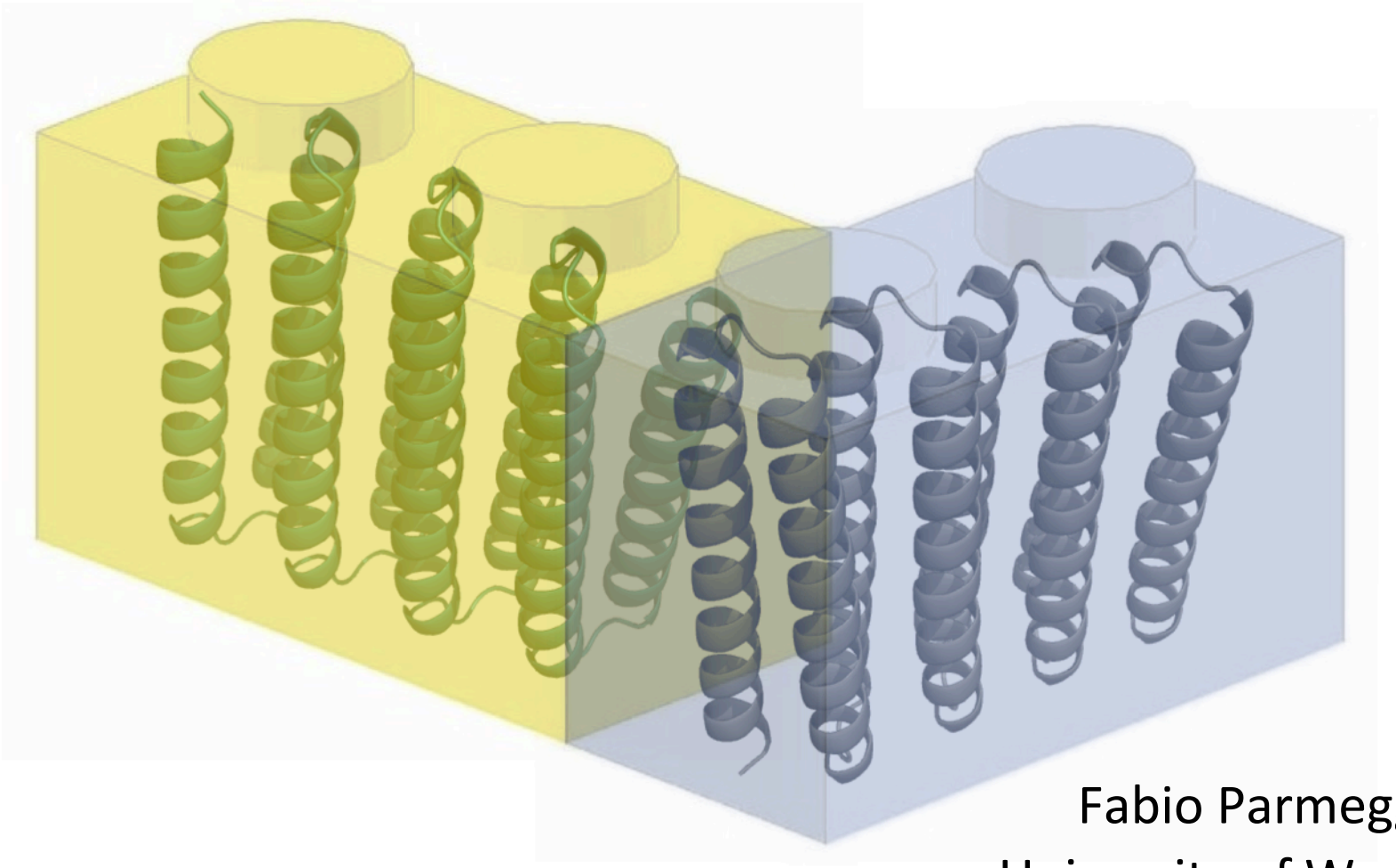
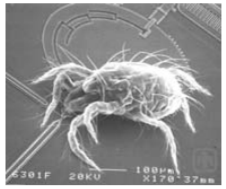
Design of Repeat Proteins:
Larger Structures from simpler topologies



Fabio Parmeggiani
University of Washington

The Scale of Things – Nanometers and More

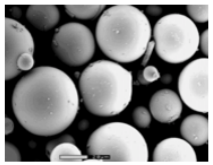
Things Natural



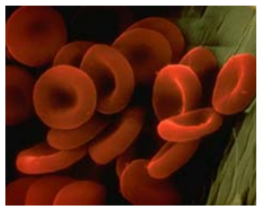
Dust mite
200 μm

Ant
~ 5 mm

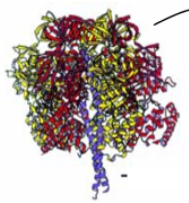
Human hair
~ 60-120 μm wide



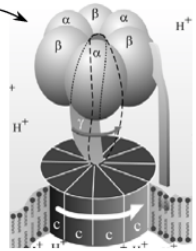
Fly ash
~ 10-20 μm



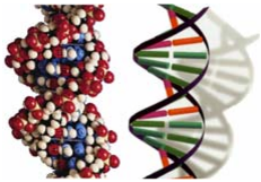
Red blood cells
(~7-8 μm)



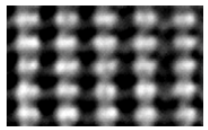
~10 nm diameter



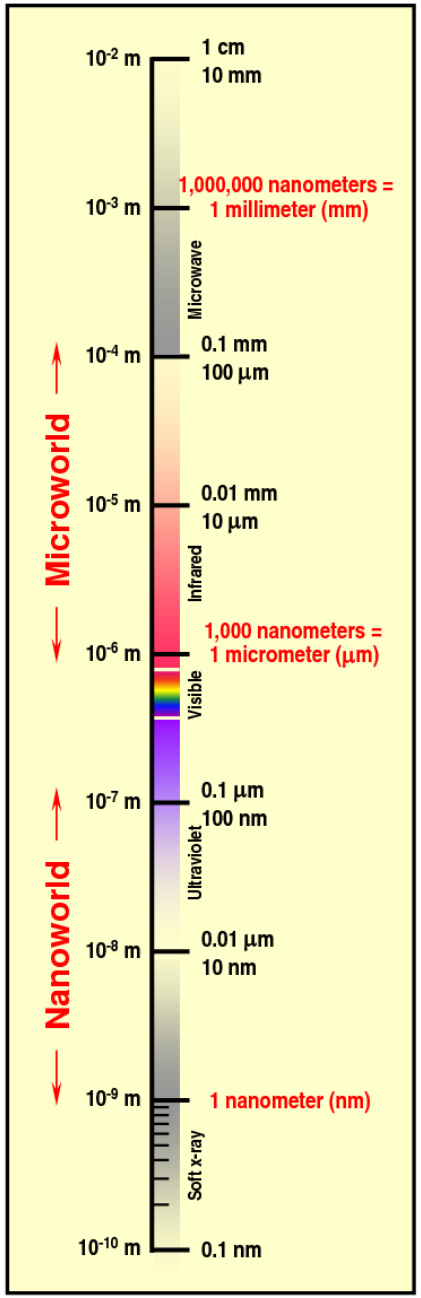
ATP synthase



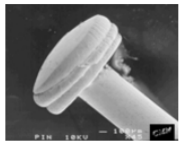
DNA
~2-1/2 nm diameter



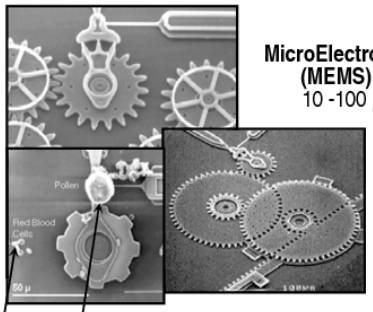
Atoms of silicon
spacing 0.078 nm



Things Manmade

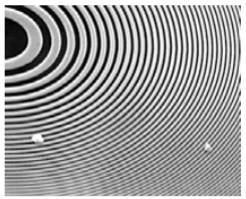


Head of a pin
1-2 mm

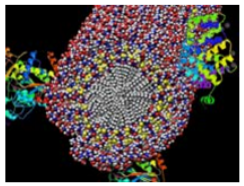


MicroElectroMechanical (MEMS) devices
10 -100 μm wide

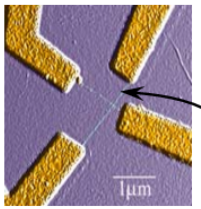
Pollen grain
Red blood cells



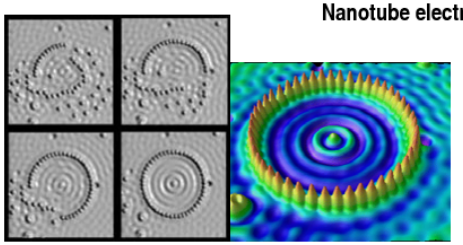
Zone plate x-ray "lens"
Outer ring spacing ~35 nm



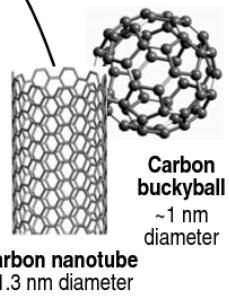
Self-assembled,
Nature-inspired structure
Many 10s of nm



Nanotube electrode

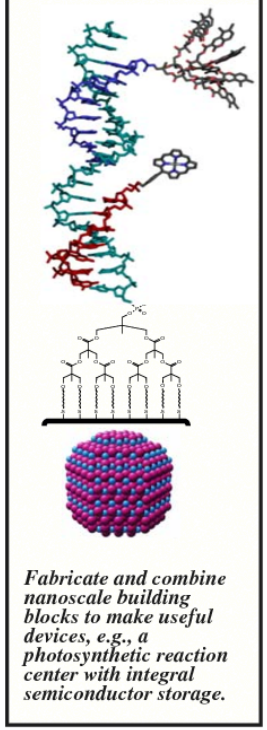


Quantum corral of 48 iron atoms on copper surface positioned one at a time with an STM tip
Corral diameter 14 nm



