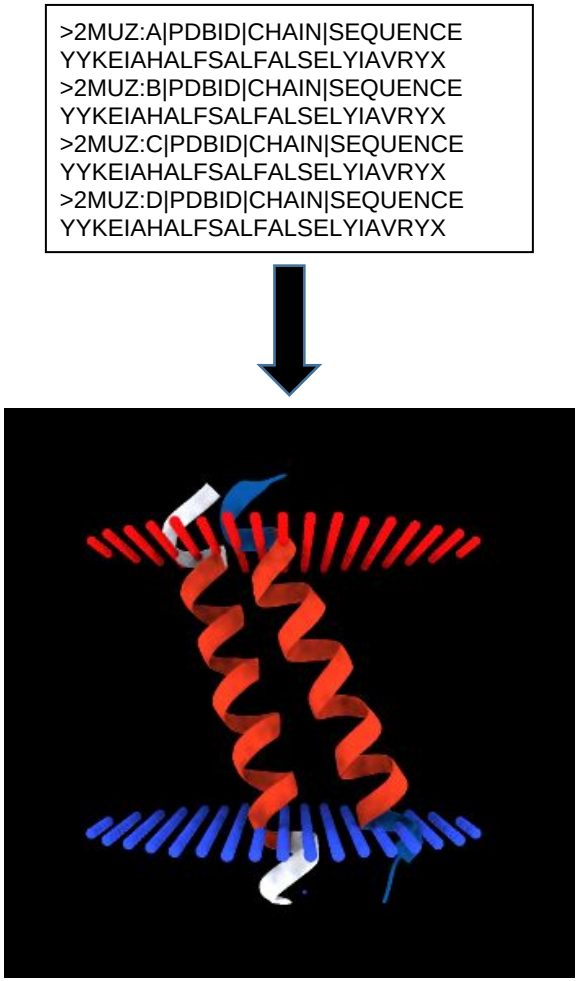
Modern Protein Design

Anika Naidu, Kevin Goncalves, Mila Schultz

CS 371 January 25, 2017

Why design proteins *de novo*?

- Push the limits of our understanding of how proteins look and work.
- Help create a toolset for engineering proteins of interest.
- Limit reliance on previous information to design a protein.

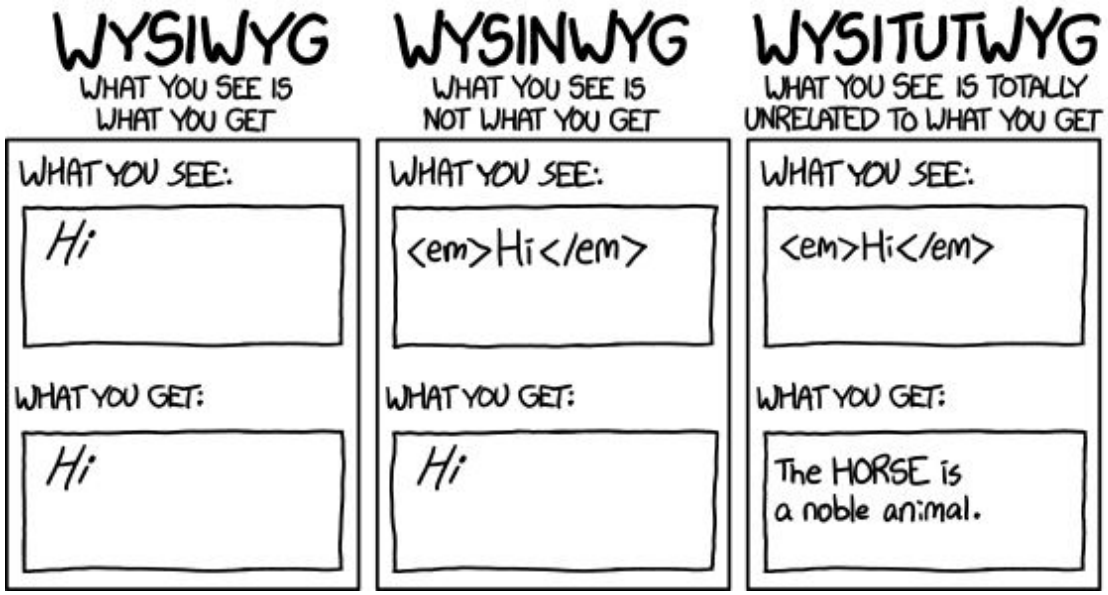


Protein Design

- Given: Desired protein function (which leads to desired approximate structure)
- Problem: Construct 1D amino acid sequence which assembles or conforms into desired 3D structure

Challenges in Protein Design

- What you make is not necessarily what you will see.
- Intelligent design may require intuition on how the protein works.
- Not much previous work on design of membrane proteins



Wikipedia

1. Accurate design of megadalton-scale two-component icosahedral protein complexes

Mila

2. De novo design of a transmembrane Zn^{2+}-transporting four-helix bundle

Kevin

3. Accurate de novo design of hyperstable constrained peptides

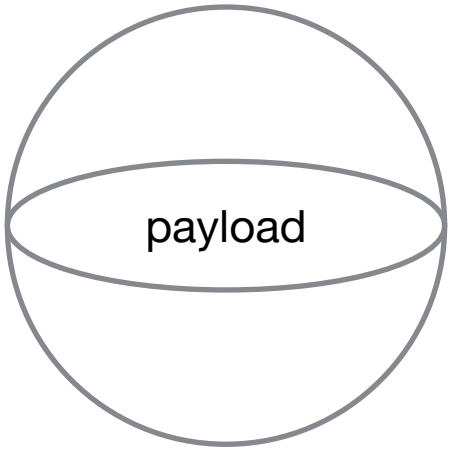
Anika

PROTEIN DESIGN

Accurate design of megadalton-scale two-component icosahedral protein complexes

Jacob B. Bale,[1,2] Shane Gonen,[1,3]* Yuxi Liu,[4]* William Sheffler,[1] Daniel Ellis,[5] Chantz Thomas,[6] Duilio Cascio,[4,7,8] Todd O. Yeates,[4,7] Tamir Gonen,[3] Neil P. King,[1,5]† David Baker[1,5,9]†

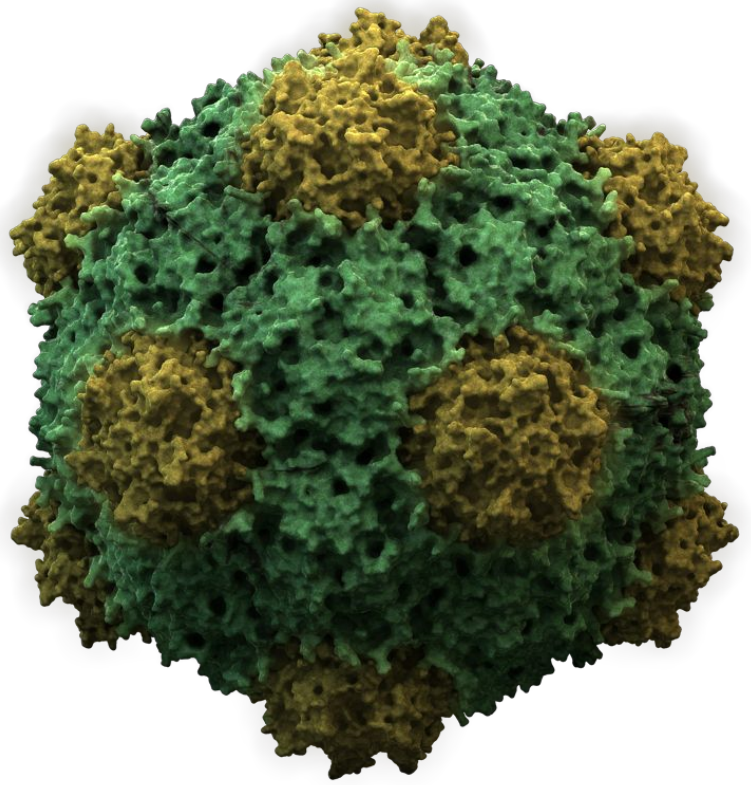
Goal: Design hollow sphere



- Authors want to build a huge protein complex
- Unlike traditional protein design, which focuses on designing individual proteins
- Not trivial

Large-scale complexes

- Inspired by self-assembling proteins that occur in nature
  - Cages
  - Capsids
- Scale:
  - Multiple nanometers

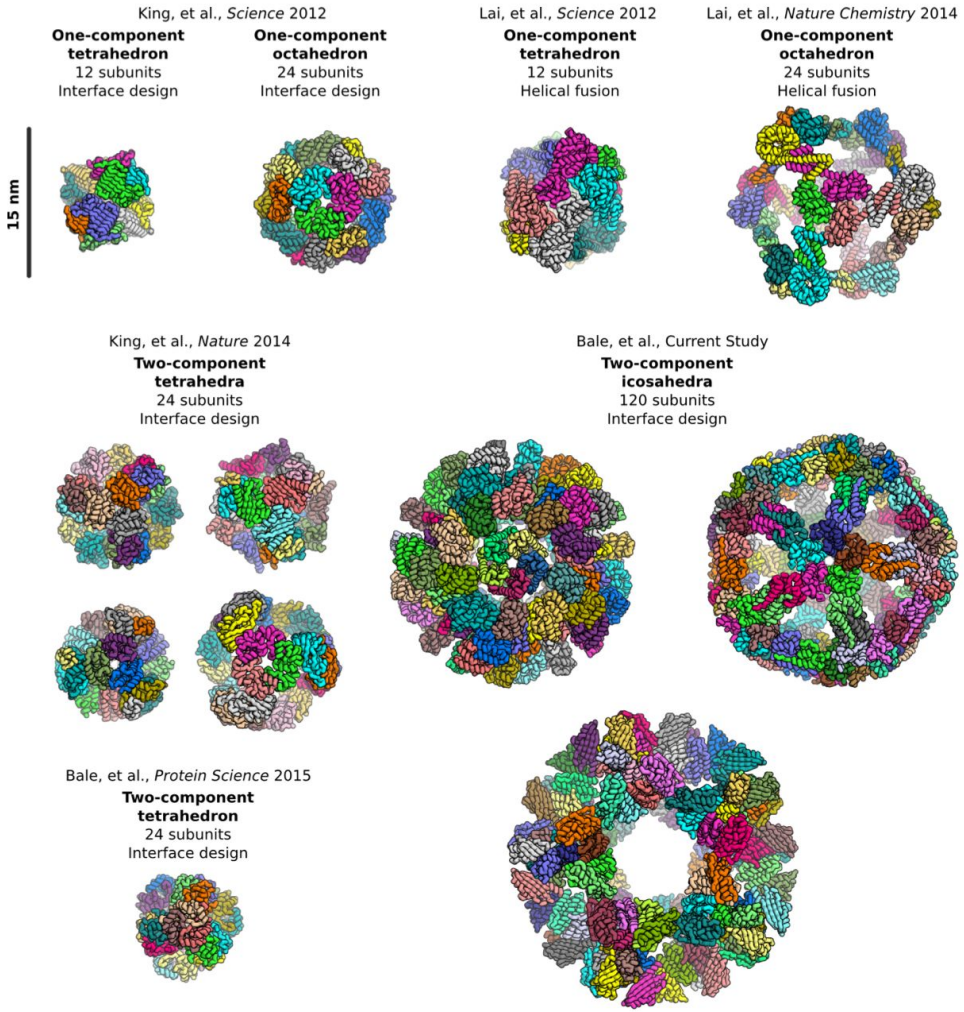


Cowpea Mosaic Virus

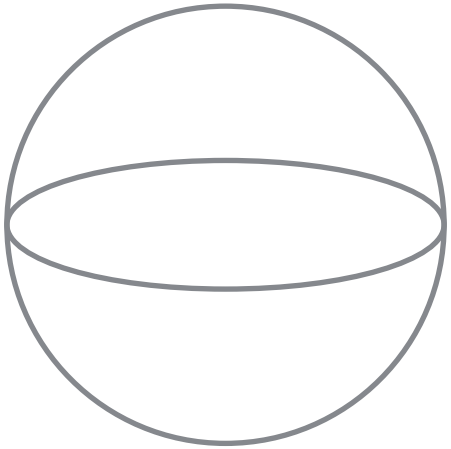
Structure of the icosahedral Cowpea mosaic virus (CPMV) based on PDB ID 2BFU, by Thomas Splettstoesser (www.scistyle.com), licensed with CC 3.0

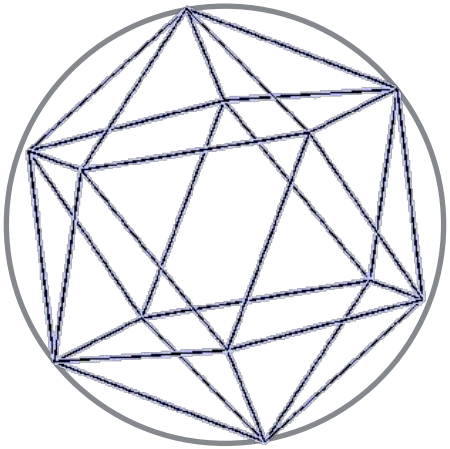
Why build these machines?

