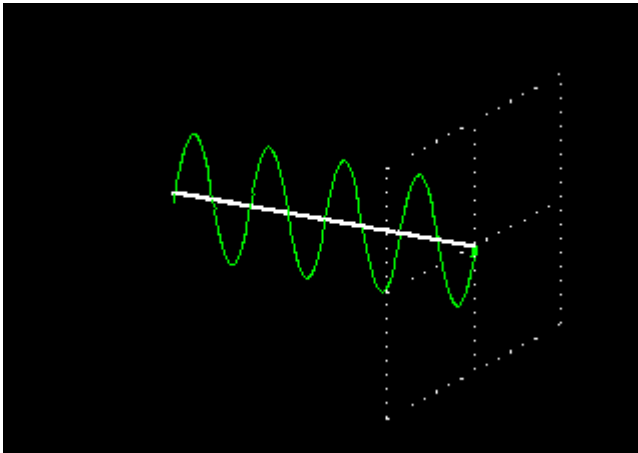
Di-IV_5　　　Di-V_7

RMSD < 2Å

Experimental characterization: Circular Dichroism (CD)
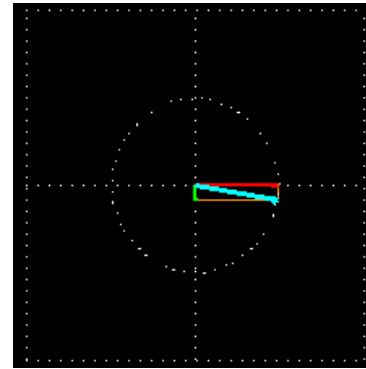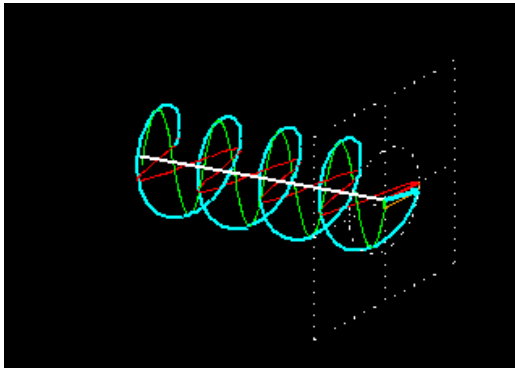
Light is essentially a electromagnetic field.

- The natural light fluctuates in every directions.

- The linear polarized light fluctuates in one direction.

Experimental characterization: Circular Dichroism (CD)

Circularly polarized light is the linear combination of two linearly polarized light.

- There are <u>left</u> and <u>right</u> circularly polarized.

- The combination of two mirror symmetric circularly polarized light is reduced to a linearly polarized light.