Carbon buckyball
~1 nm diameter

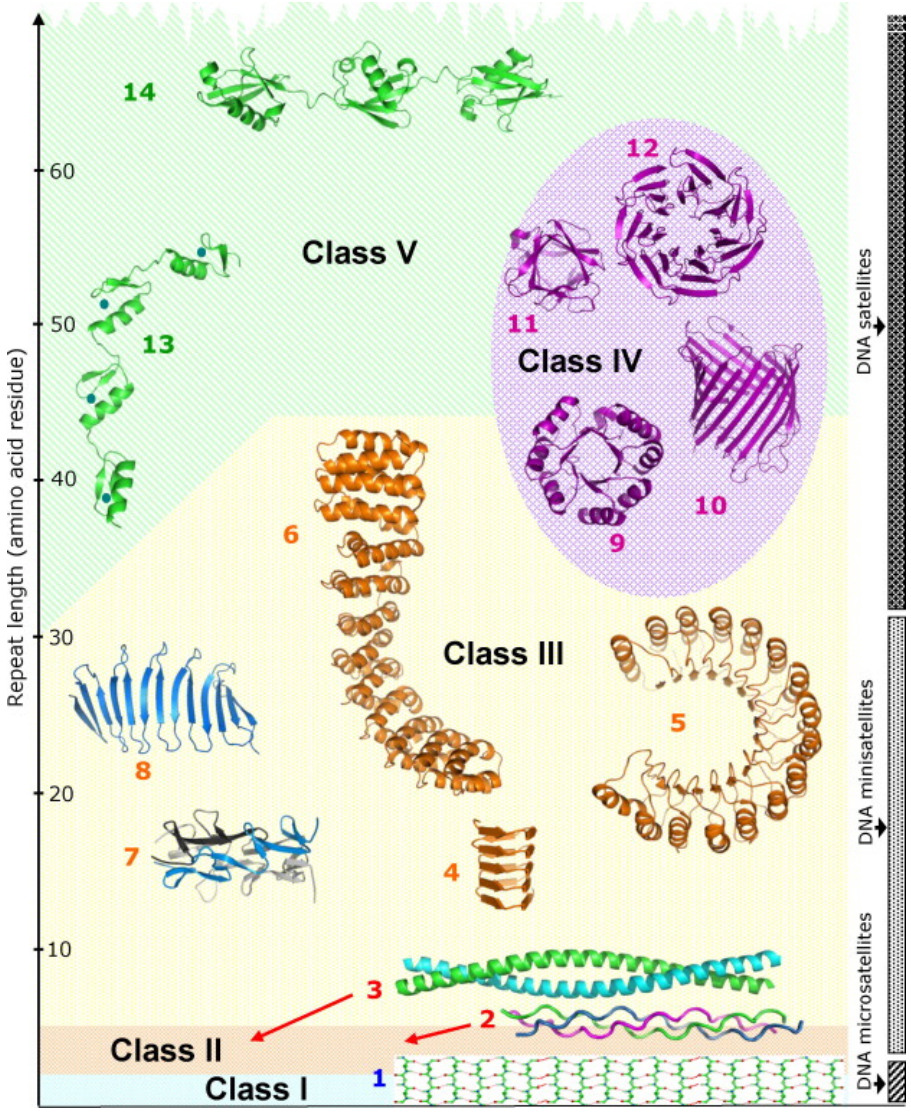
Carbon nanotube
~1.3 nm diameter



Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy

How to design large proteins?

A modular approach



Repeat protein:
one chain with internal repeated sequences

Repeat proteins among the most abundant families in database (LRR, ANK, WD40)

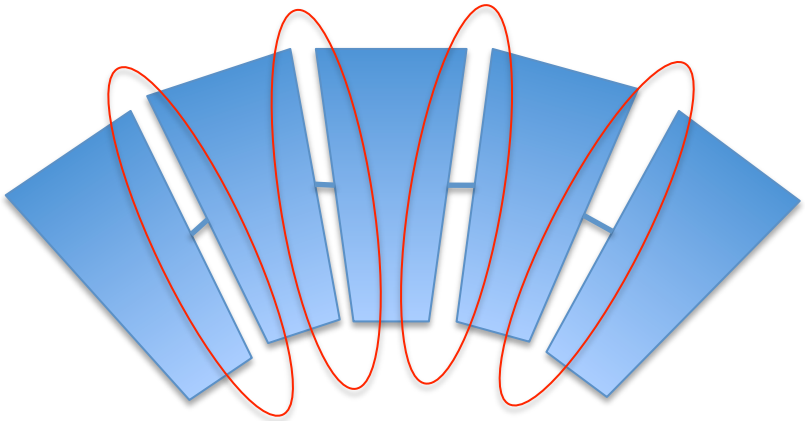
Involved often in protein-protein interaction

Generated by gene duplication events (early or late)