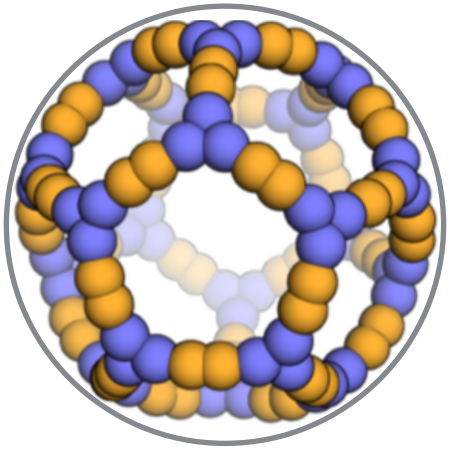
- Delivery mechanism for:
  - Vaccines
  - Drugs
  - Fluorescence
  - And more…

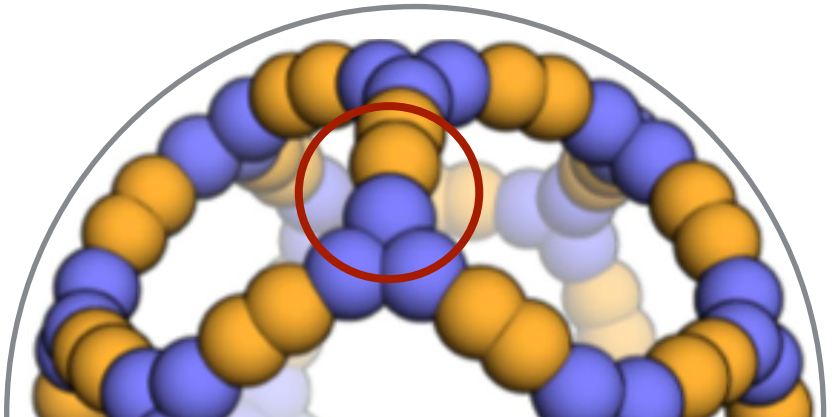


Combine multiple components into a symmetrical 3D volume





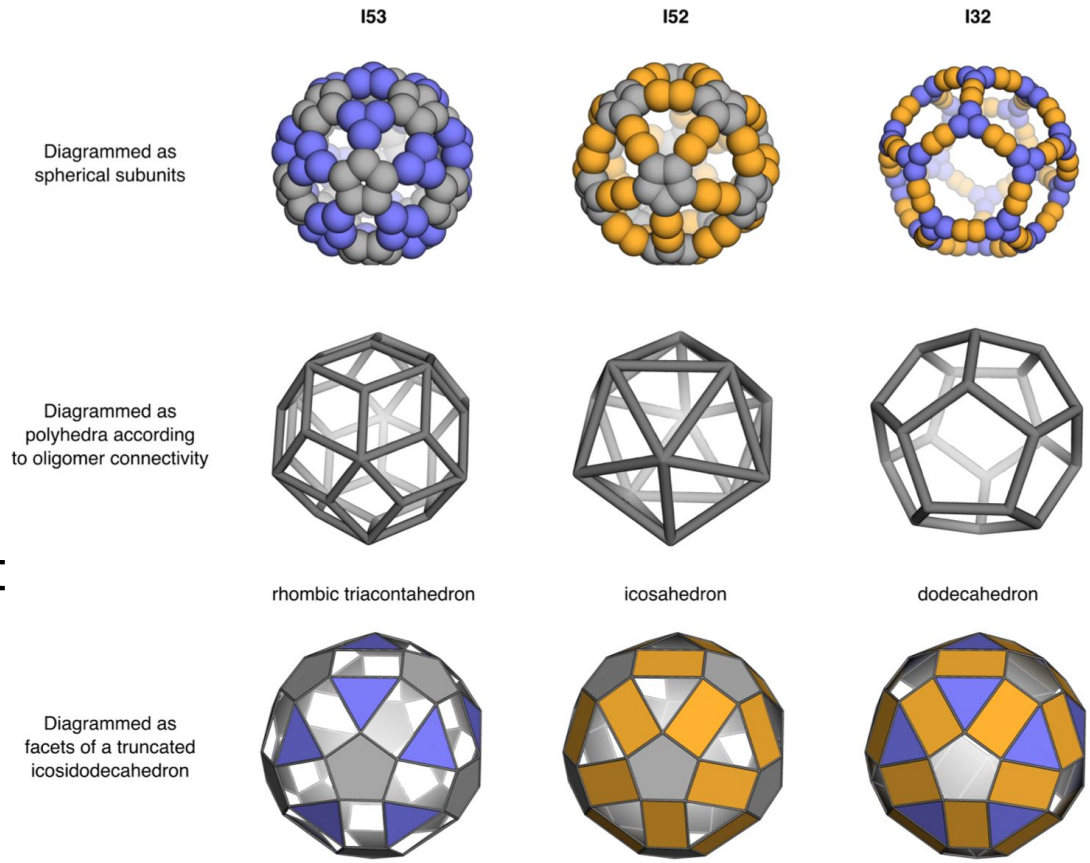




https://www.math.nmsu.edu/~pmorandi/math112s00/Icosahedron.html

Icosahedral Symmetry

- Previous work used tetrahedra and octahedra
- Chosen for largest relative interior volume, high symmetry
- Rotational axes of symmetry call for different subunit shapes
  - dimeric, trimeric, pentameric



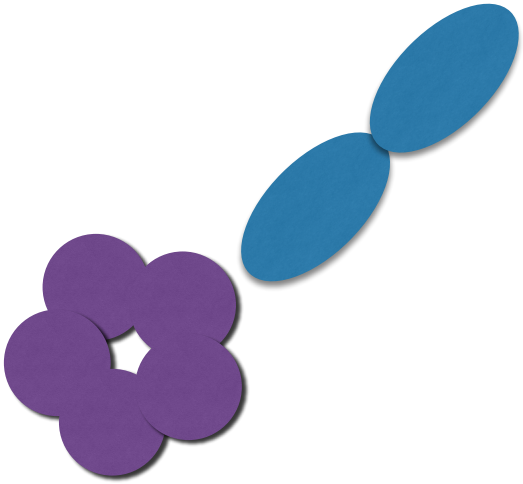
Video: 3D Printed Model

https://youtu.be/QtddFlkQNmc?t=1m44s

Pick Building Blocks

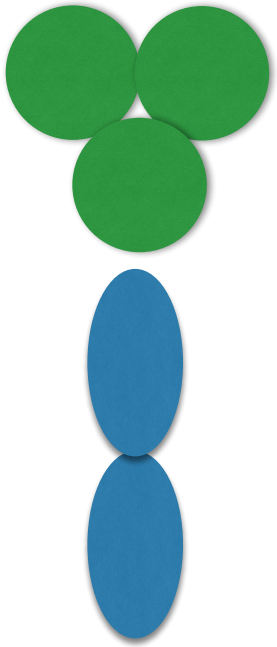
- Look for homooligomeric structures in PDB (complex of several identical proteins)

- Use *de novo* homooligomeric structures



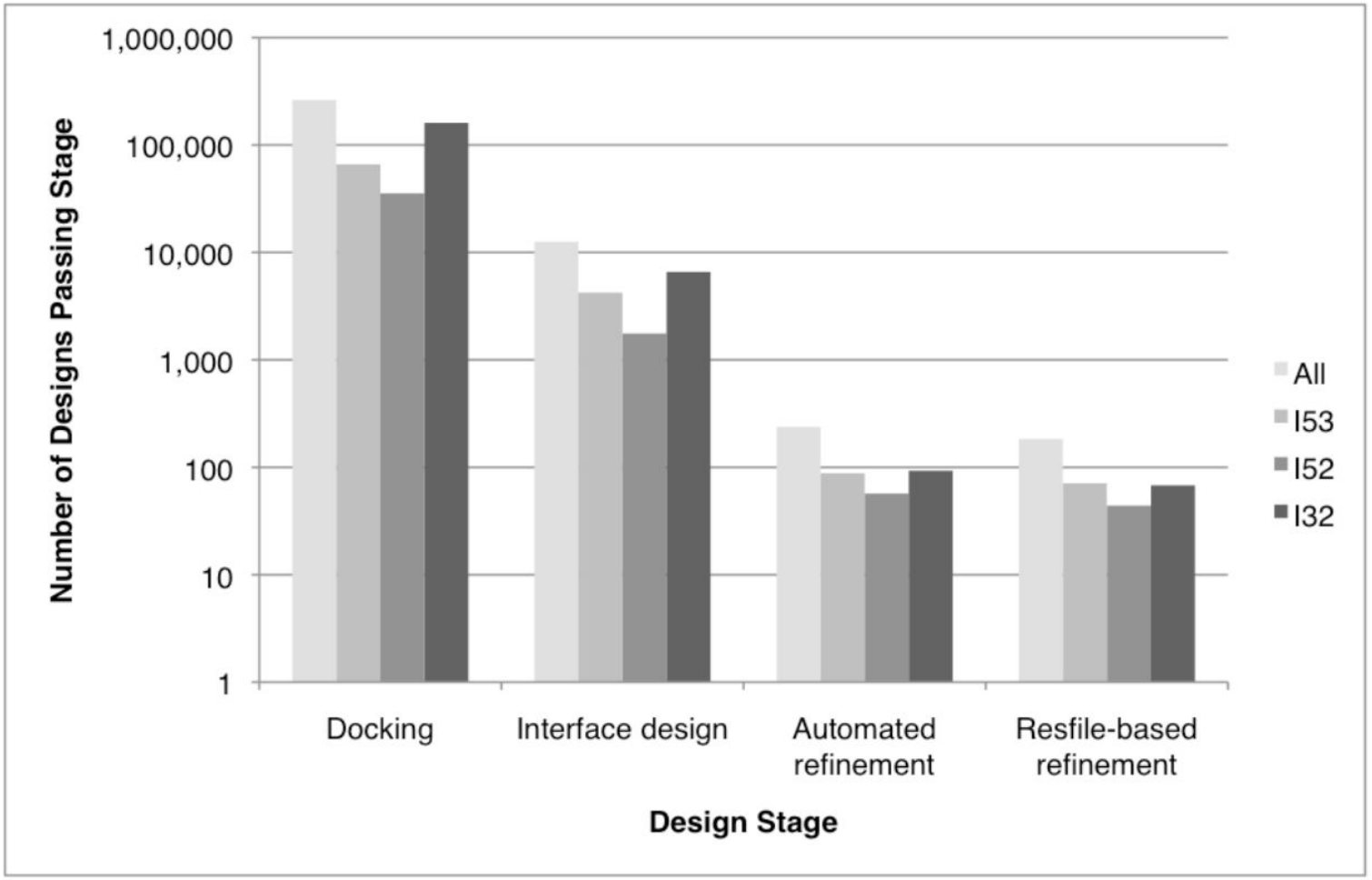
50,400 pairs

14,400 pairs



276,150 pairs

Stages of Design Filtering



Filtering the Models

1. Symmetric Docking: Arrange the building blocks into favorable configurations

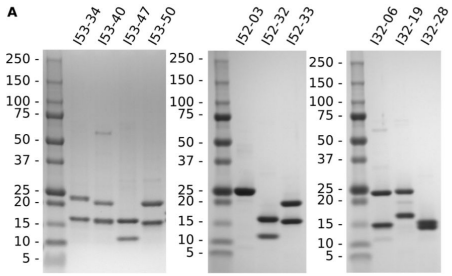
2. Interface Design: Pick rotamers (side chains) to optimize interface between building blocks

3. Automated Reversion: Revert unnecessary changes to building blocks

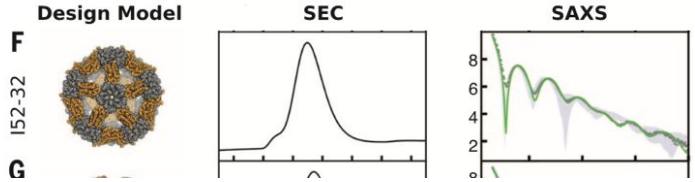
4. Resfile-Based Refinement: Visual inspection of the models (I53: 71, I52: 44, I32: 68)

Structures Confirmed Experimentally

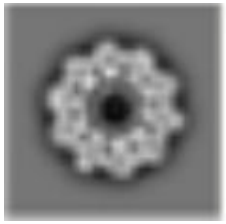
- Cloned in E. Coli
- Electrophoresis



- Size-Exclusion Chromatography

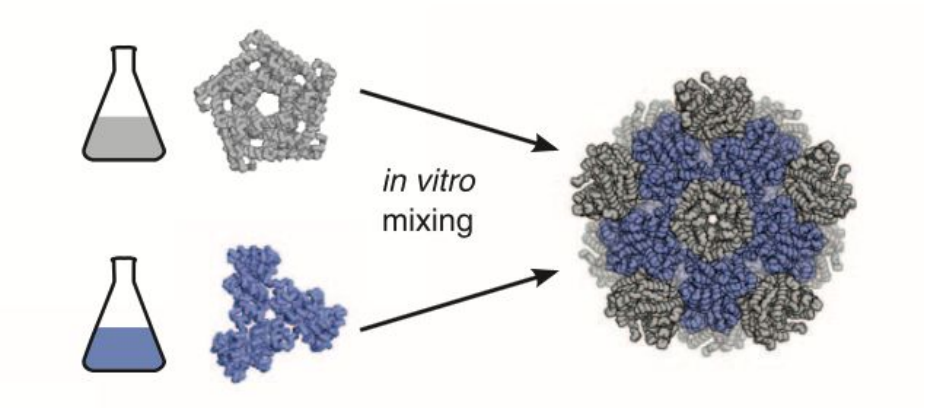


- Small-Angle Light Scattering
- Electron Microscopy (8)