Kajava, J Struct Biol 2011

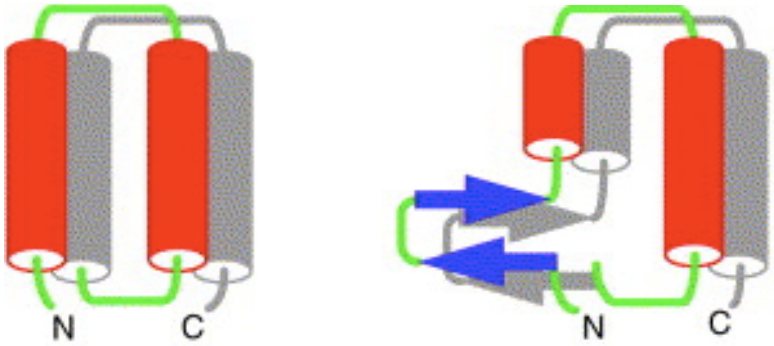
Protein repeats as building blocks

Repeat



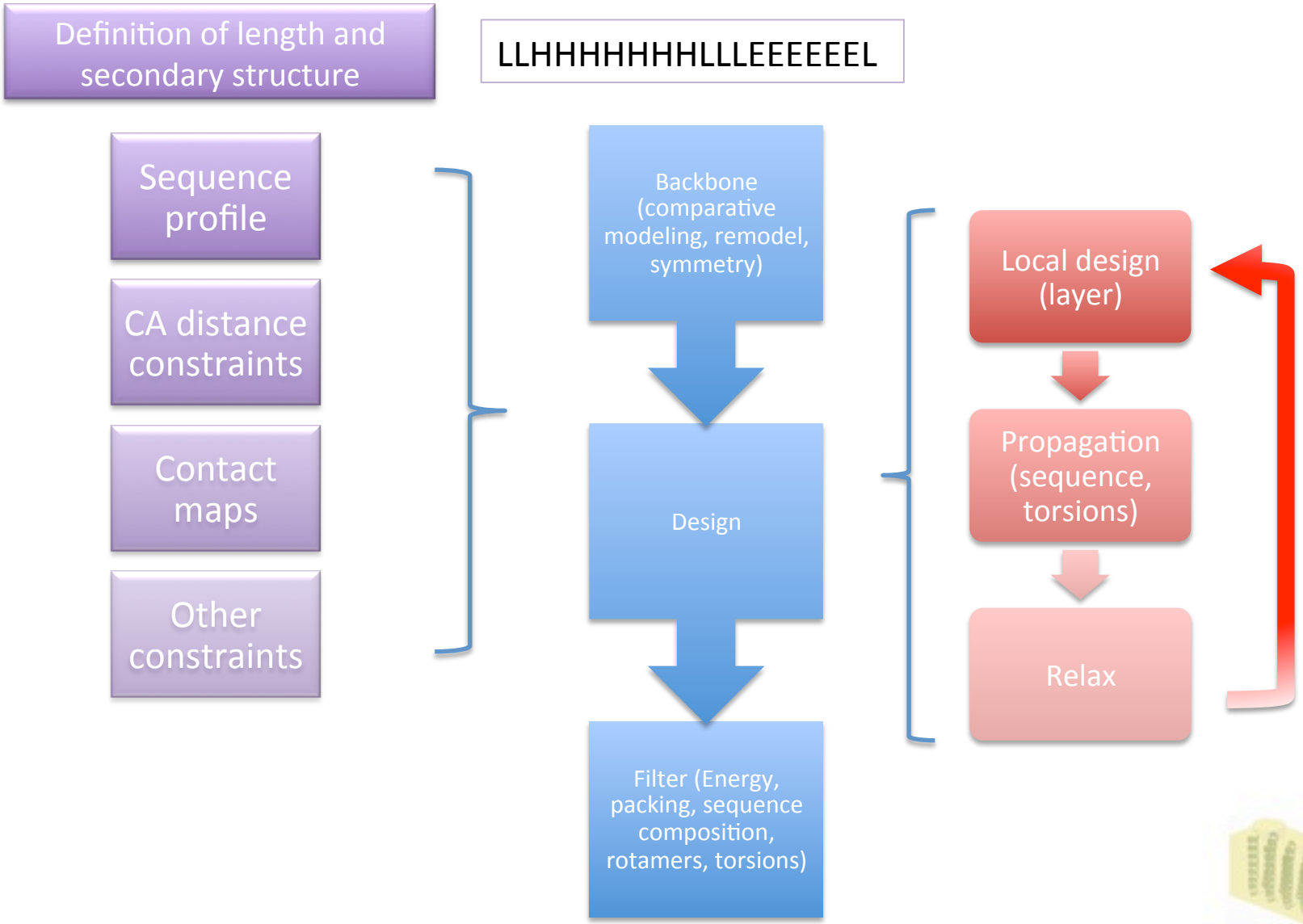
Repeat protein
Single protein chain

Conserved interface

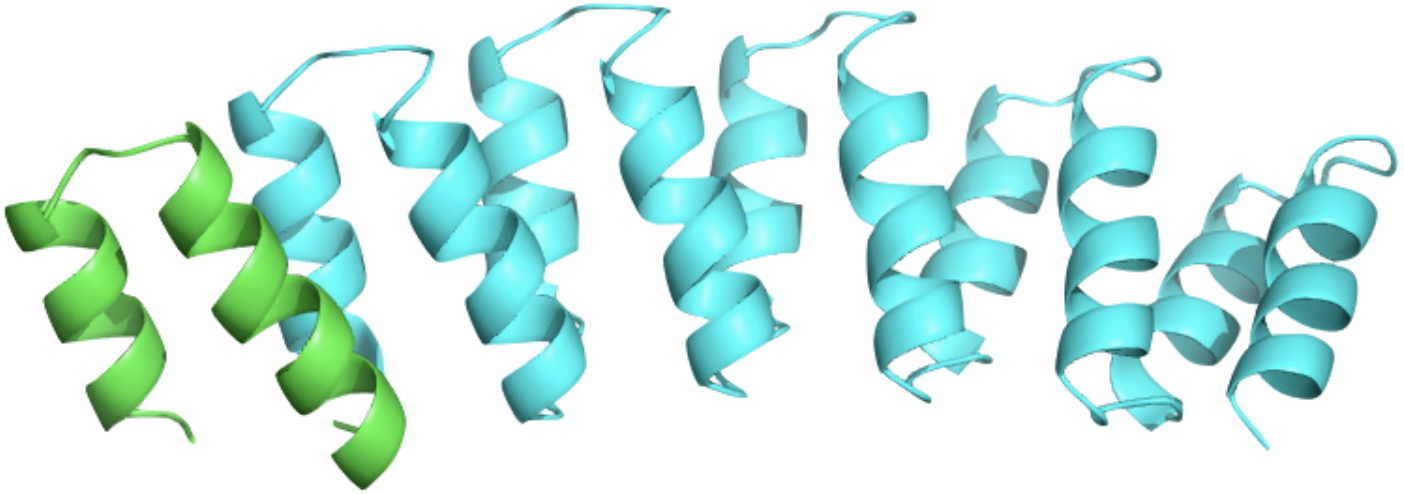


Mod. from Kobe, B. and Kajava, A. V.
Trends Biochem. Sci. 2000

Design protocol



Remodel: repeat mode



blueprint

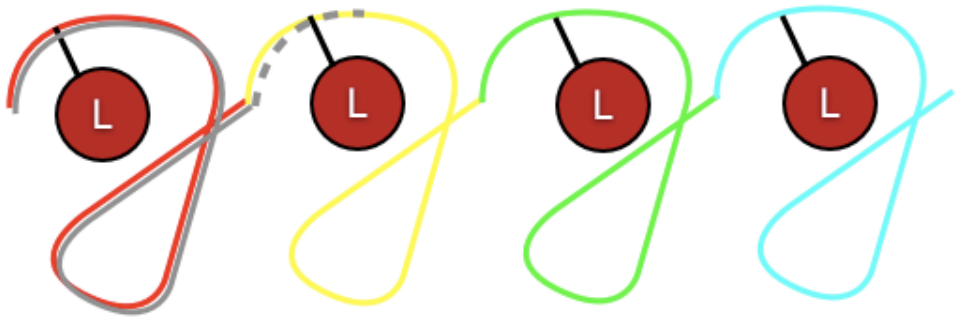
```
1 A H
0 x H
0 x H
0 x H
0 x H
0 x H
0 x H
0 x H
0 x H
0 x L
0 x L
0 x L
```

or

```
1 Y H
2 Q H
3 A H
4 A .
5 V .
6 D .
7 G .
8 A .
9 R .
10 K L
11 A L
12 N L
```

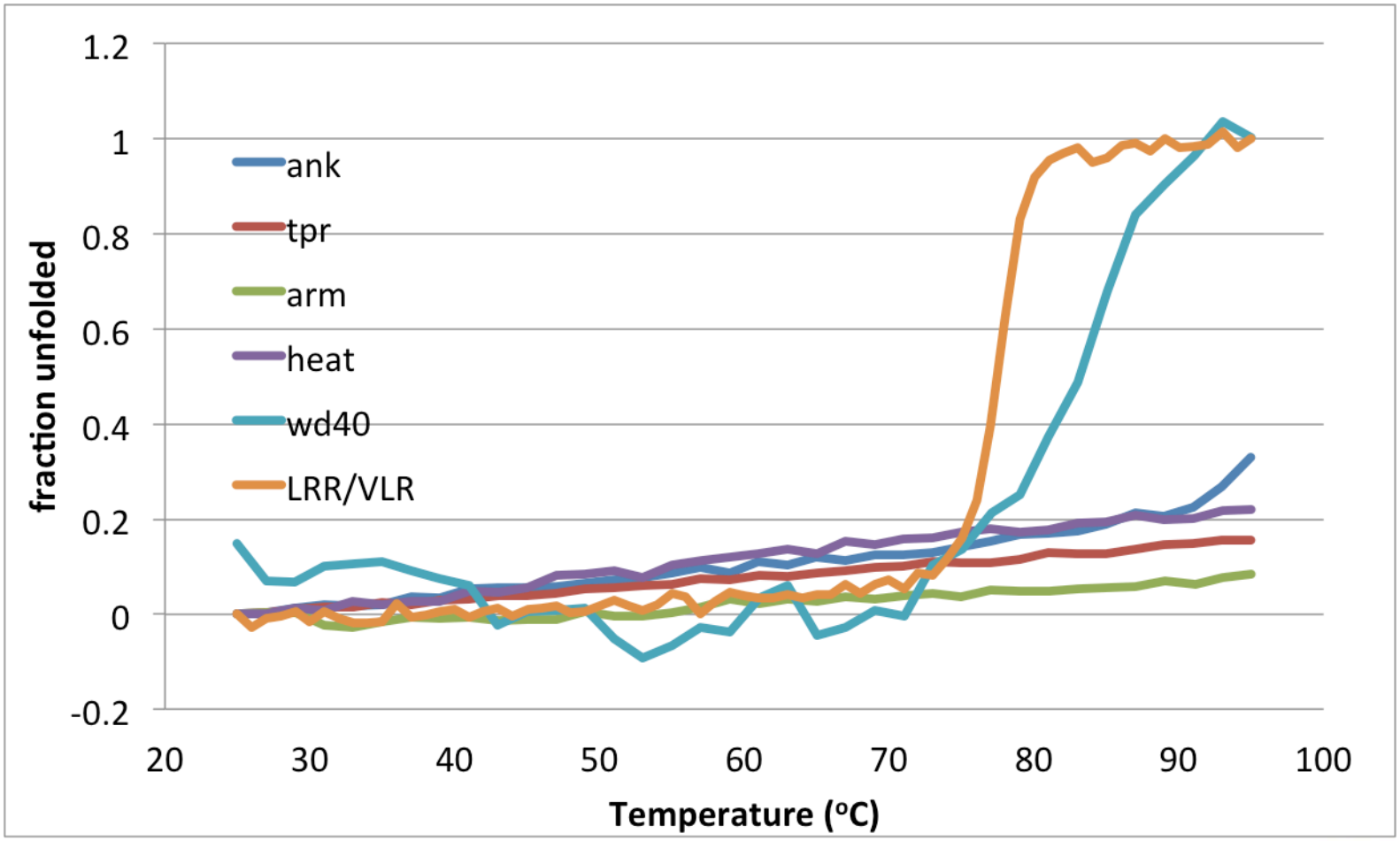
Rotamer link + NCS constraints

-repeat_structure 4

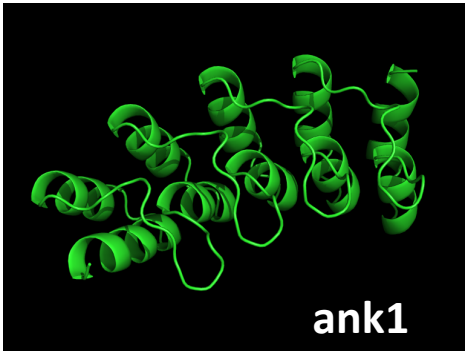


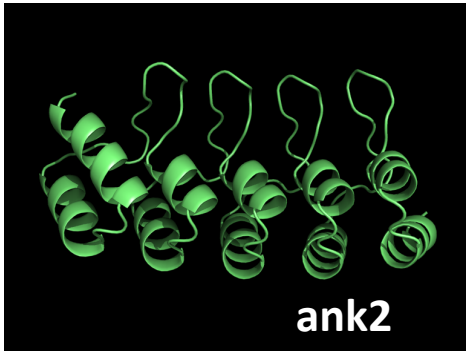
Possu Huang

Design of existing repeat protein families

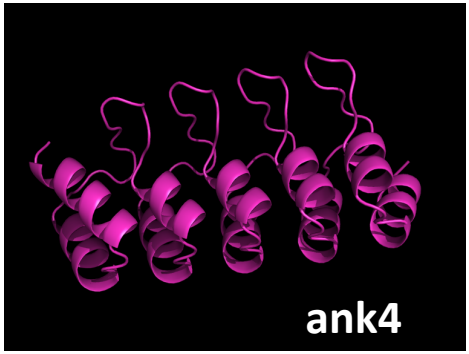


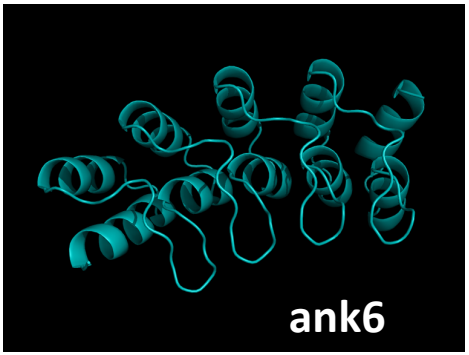
family	tested	expressed	soluble	folded	monomer
ankirin	6	6	6	6	6
tpr	6	6	6	5	3
armadillo	8	7	7	3	3
heat	8	8	5	2	1
Wd40 #	12	9	7	2	1*
LRR/VLR	2	2	2	2	2
Kelch #	10	0			
hexapep	3	0			
pentapep	5	0			

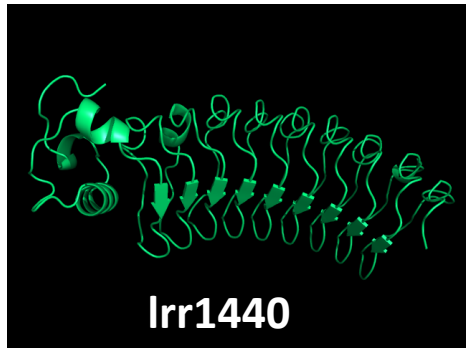

ank1

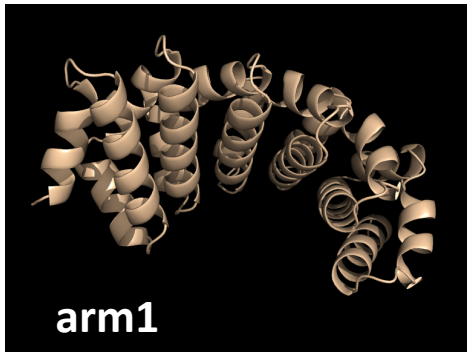

ank2

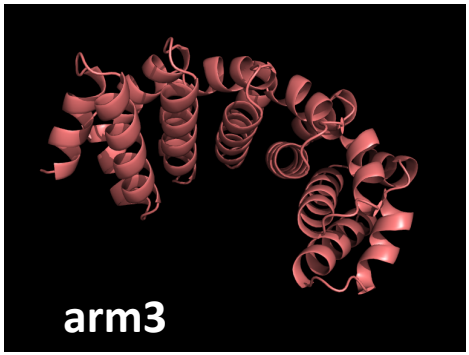

ank3

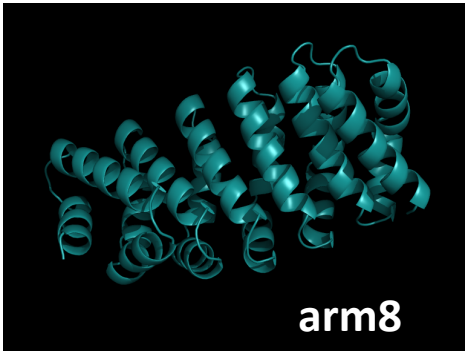

ank4

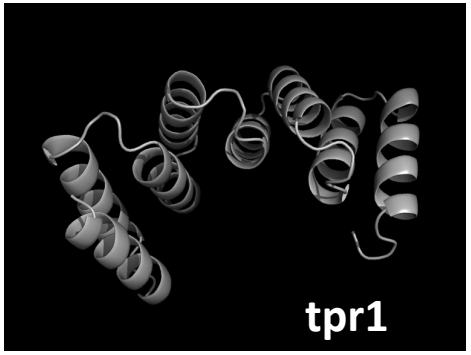

ank6

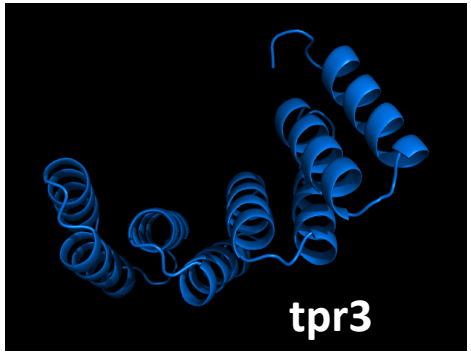

lrr1440

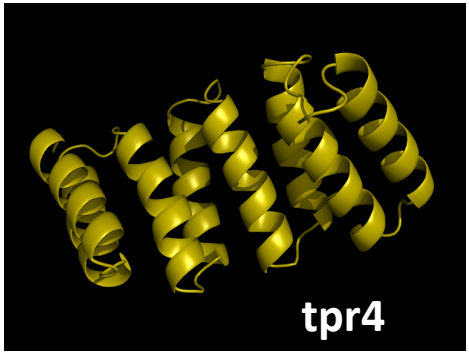

arm1


arm3

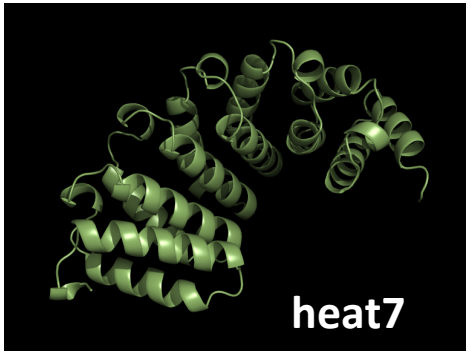

arm8


tpr1


tpr3

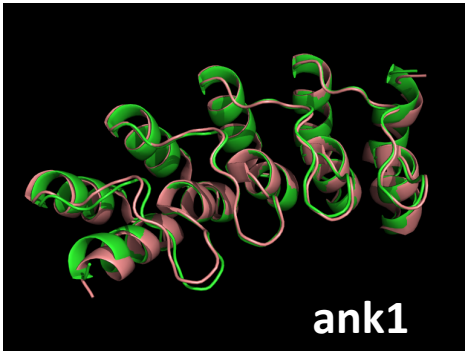

tpr4


heat2

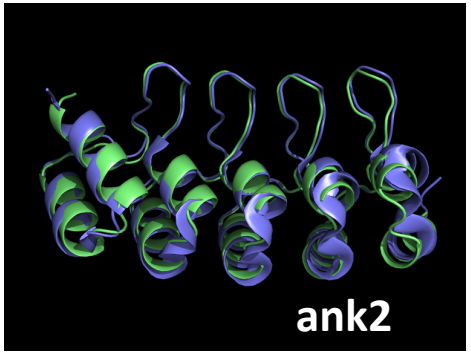

heat7


wd40_8