- X-Ray Crystallography (3)

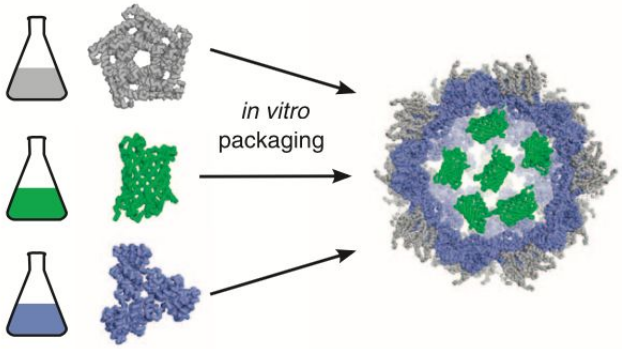
In Vitro Construction



- I53-50 variant
- Assembly was ~halfway complete within 1-10 min depending on the concentrations
- Similar timeframe to viral capsids
- Controlled assembly with solvent concentration

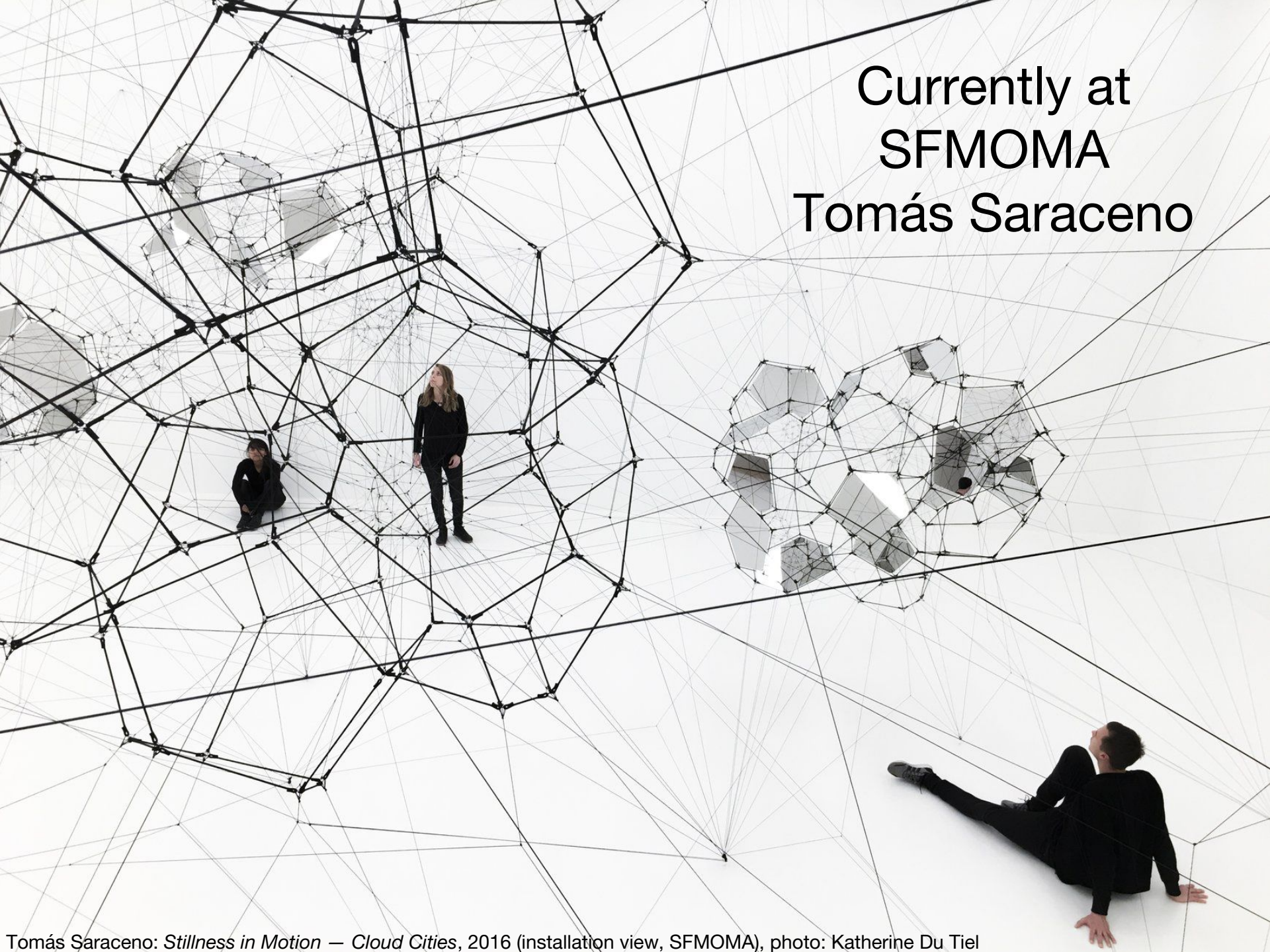
Application: Packaging Green Fluorescent Protein (GFP)

- Mutate I53-50 to add positively-charged residues on the inside of the building blocks
- Mixed with “supercharged GFP”
- 7-11 GFPs are packaged per icosahedral assembly
- Precise GFP luminance



Caveats & Further Work

- There was clearly hand-tuning in the filters of the different levels
- Very few successful designs
- Each design had to be identified by hand
- Should try to store an active payload
- No 120-unit complexes with strict icosahedral symmetry that assemble in this way have been identified in nature so far

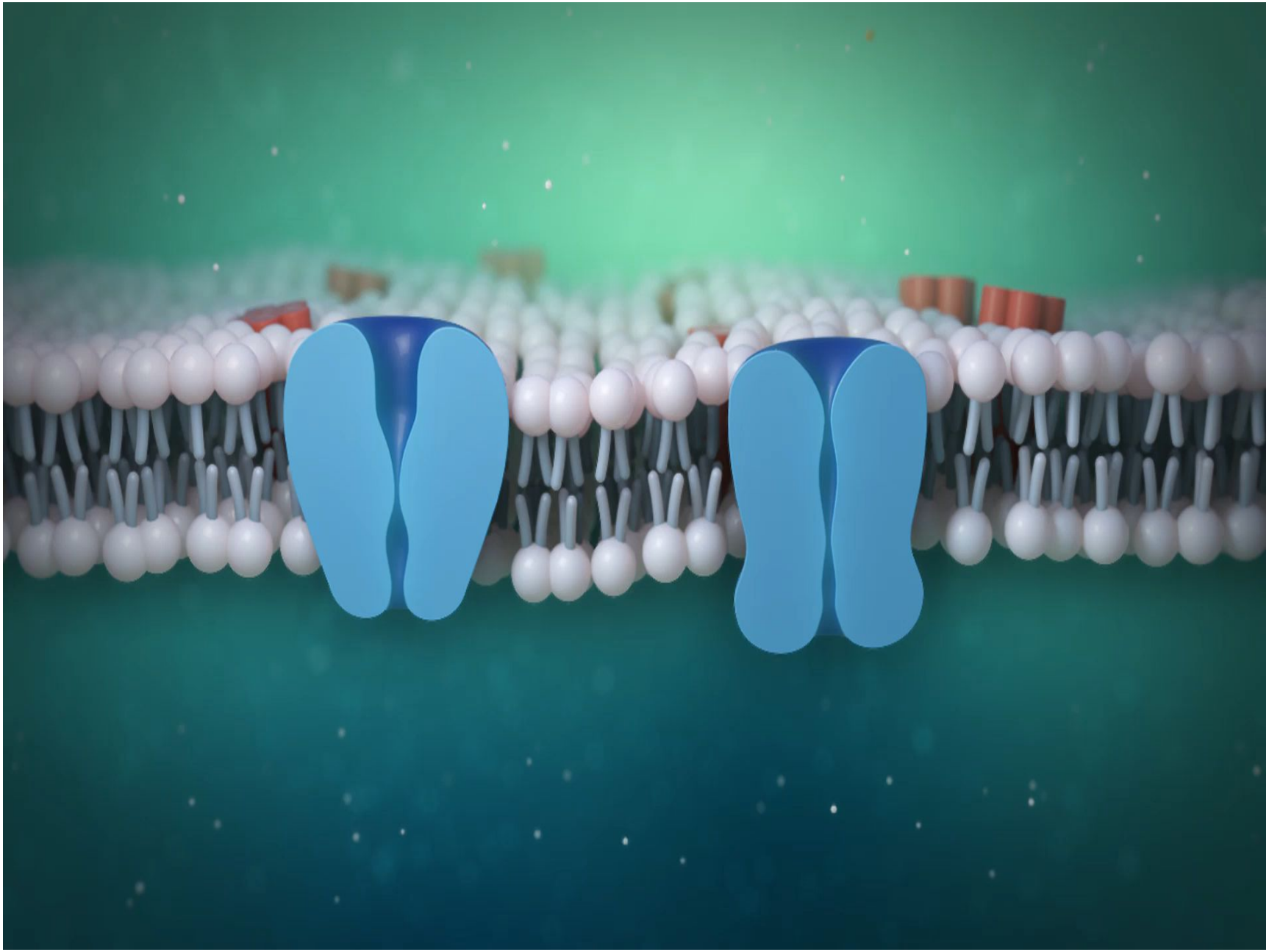


Tomás Saraceno: *Stillness in Motion — Cloud Cities*, 2016 (installation view, SFMOMA), photo: Katherine Du Tiel

Unless otherwise cited, figures and images are from:
Accurate design of megadalton-scale two-component icosahedral protein complexes
BY JACOB B. BALE, SHANE GONEN, YUXI LIU, WILLIAM SHEFFLER, DANIEL ELLIS, CHANTZ THOMAS, DUILIO CASCIO, TODD O. YEATES, TAMIR GONEN, NEIL P. KING, DAVID BAKER
SCIENCE 22 JUL 2016 : 389-394