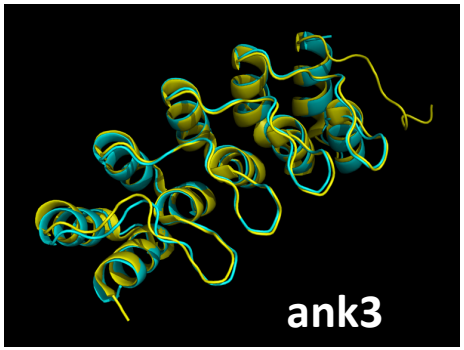

wd40_6



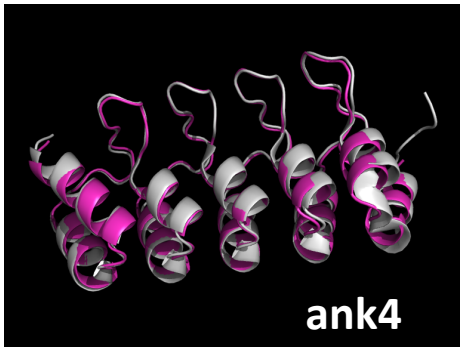
RMSD Cα: 0.5Å
126 residues



RMSD Cα: 0.9Å
126 residues



RMSD Cα: 0.7Å
126 residues



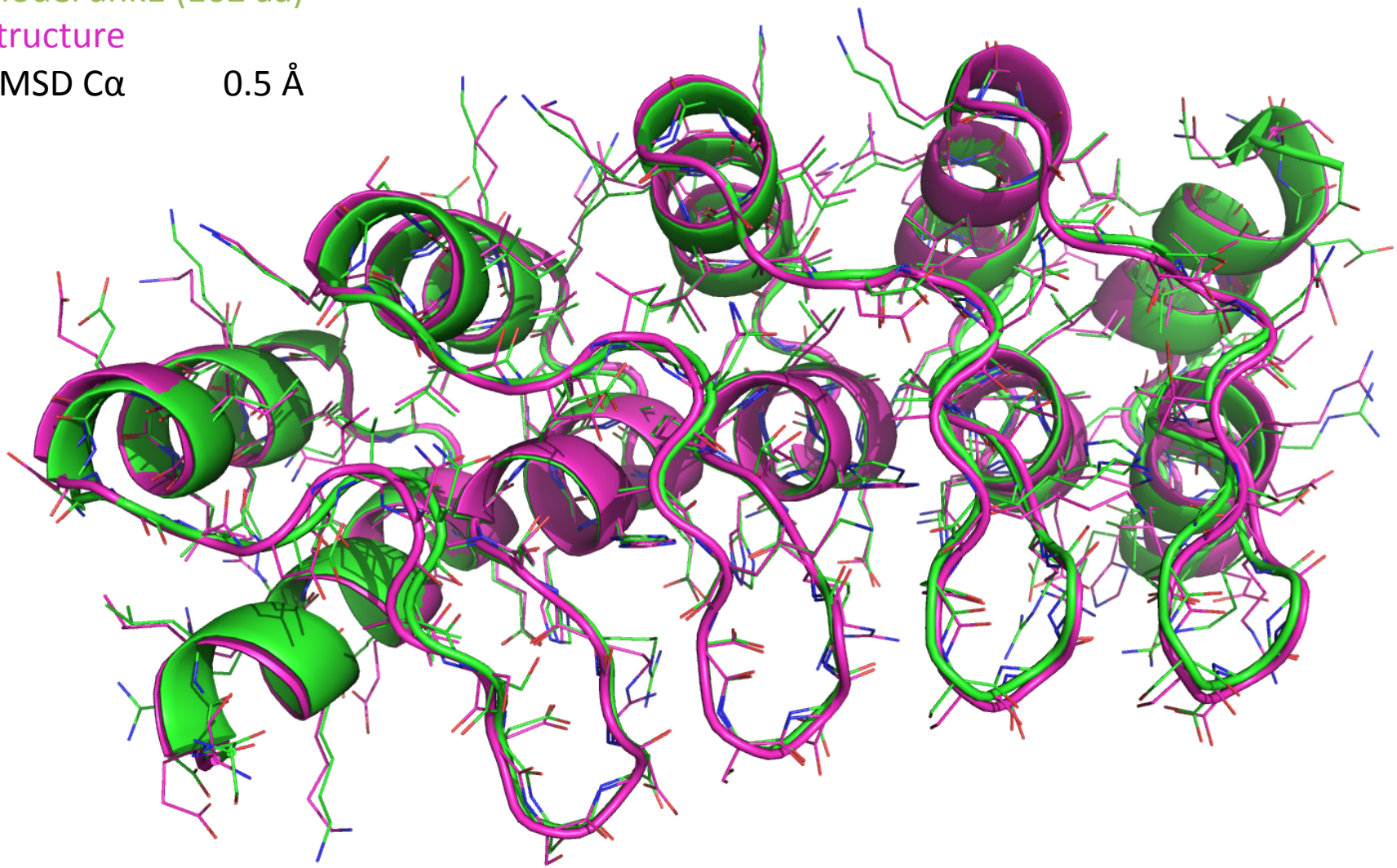
RMSD Cα: 0.7Å
126 residues

RMSD Cα: 1.0Å
252 residues

Comparison with
crystal structures

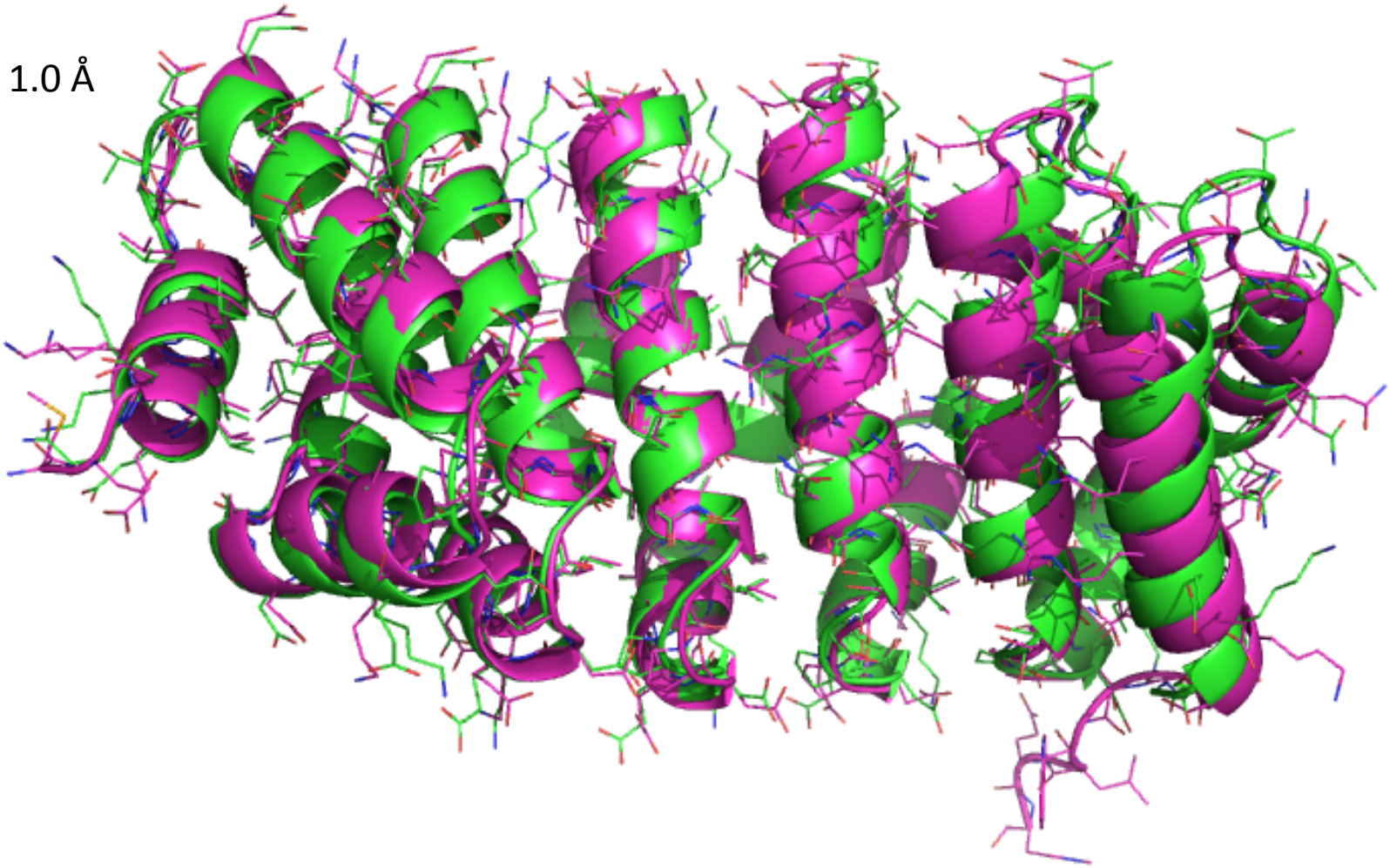
Protocol validation: crystal structure

Model ank1 (162 aa)
Structure
RMSD Cα 0.5 Å



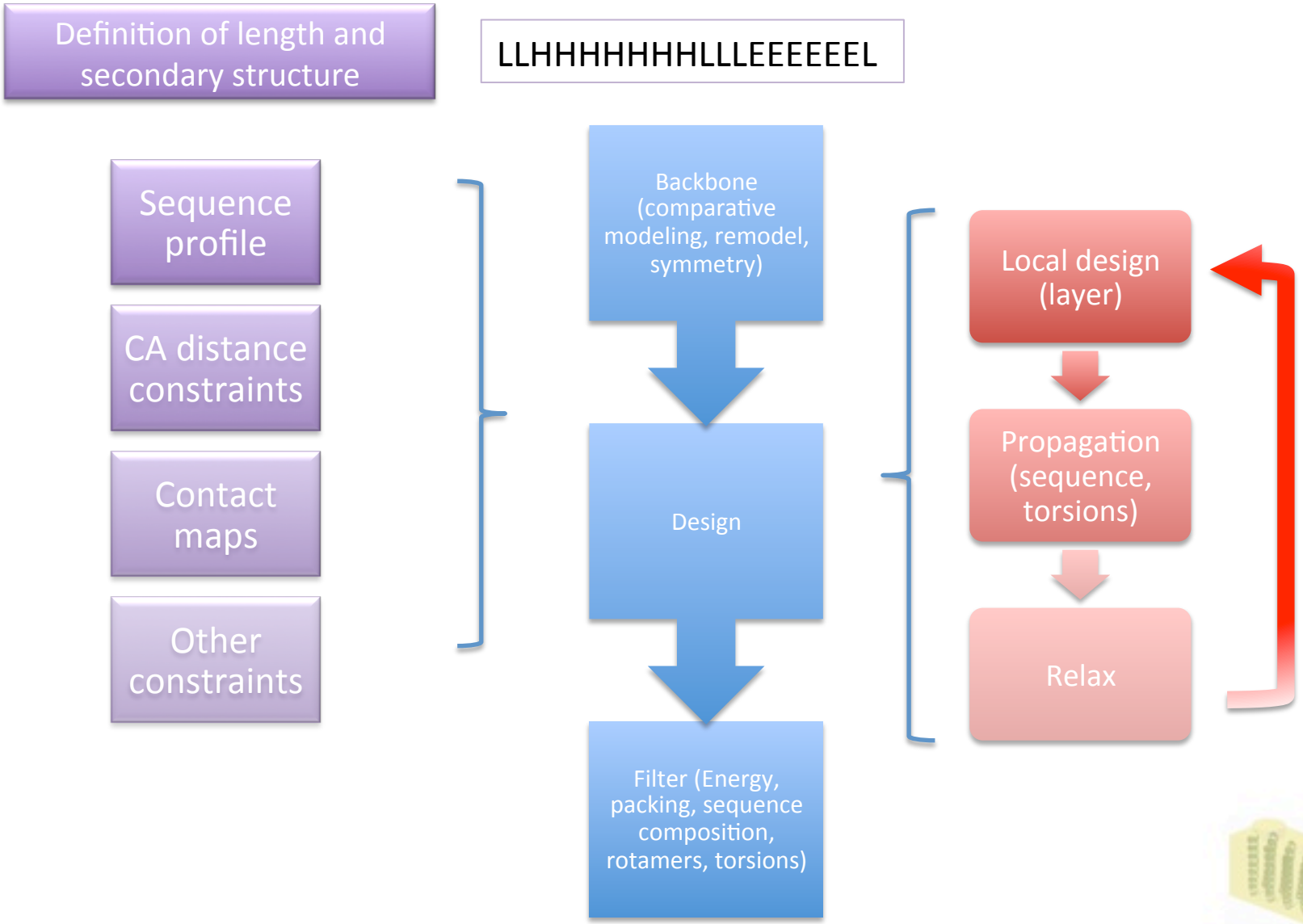
Protocol validation: crystal structure

Model arm8 (252 aa)
Structure