De Novo Design of a Transmembrane Zn^{2+}-Transporting Four-Helix Bundle

CS 371 Presentation by Kevin Goncalves

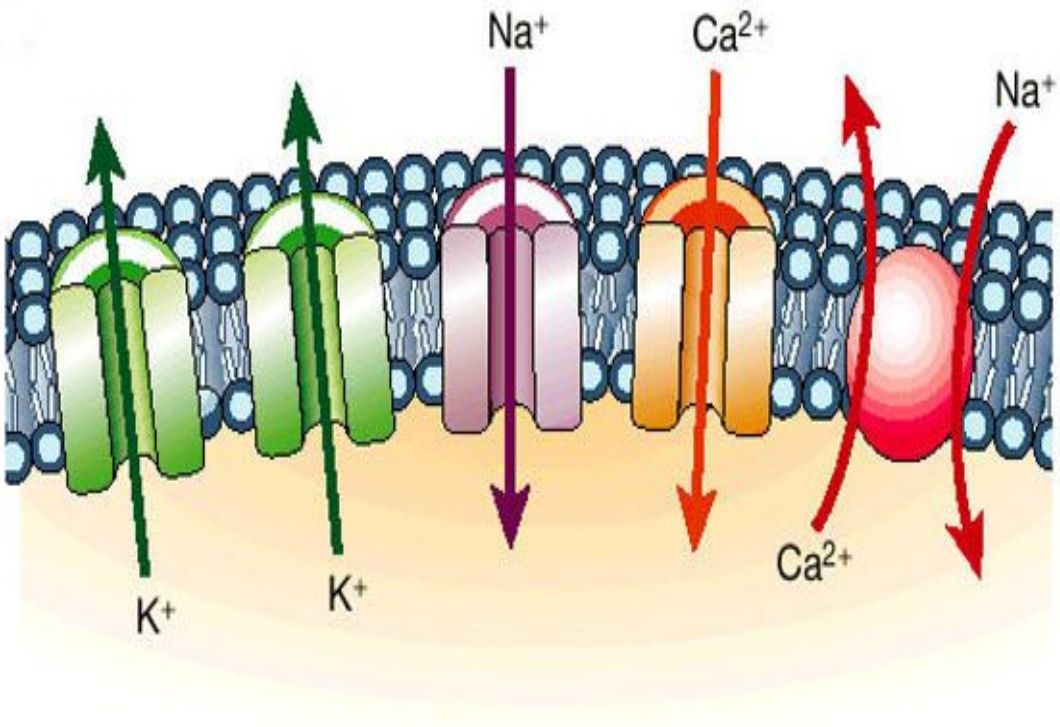


https://vimeo.com/167920911

Credit: Dennis Wei on Vimeo

Significance of Designing Ion Channels and Transporters

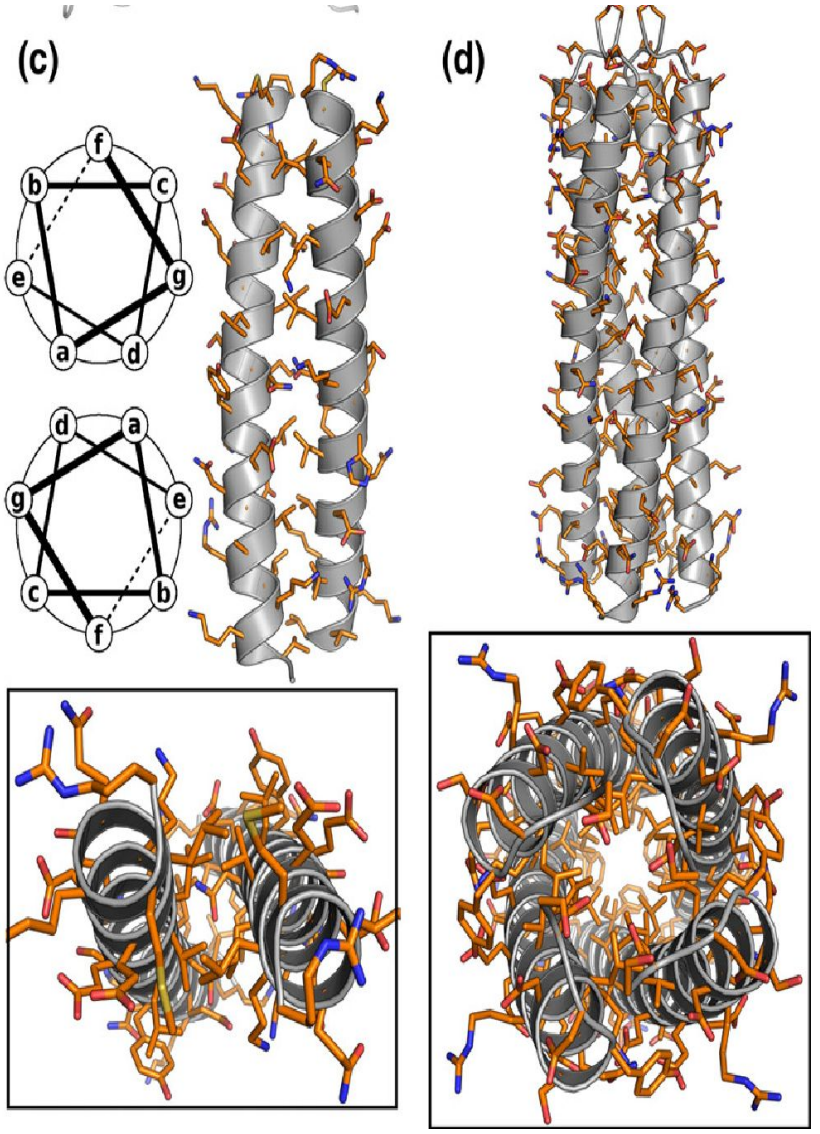
- Ion channels and transporter determines what goes through the membrane:
  - Must be selective
  - Can couple gradients for transport



Rasband et al. Nature Education 2010

Previous Work by Degrado et al.

- Previously designed helical bundles
- Crystallized structures reveal structural features for designing coiled-coils.



Grigoryan et at. J. Mol. Bio. 2011

PROTEIN DESIGN

De novo design of a transmembrane Zn^{2+}-transporting four-helix bundle

Nathan H. Joh,[1] Tuo Wang,[2] Manasi P. Bhate,[1] Rudresh Acharya,[3] Yibing Wu,[1] Michael Grabe,[1*] Mei Hong,[2*] Gevorg Grigoryan,[4*] William F. DeGrado[1*]

The design of functional membrane proteins from first principles represents a grand challenge in chemistry and structural biology. Here, we report the design of a membrane-spanning, four-helical bundle that transports first-row transition metal ions Zn^{2+} and Co^{2+}, but not Ca^{2+}, across membranes. The conduction path was designed to contain two di-metal binding sites that bind with negative cooperativity. X-ray crystallography and solid-state and solution nuclear magnetic resonance indicate that the overall helical bundle is formed from two tightly interacting pairs of helices, which form individual domains that interact weakly along a more dynamic interface. Vesicle flux experiments show that as Zn^{2+} ions diffuse down their concentration gradients, protons are antiported. These experiments illustrate the feasibility of designing membrane proteins with predefined structural and dynamic properties.

Uniting Intuition of Dynamics with Design

Design Zn^{2+}-Transporting Four-Helix Bundle (Rocker).

Use Molecular Dynamics Simulations to probe Rocker's structure.

Experimentally define structure and dynamics of Rocker.

Uniting Intuition of Dynamics with Design

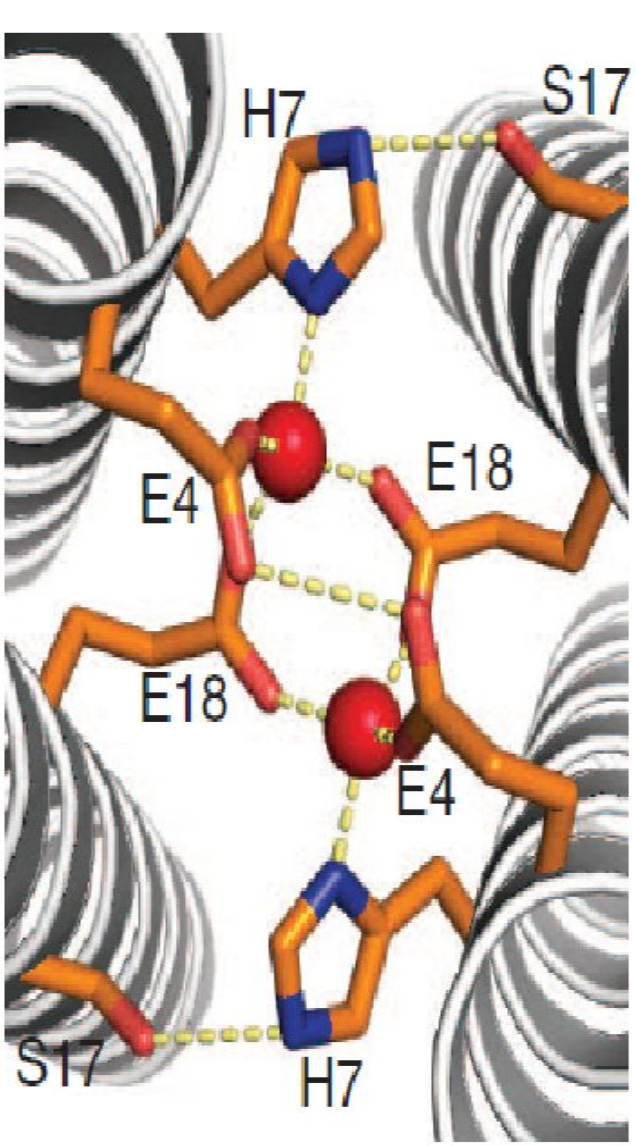
Design Zn^{2+}-Transporting Four-Helix Bundle (Rocker).

Use Molecular Dynamics Simulations to probe Rocker's structure.

Experimentally define structure and dynamics of Rocker.

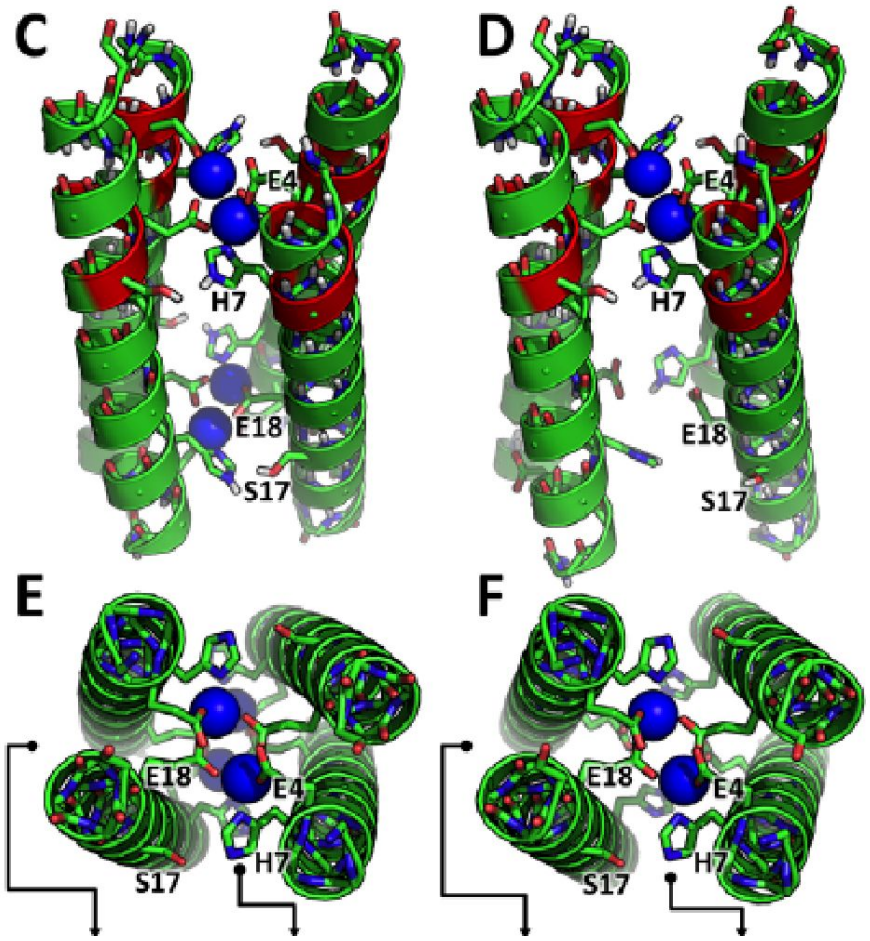
Rocker Design Stra

- Introduce new structural features
  - Two Zn^{2+} binding sites 2His4Glu
  - Backbone conformation optimized for Zn^{2+} binding
- Resulting design is very different from YiiP, the only natural Zn^{2+} transporter with high-res structure.