RMSD Cα 1.0 Å

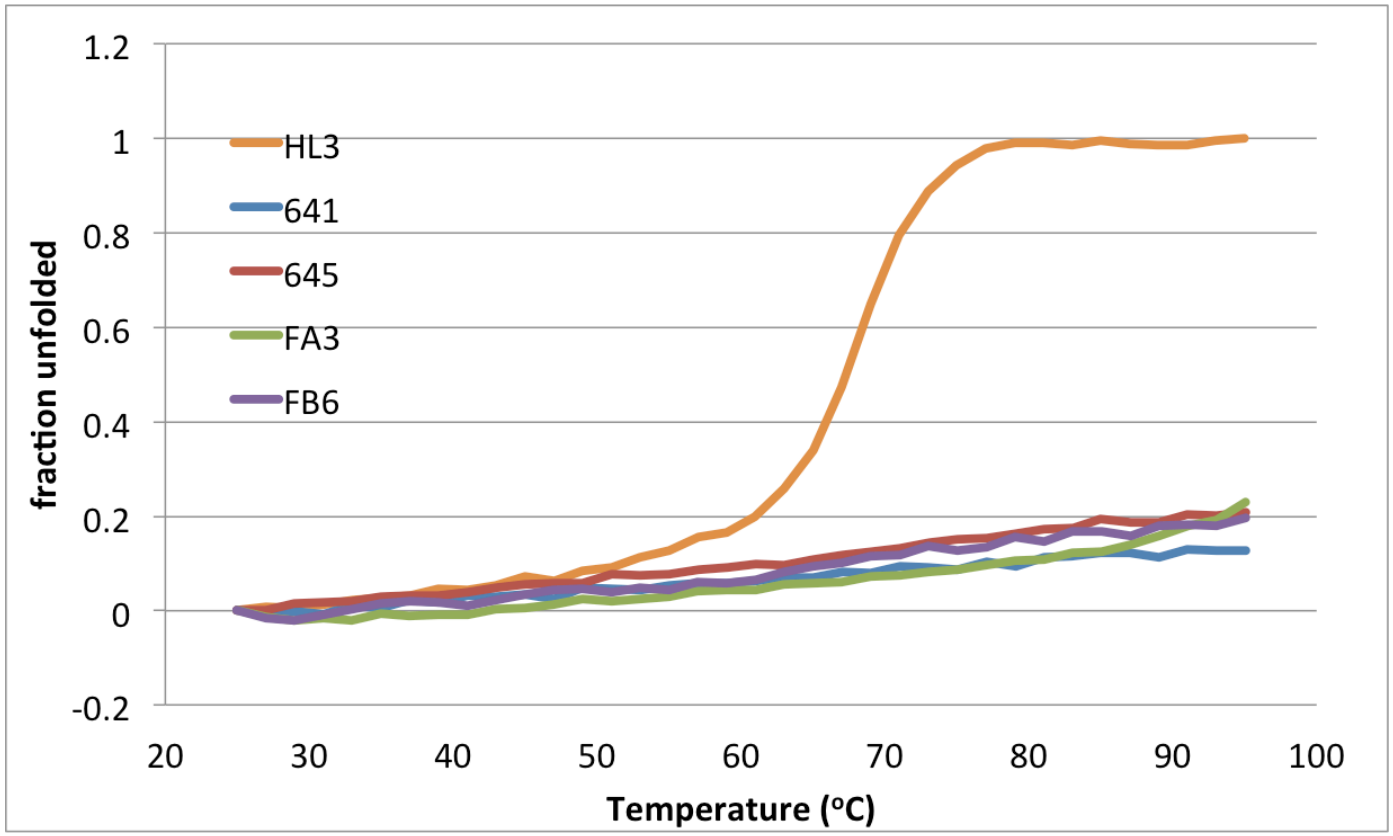


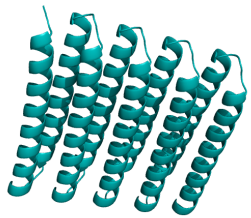
De novo repeat protein design

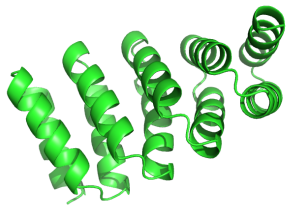
De novo designs

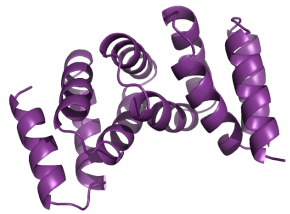


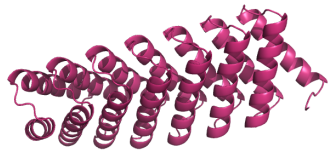
De novo designs: two helices topologies

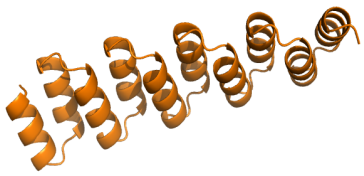






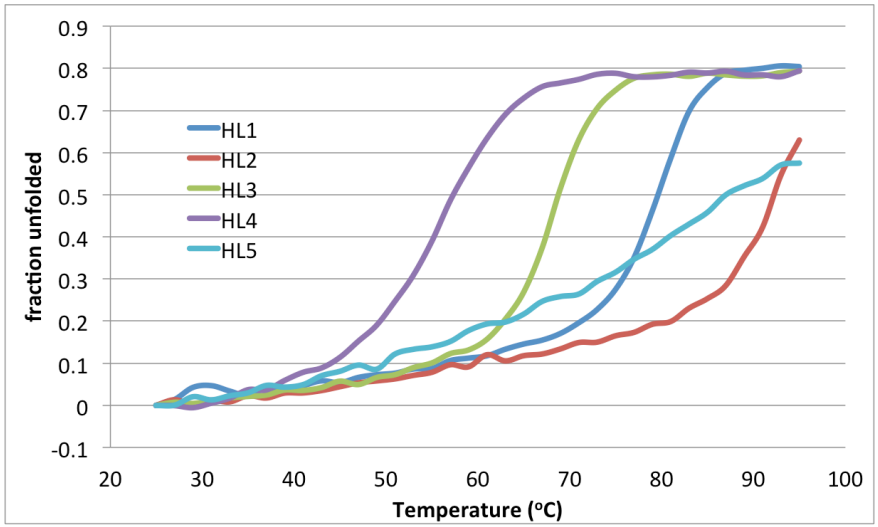


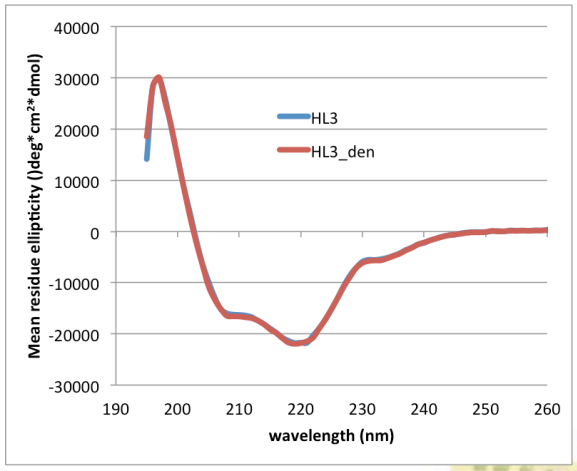




family	tested	expressed	soluble	folded	monomer
HL	9	9	9	4	2
6xx *	16	16	16	2	2
FA-B	12	10	10	5	5