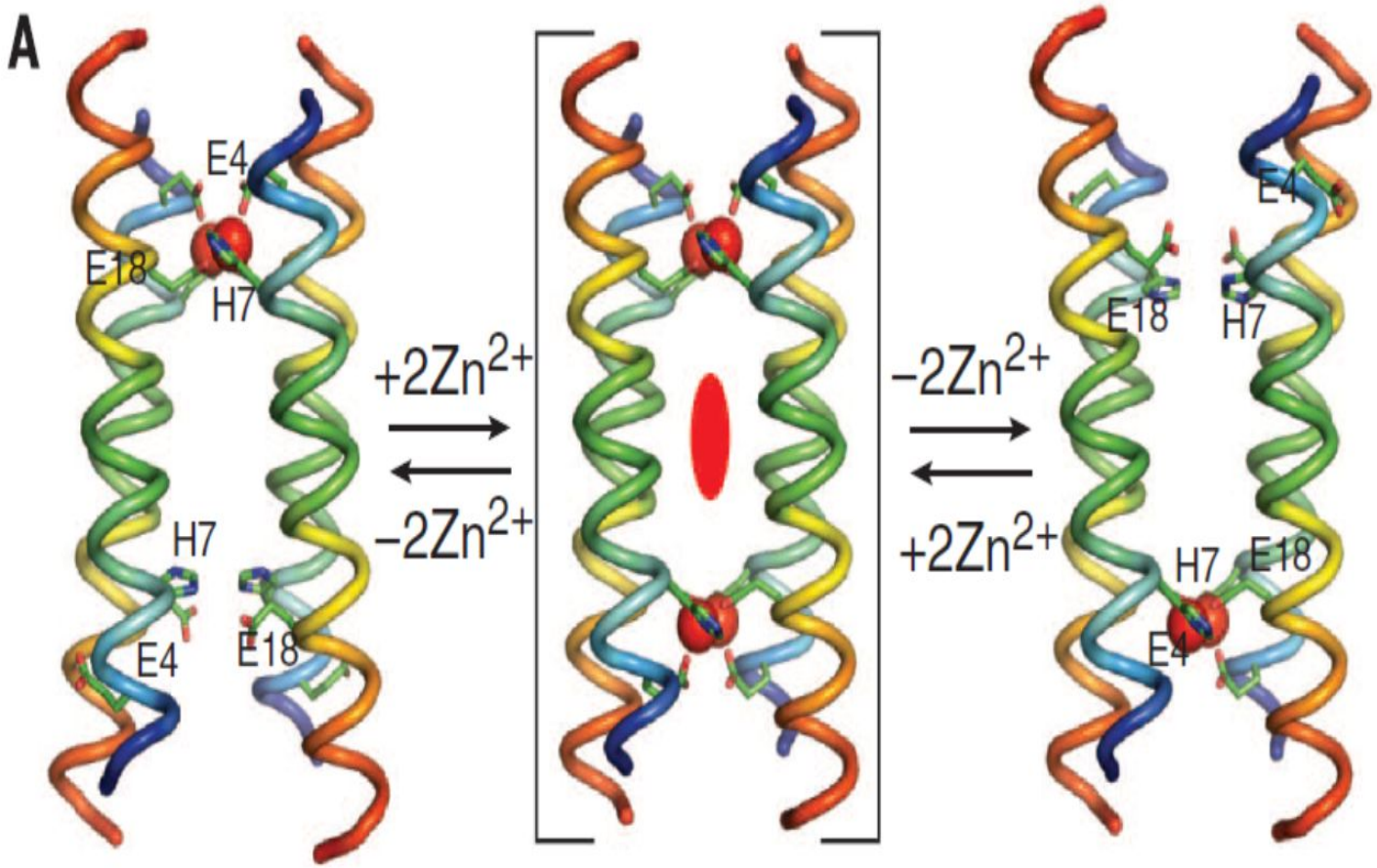


Rocker Design Strategy

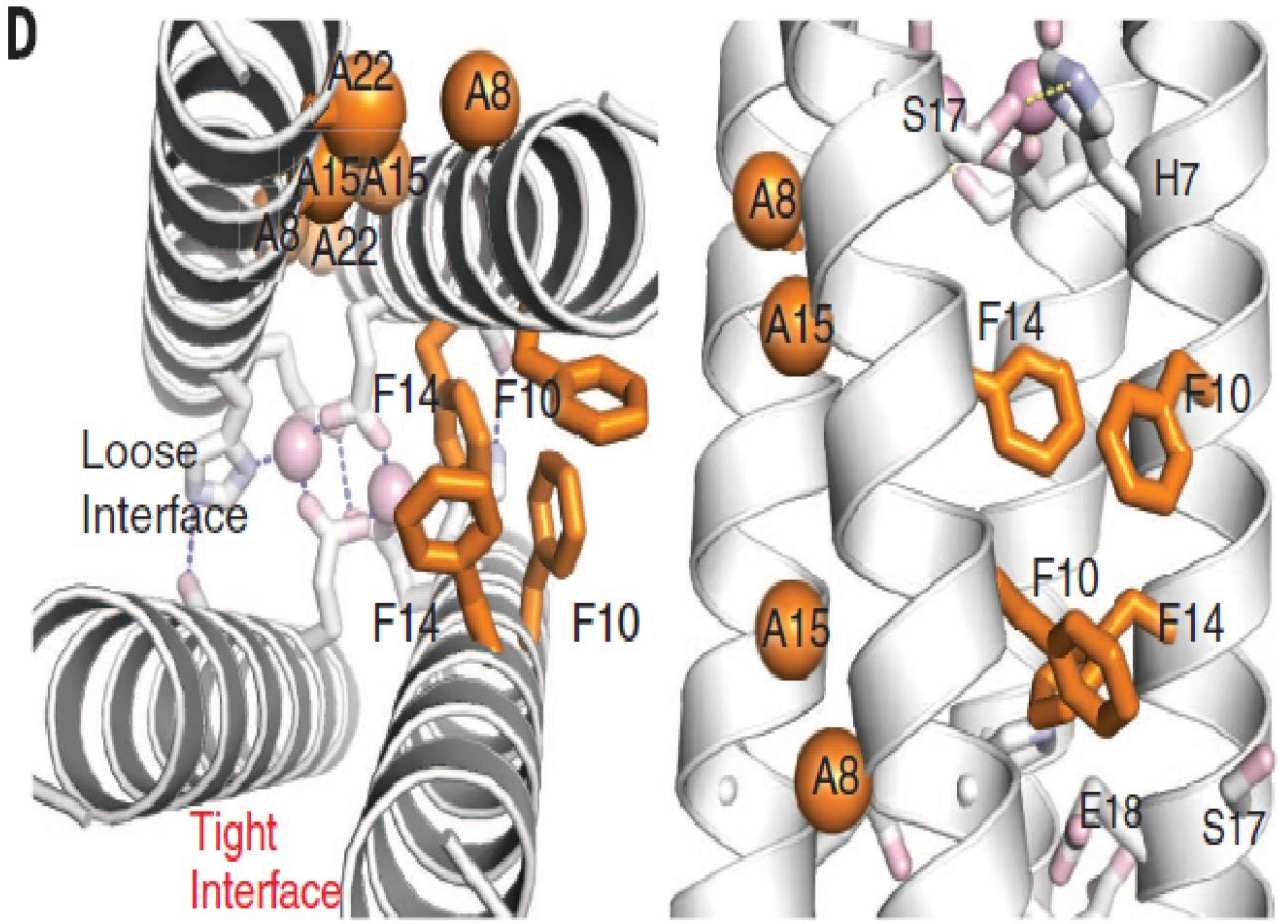
- First created a stable tetramer and then introduced asymmetric qualities
- Negative design concept:
  - Introduces interactions that destabilize symmetric conformations.



Proposed Dynamic Mechanism



Final Design



- VALOCIDY: estimated free energy of ~1000 asymmetric sequences
- Final sequence chosen:
  - Largest VALOCIDY free-energy difference
  - Pore large enough for ion to pass through.

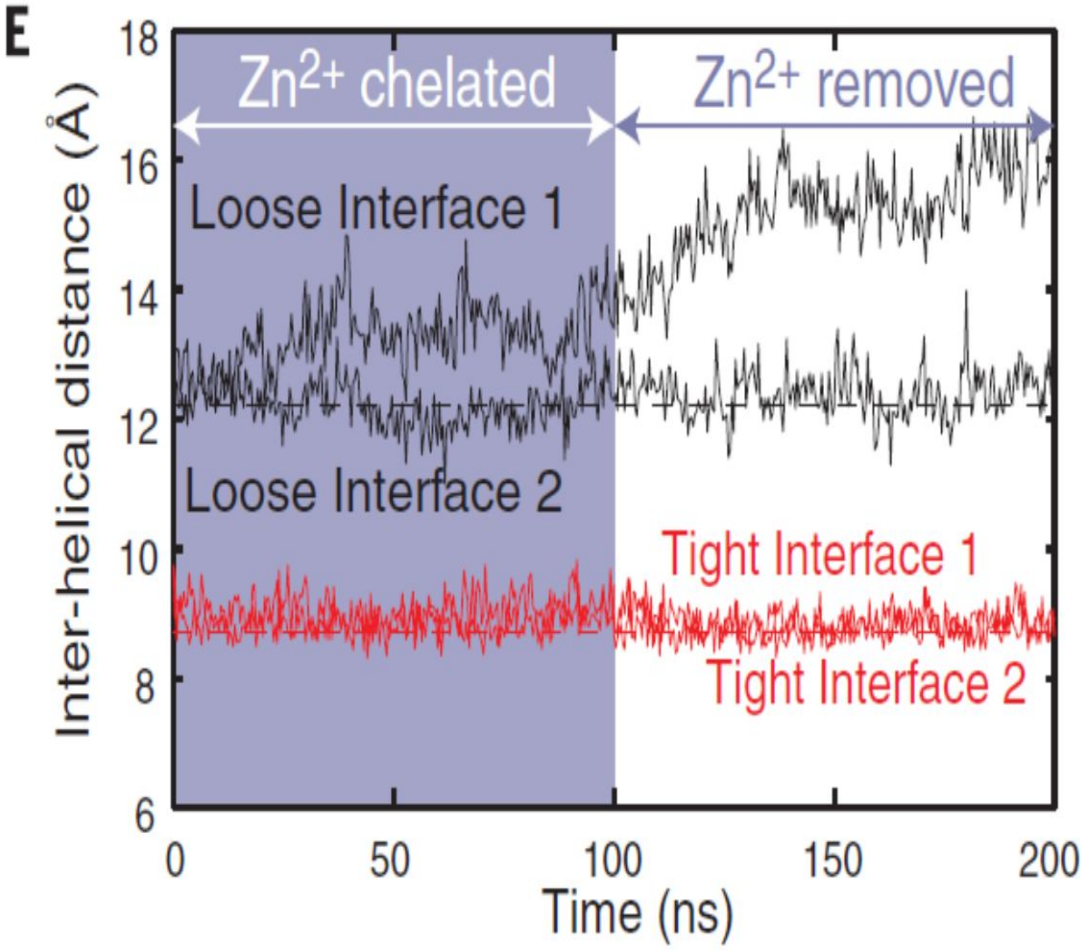
Uniting Intuition of Dynamics with Design

Design Zn^{2+}-Transporting Four-Helix Bundle (Rocker).

Use Molecular Dynamics Simulations to probe Rocker's structure.

Experimentally define structure and dynamics of Rocker.

Simulating Interhelical Distances



- Confirmed tight vs loose interfaces via simulation
- Loose interface interhelical distance changes by 3 Å, in some cases
  - Longest distance when Zn^{2+} not bound.
- Need longer scale for full transport cycle simulation

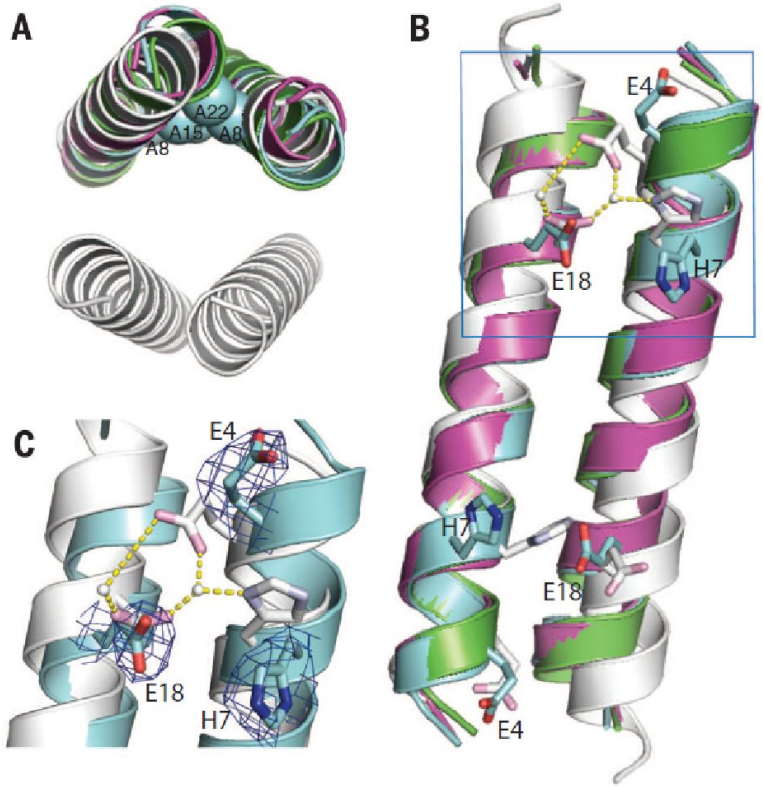
Uniting Intuition of Dynamics with Design

Design Zn^{2+}-Transporting Four-Helix Bundle (Rocker).

Use Molecular Dynamics Simulations to probe Rocker's structure.

Experimentally define structure and dynamics of Rocker.

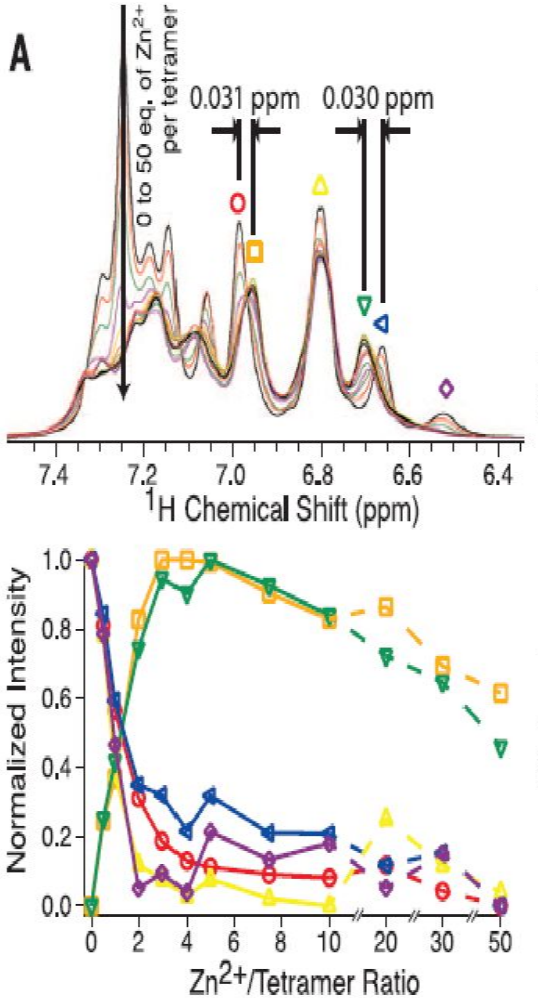
Experimental Validation of Design



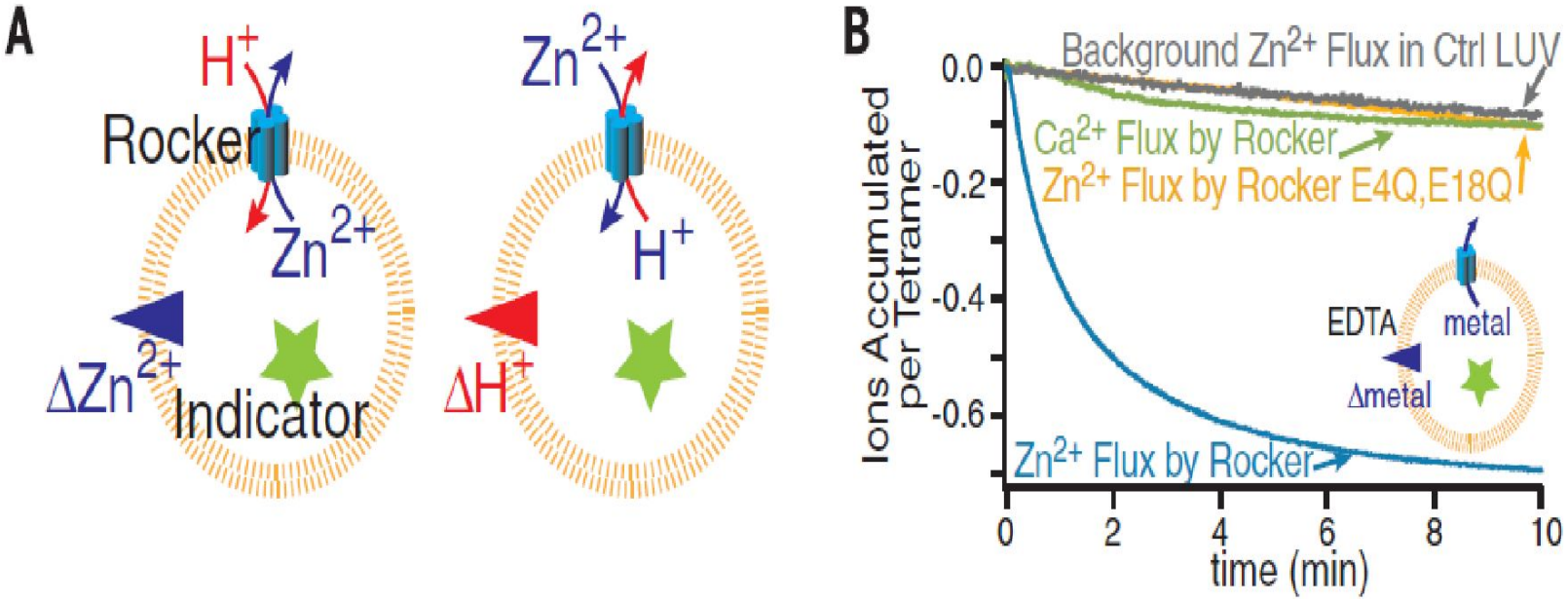
- Analytical Ultracentrifugation (AUC)
  - Confirmed that Rocker tetramerizes in the presence of Zn^{2+}
- X-ray Crystallization
  - Structures solved for dimer without Zn^{2+} bound
  - up to 2.7 Å resolution
- Dimer conformations nearly identical

NMR Supports Rocker Tetrameric Form

- Peak shifts in ^{1}H NMR structures, suggests that side chains shift when Zn^{2+} is added to the solution.
- Moreover, Zn^{2+} shifts only up to 2:1 ratio of Zinc to tetramer.
  - Suggests that tetramer formation is dependent on Zn^{2+}
- Limitation: NMR is a coarse way of defining Rocker Structure.



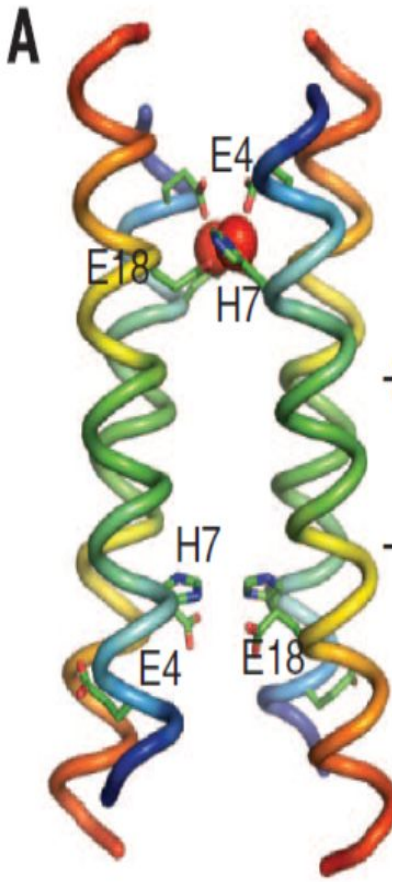
Kinetics of Rocker



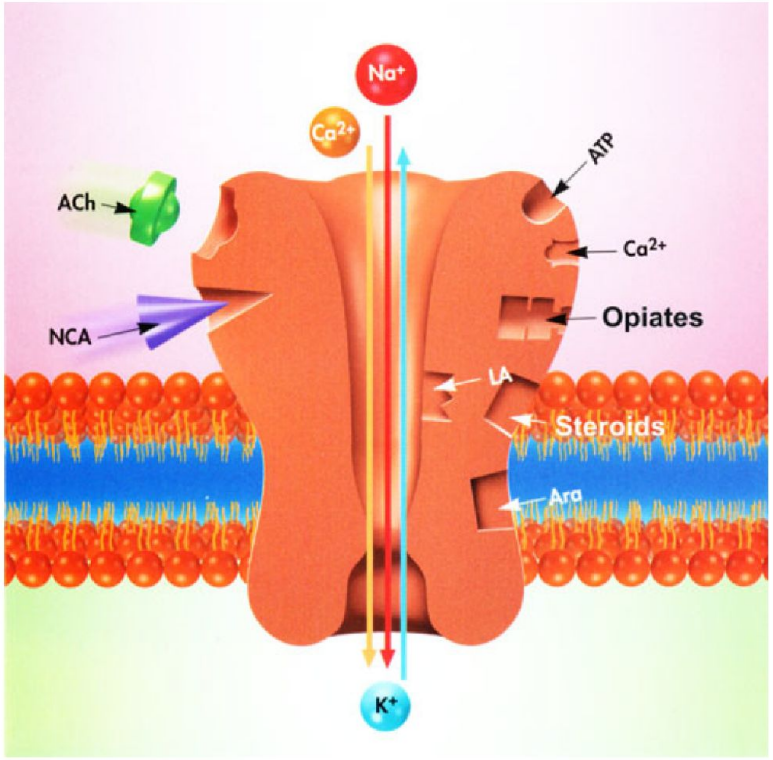
- Testing function via flux assays in vesicles
- Zn^{2+} (and CO^{2+}) allowed to enter, Ca^{2+} does not
- Proton-Zn^{2+} co-transportation occurs

Limitations

- Serendipitous design process
  - Lots of previous information on Rocker
- No x-ray structure of Rocker tetramer
- Kinetics demonstrate lack of selectivity and low rate of transport
- Would be cool to see full transporter simulation



Future Directions and Applications



- Improve selectivity and activity of Rocker
- Apply design principles to create custom biosensors
- Could create custom ion transporter with optimized rates

Steller et al. Anal Bioanal Chem 2012

References

- Figures from (unless otherwise stated):
  - Joh, N. H., Wang, T., Bhate, M. P., Acharya, R., Wu, Y., Grabe, M., Hong, M., Grigoryan, G., and DeGrado, W. F. (2014) De novo design of a transmembrane Zn2+-transporting four-helix bundle., Science 346, 1520–4
- Rasband, M. N. (2010) Ion Channels and Excitable Cells. Nature Education 3(9):41
- Grigoryan, G., and Degrado, W. F. (2011) Probing designability via a generalized model of helical bundle geometry., *J. Mol. Biol. 405*, 1079–100.
- Steller, Kreir, and Salzer. (2012) Natural and artificial ion channels for biosensing platforms, *Analytical and Bioanalytical Chemistry 402*, 209–230.

ARTICLE

doi:10.1038/nature19791

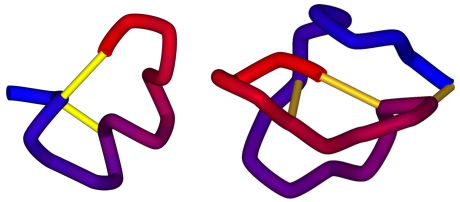
Accurate *de novo* design of hyperstable constrained peptides

Gaurav Bhardwaj[1,2]*, Vikram Khipple Mulligan[1,2]*, Christopher D. Bahl[1,2]*, Jason M. Gilmore[1,2], Pet[illegible][3], Olivier Cheneval[3], Garry W. Buchko[4], Surya V. S. R. K. Pulavarti[5], Quentin Kaas[3], Alexander Eletsk[illegible][5], Po-Ssu Huang[1,2], William A. Johnsen[6], Per Jr Greisen[1,2,7], Gabriel J. Rocklin[1,2], Yifan Song[1,2,8], Thomas W. Linsky[1,2], Andrew Watkins[illegible], Stephen A. Rettie[2], Xianzhong Xu[5], Lauren P. Carter[2], Richard Bonneau[10,11], James M. Olson[6], Evangelos Coutsias[12], Colin E. Correnti[6], Thomas Szyperski[5], David J. Craik[3] & David Baker[1,2,13]

Constrained Peptides

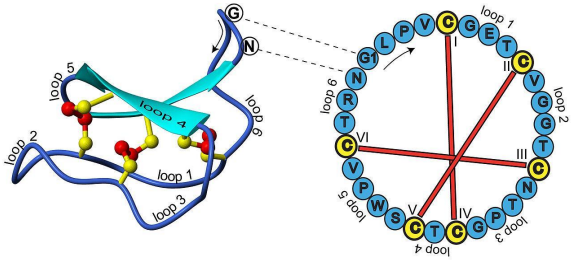
- Small proteins that have a "constrained" structure
  - disulfide bonds
  - backbone cyclization
- Extremely stable

Conotoxins



Knottins

Cyclotides



Sources: Wikipedia, Cochran Lab

Desirable Drug Candidates

Current Drugs

Small Molecule	Biologics
<500 Da	>5000 Da
Can be delivered orally	High specificity
Can permeate membranes	High potency
Stable	Natural components
Low Cost	

- Constrained peptides = 500 – 2000 Da
- A middle ground
- Combines advantages of small molecules and biologics

Problem: Lack of variety in nature

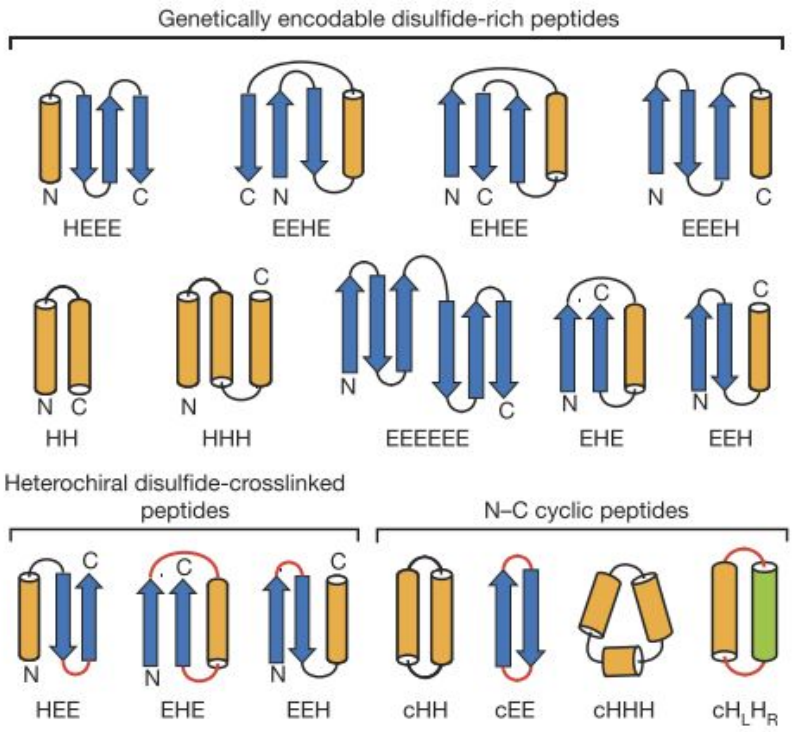
- Methods to re-engineer existing constrained peptides have had limited success

- Difficult to design constrained peptides with the exact structure to complement drug targets using portions of existing peptides

Goal:
Develop method for creating constrained peptides with new and unusual structures