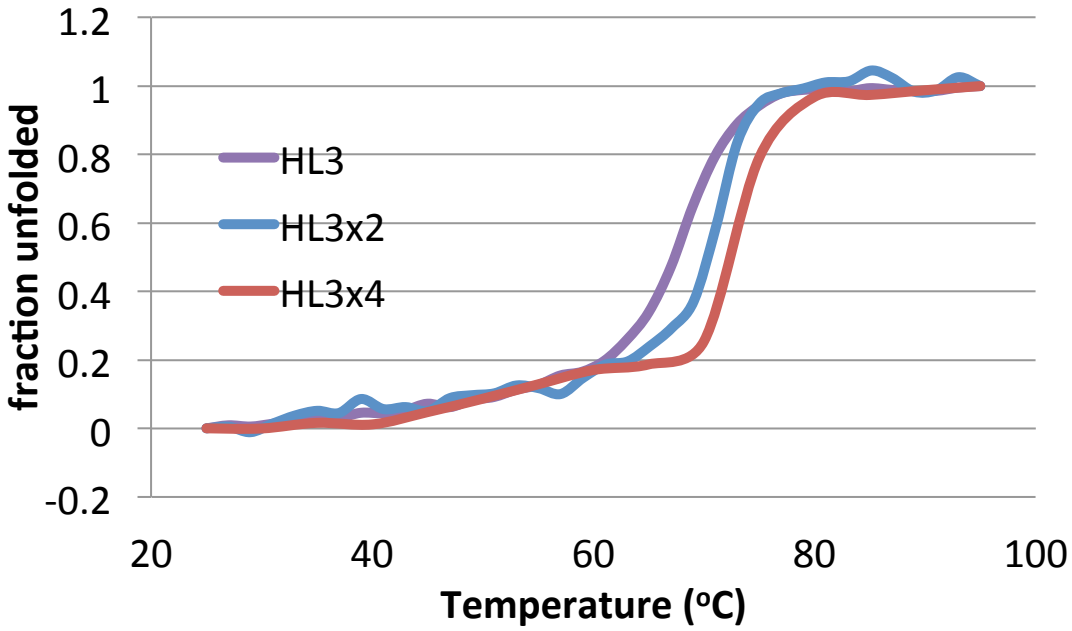
De novo design: HL

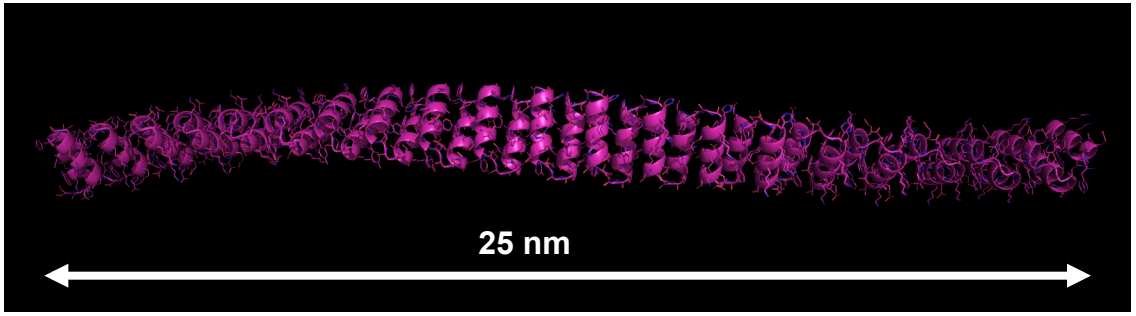


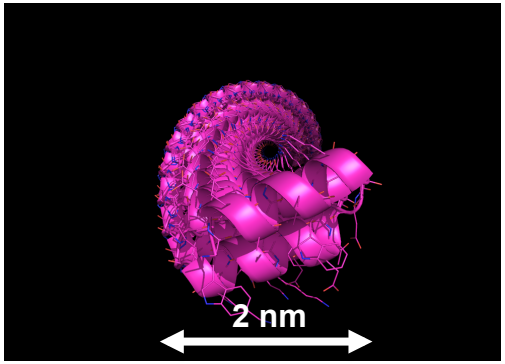


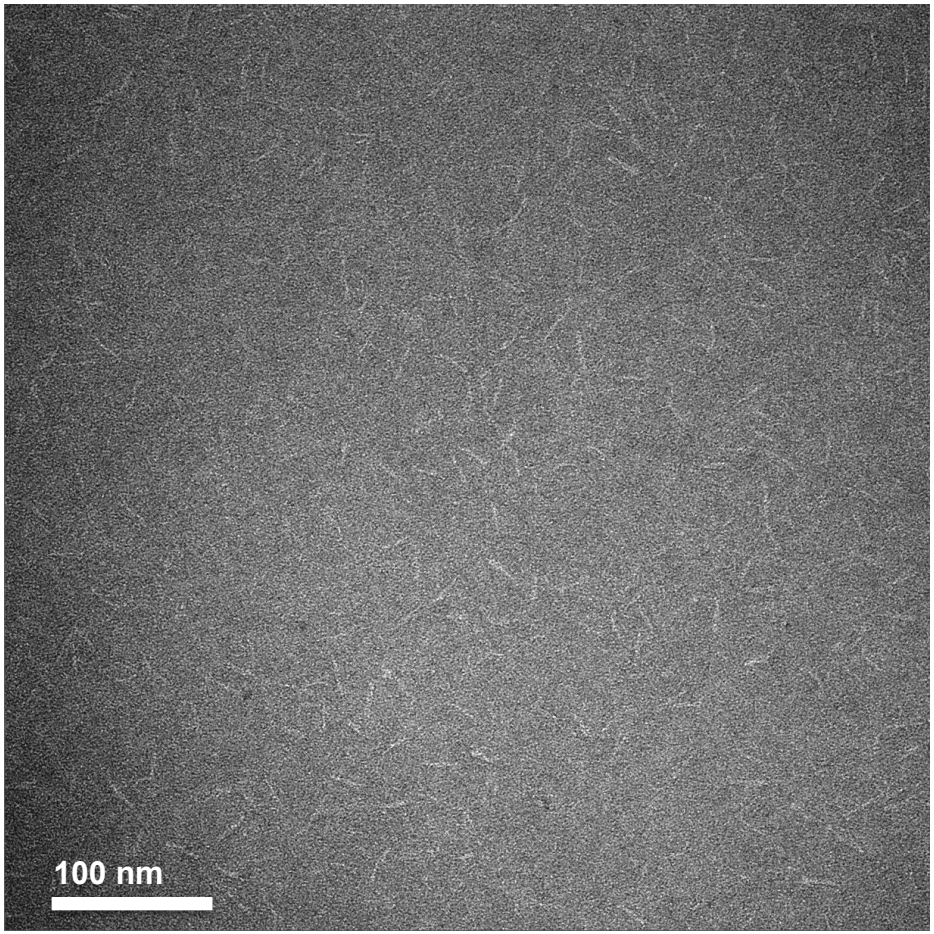
Controlling size

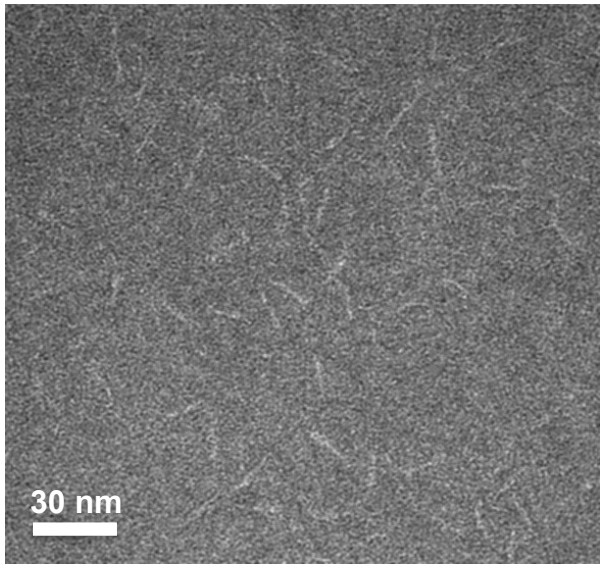
- HL3: 5 repeats
- HL3x2: 10 repeats
- HL3x4: 20 repeats