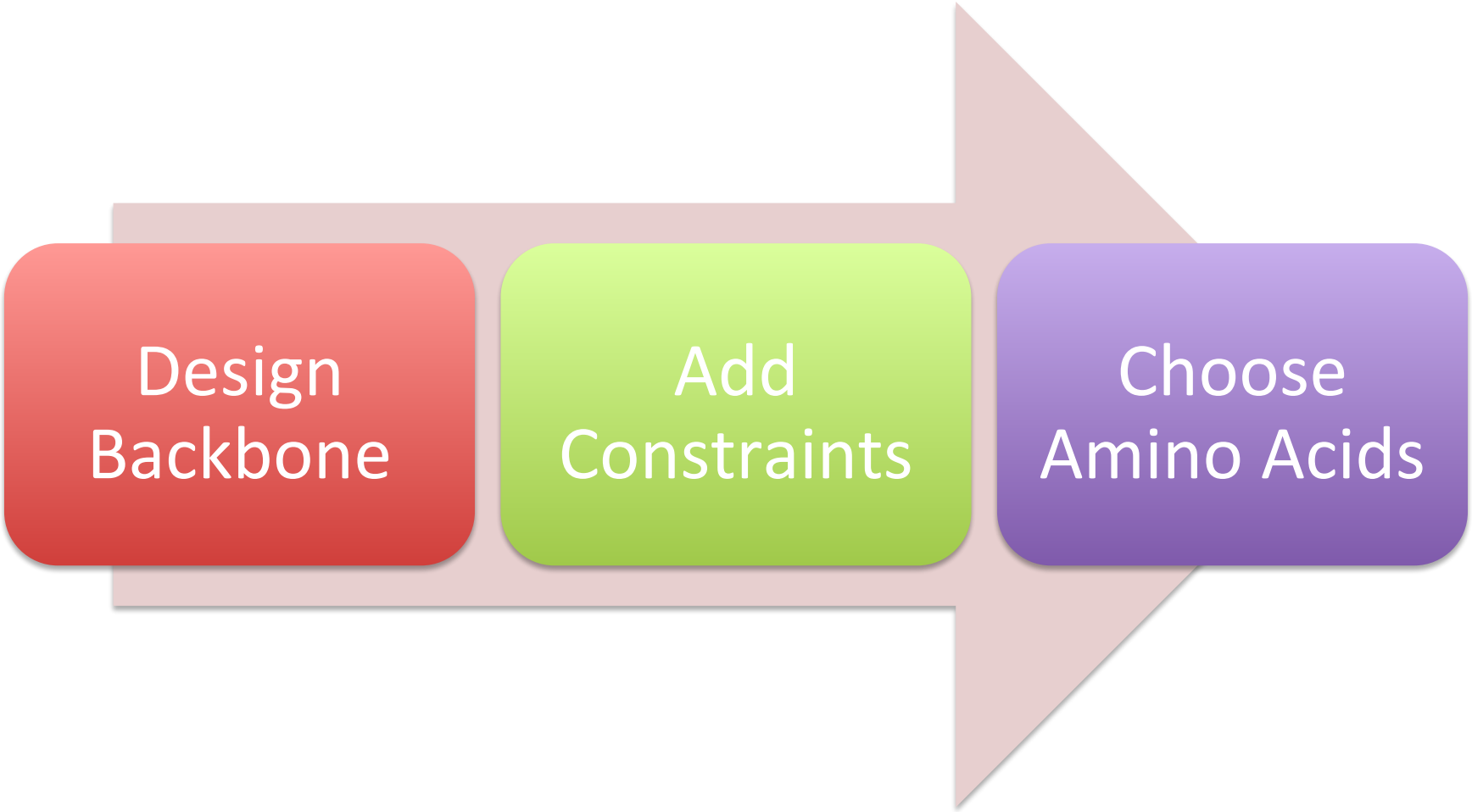
Overview

- Designed a structurally diverse array of peptides
  - Designed general topologies, wanted to find specific geometries
  - Unusual backbone combinations
  - Combination of natural and unnatural amino acids



Source: Bhardwaj *et al*, 2016

General Computational Approach

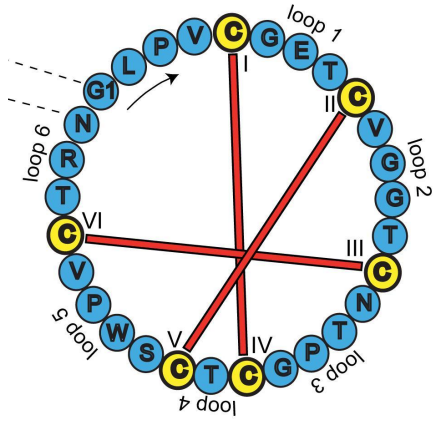


Step 1: Design Backbone

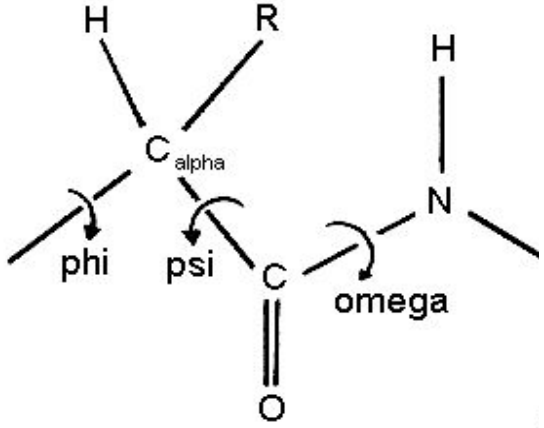
- Method 1: Assemble short fragments of known proteins together
  - Monte Carlo method with Rosetta energy function
    - Large number of protein fragments from database
    - Choose whether or not to add specific fragment to backbone based on resulting free energy (lowest free energy is most stable)
  - Limited to naturally-occurring conformations

Step 1: Design Backbone

- Method 2: GenKIC (for cyclic designs)
  - "Generalized Kinematic Closure"
  - Analytically solve kinematic equations find torsional angles that will close the loop
  - Filter out solutions with unfavorable amino acid interactions
  - Choose 'best' solution based on Rosetta free energy



Peptide Torsional Angles

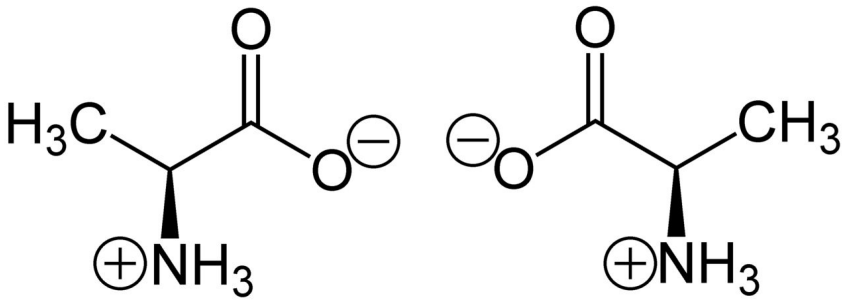


Sources: Wikipedia, www.cyst.bbk.ac.uk

Modification: Heterochiral peptides

- Chiral molecules: mirror images
  - For amino acids, called L- or D- amino acids
  - D- amino acids not usually found in nature
- Heterochiral: includes both L- and D- amino acids
- Manipulated Rosetta energy function to support D-amino acids
  - Invert torsional potential used for equivalent L-amino acids

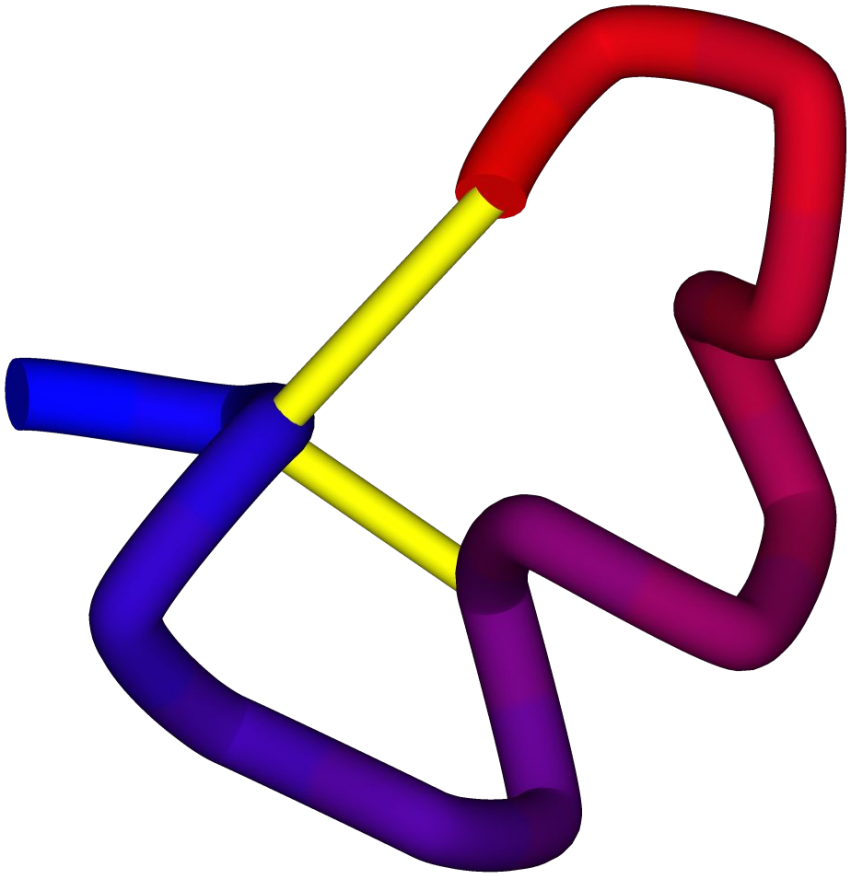
L- vs. D- Alanine



Source: Wikipedia

Step 2: Add Constraints

- Scanned for sites to add disulfide bonds to stabilize backbone conformation
- Ranked the set of all sterically possible combinations of disulfide bonds according to their effect on energy of unfolded state (lower energy = higher ranking)
- Chose 1-3 of highest ranking bonds



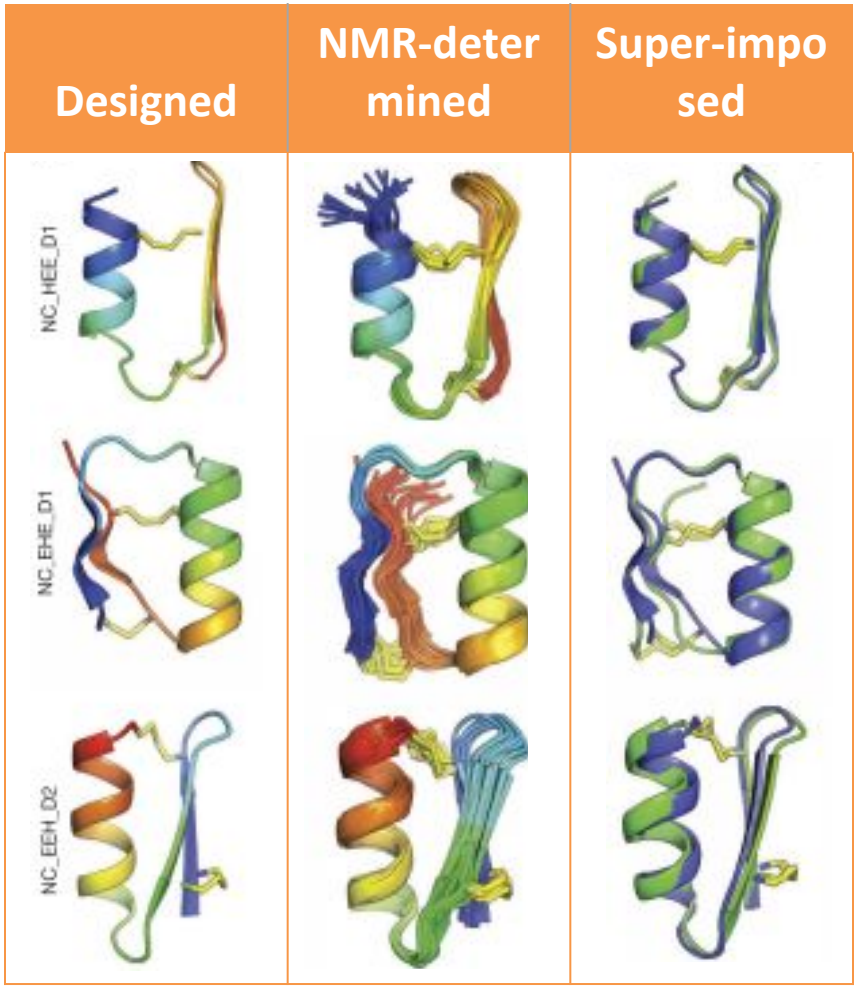
Source: Wikipedia

Step 3: Choose Amino Acids

- Optimize rotamers (allowing changes to side chains)
- Position groups: core, boundary, surface
  - Dictated which amino acids allowed at each position
  - Ex: hydrophobic amino acids only allowed at core
- Relax backbone (allow it to move)
- ~80,000 structures per topology → filtered based on overall energy, backbone quality, and disulfide geometry

Results

- Experimental structures (NMR, crystallography) in close agreement with the designed structures
- Extremely stable – resistant to thermal and chemical denaturation
- Unique structures: searches for similar structures in PDB found minimal matches



Source: Bhardwaj *et al*, 2016

Limitations

- Energy functions are not perfect
- Method is not general enough to use for all types of backbones
  - Two different methods for backbone design
  - Must make specific alterations for heterochirality, cyclization
- Too much filtering at end!
- Validation is not integrated into pipeline

References

- Bhardwaj G et al. Accurate de novo design of hyperstable constrained peptides. *Nature.* 2016, Oct : 329 – 335.
- Craik DJ et al. The future of peptide-based drugs. *Chemical Biology & Drug Design*. 2013, Jan: 136 – 147.