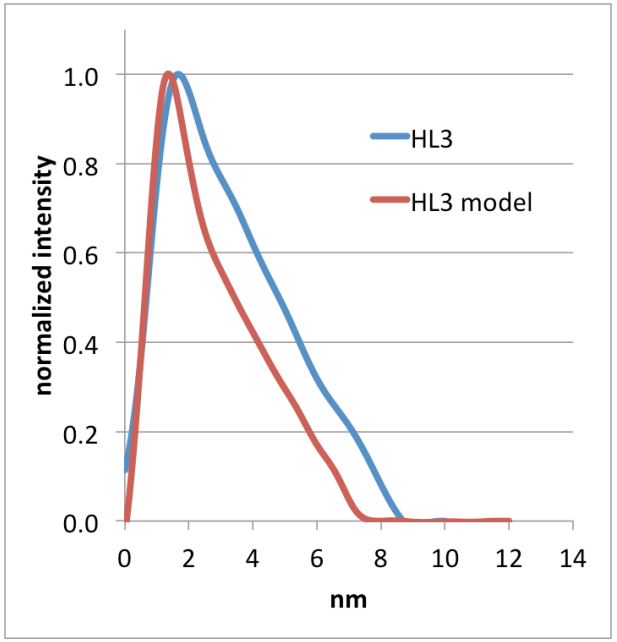


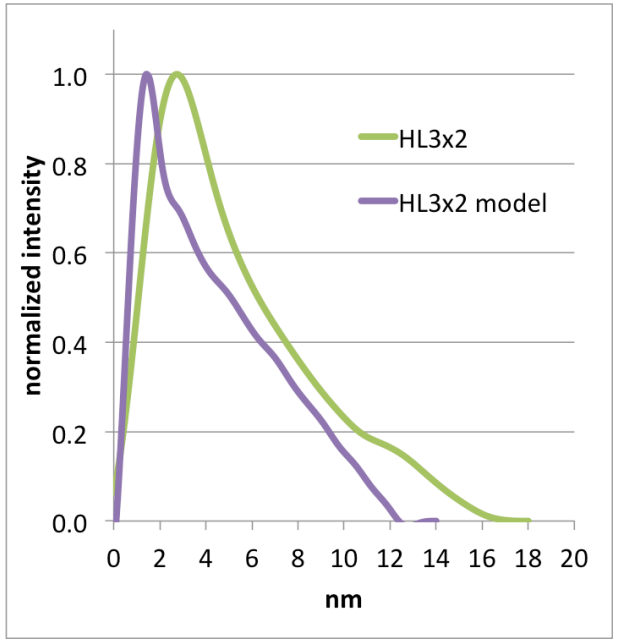


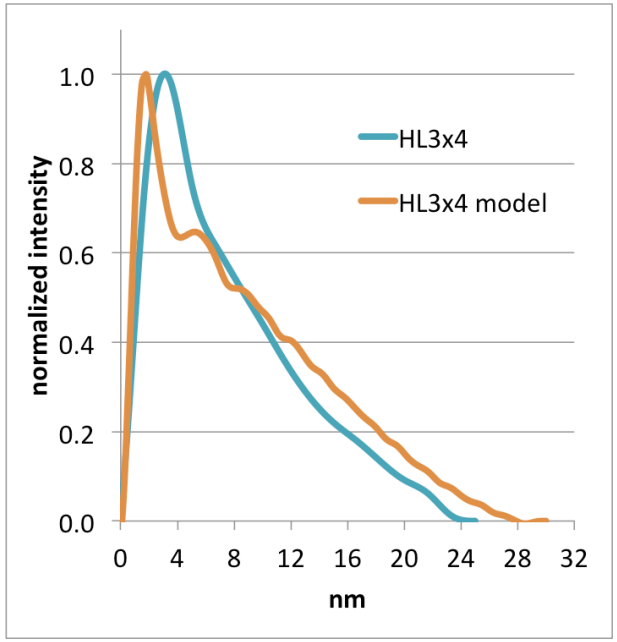


Brent Nannega

SAXS





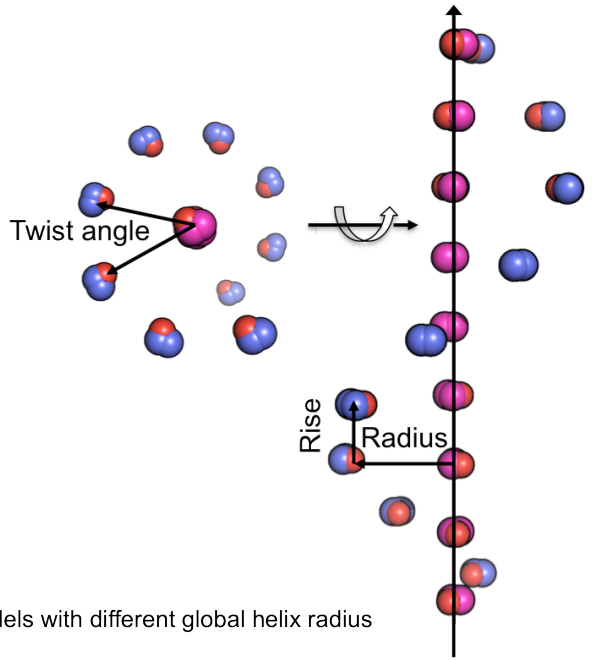


Danilo Pozzo, UW

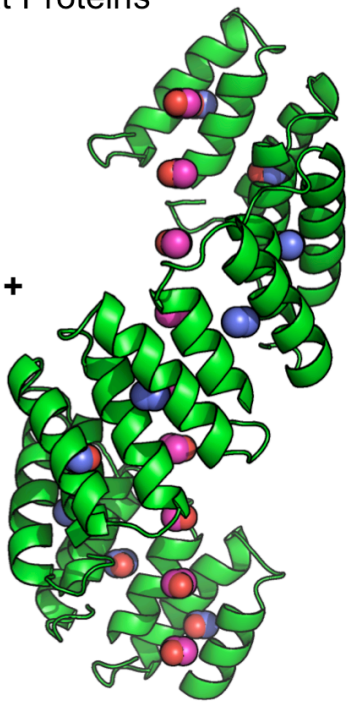
Shape control

Global super-helical parameters
Radius – Rise – Twist

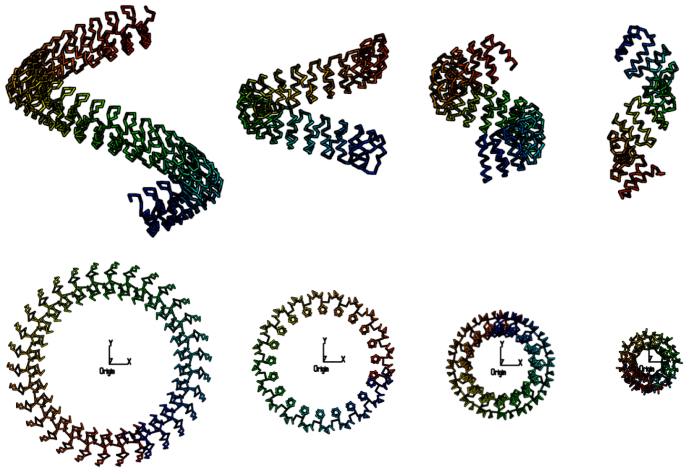
Parametric design
of Repeat Proteins



Rosetta Remodel + global helical constraint

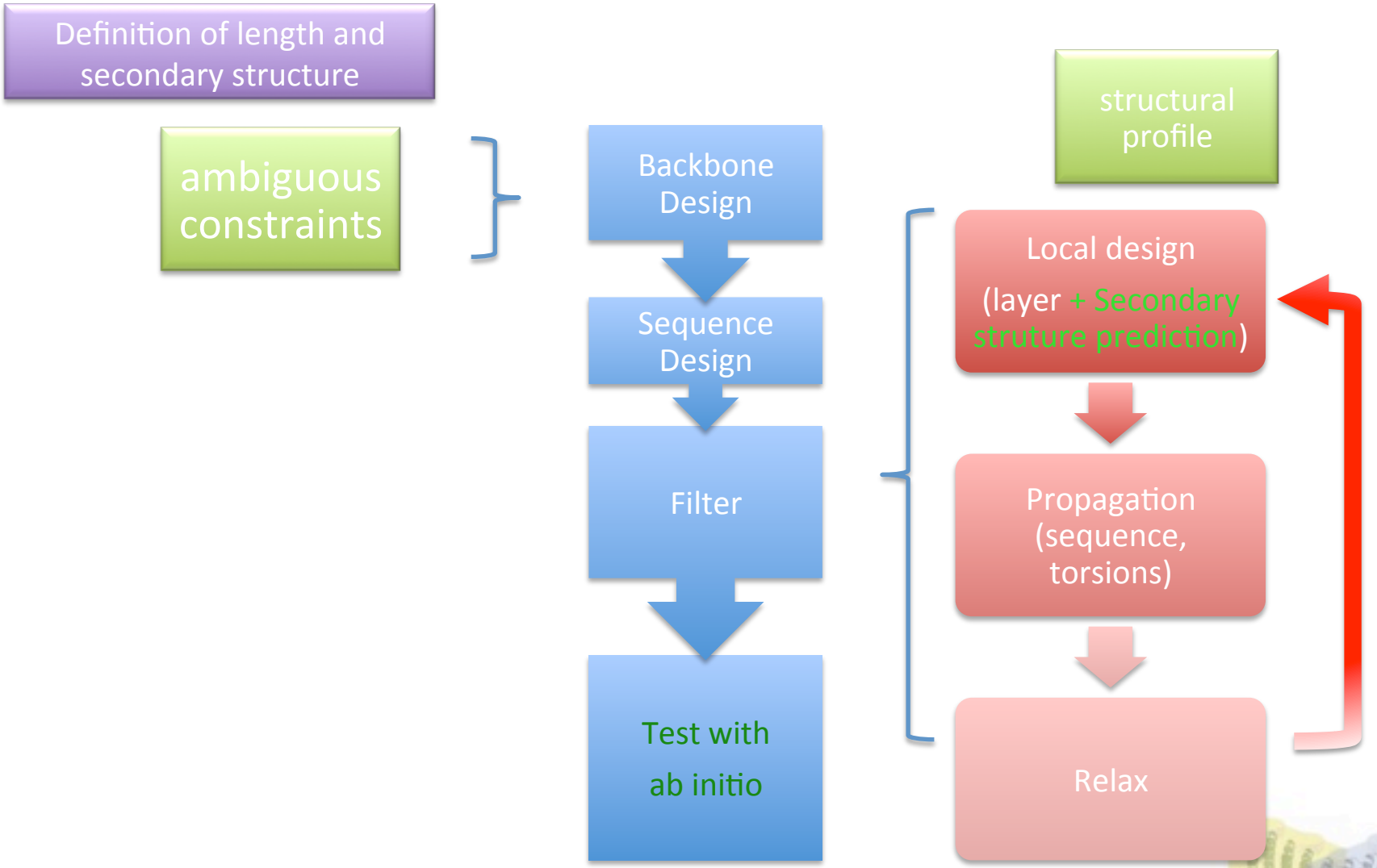


Design models with different global helix radius



Keunwan Park

Sampling topologies



Implementation in Rosetta

- Control angle in remodel
  - -helical_radius 5.0, –helical_rise 3.0, –helical_omega 1.2
- Link residues in Rosetta Scripts.
- <LinkResidues name=link_res>

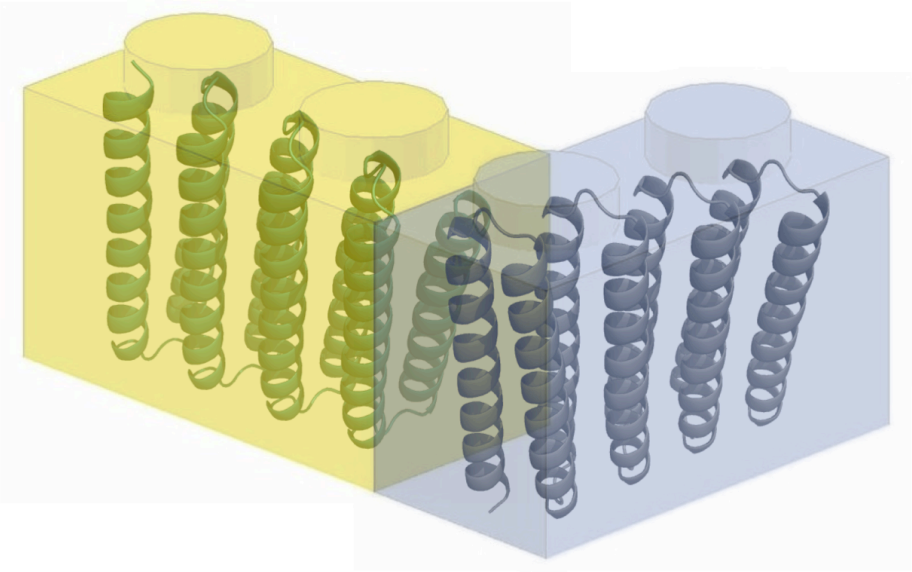
  <linkgroup group="1A-40A,41A-80A,81A-120A,121A-160A"/>

  </LinkResidues>
- Radius of gyration local energy term

  -rg_local_span 1 80
- Tune sequence to improve likelihood of abinitio success
  - Improved secondary structure match as predicted by psipred
  - Use of structural profile in packer

Summary

- Design of repeat proteins
  - Simplified design
  - Modularity
  - Stability
- Large designed proteins
- Control of shape and topology
- New scaffolds for
  - molecular recognition
  - assembly
    - Interface design
    - Fusions



Thanks

- Baker lab (UW)
  - David Baker
  - Possu Huang
  - TJ Brunette
  - Keunwan Park
  - Frank DiMaio
  - Jorge Fallas
  - Yang Hsia
  - PPF (Adam, Chris Inna, Laura, Sydney, Lara, Nikole)
- EM (Janelia Farm)
  - Tamir Gonen
  - Brent Nannega
- SAXS
  - Danilo Pozzo (UW)
- Crystallization, NESG
  - Guy Montelione (Rutgers)
  - Sergey Vorobiev (Columbia)
  - Gregory Kornhaber (Rutgers)
- Funding
  - Swiss National Science Foundation
  - Human Frontier Science Program
  - NIH
- **Rosetta Developers**