The following slides were additional in case of questions or extra time

Filtering: Symmetric Docking

- Arrange the building blocks into the overall model with some degrees of freedom
- Score the interaction at the interface between the building block pairs
- Add bias to the scoring function to favor structures that match PDB crystal structures
- Filter based on size of interface, secondary structure of interface, and orientation
- 66,115 configurations of type I53, 35,468 of type I52, and 161,007 of type I32 passed this stage

Filtering: Interface Design

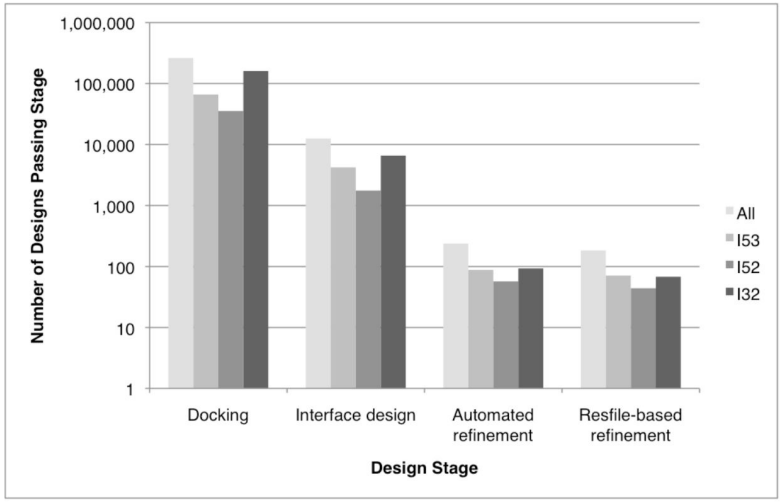
- Iterative process to perturb configuration and pack side chains
- Minimize Rosetta energy function
- Modify rotamer library to include certain motifs from earlier stage
- Filter out designs based on areas of interface contact, predicted binding energy, Rosetta energy function, and position of certain residues

Filtering: Automated Reversion

- Greedy optimization to revert changes that either:
  - Aren’t necessary for assembling the model, or
  - Result in poor packing within the building blocks

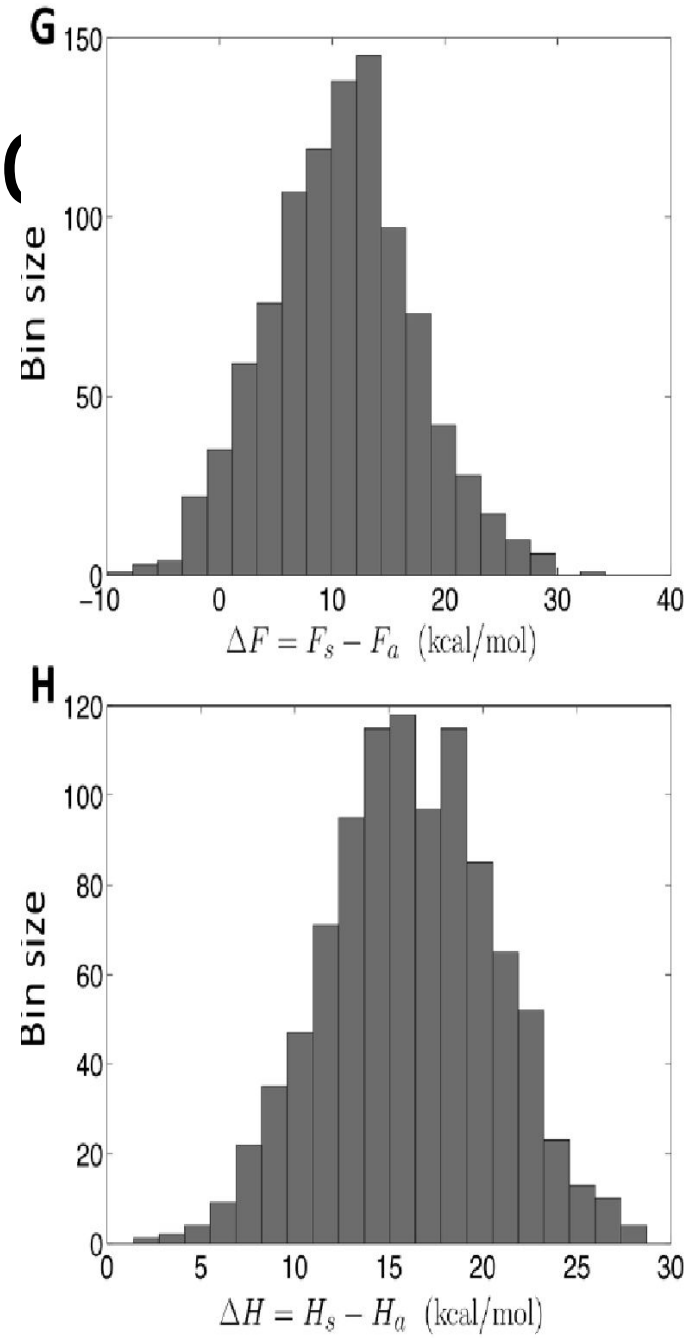
Filtering: Resfile-Based Refinement

- Visual inspection of the models
- Resulted in:
  - 71 designs of type I53
  - 44 designs of type I52
  - 68 designs of type I32



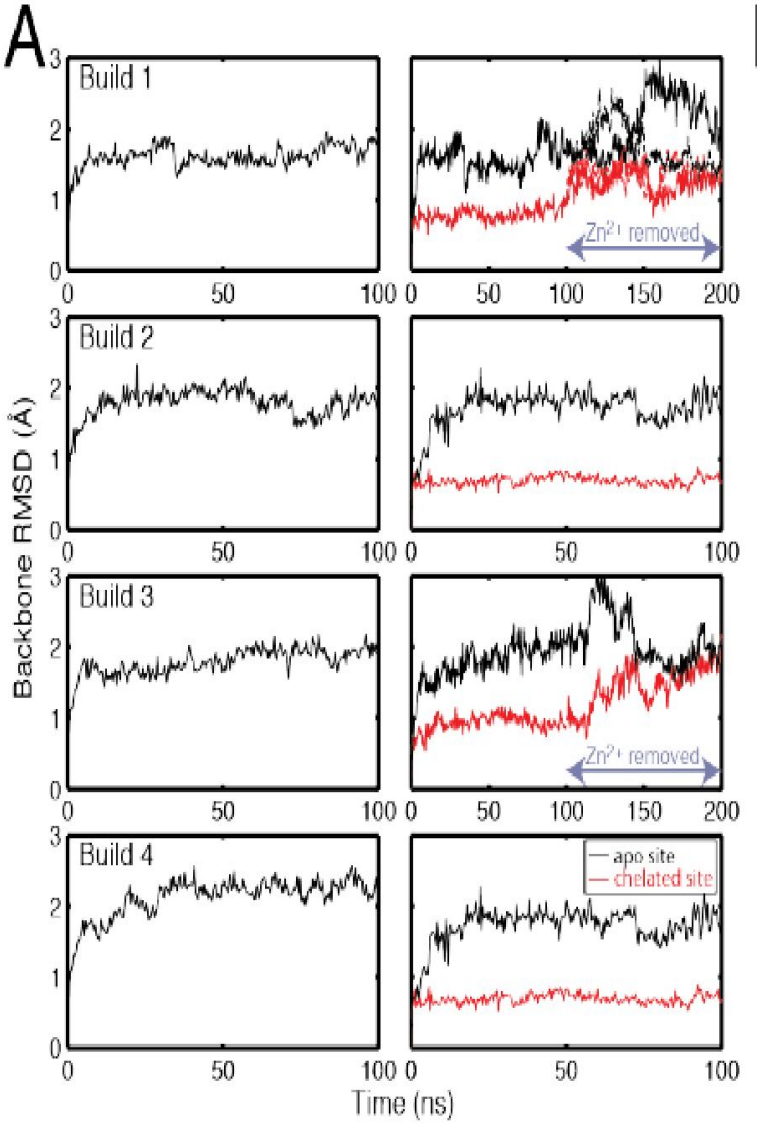
Additional Slide – VALO

rendered in green, but are generally occluded by other structural matches. **G**, VALOCIDY-estimated thermodynamic parameters for sequences optimized in the protein design stage. **G** and **H** show the distributions of differences in free energy and enthalpy, respectively, between the symmetric and asymmetric states. Most sequences prefer the symmetric state and all prefer it enthalpically.

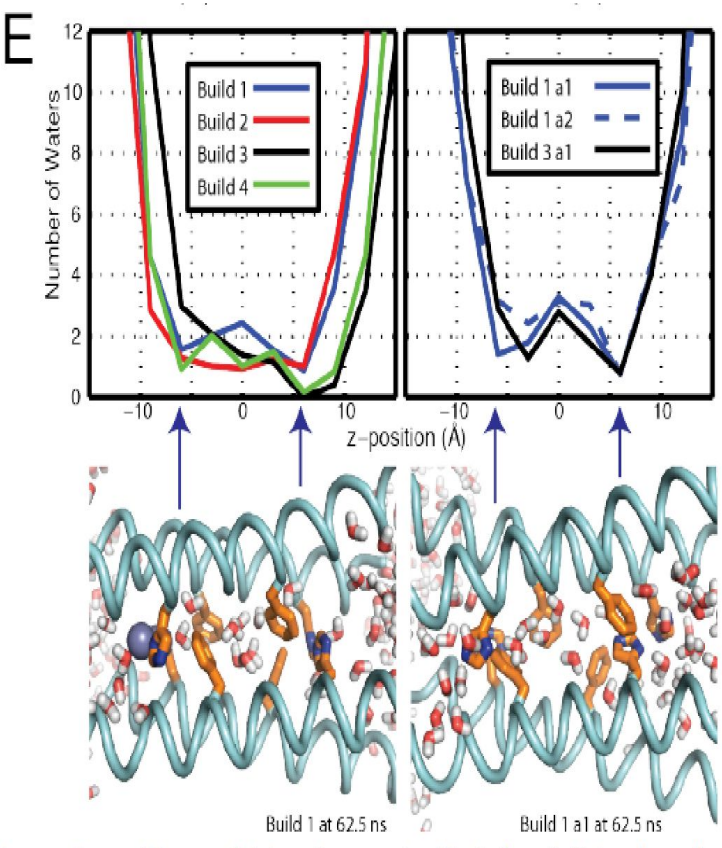


Molecular Dynamic Simulation

- Ran 4 simulations in CHARMM:
  - All are stable, with Cα root mean square deviation (RMSD) of 0.75 Å

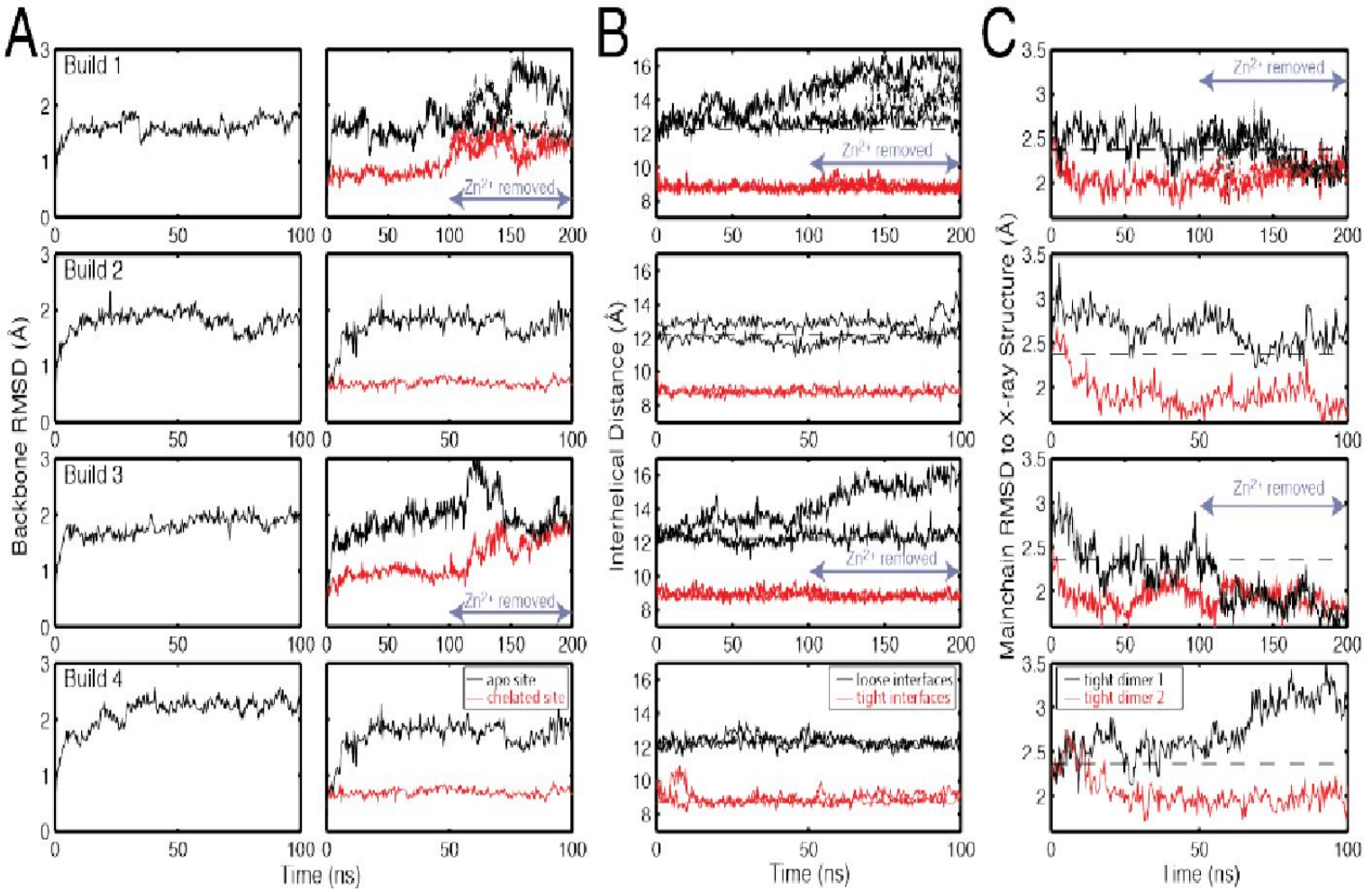


Molecular Dynamic Simulation

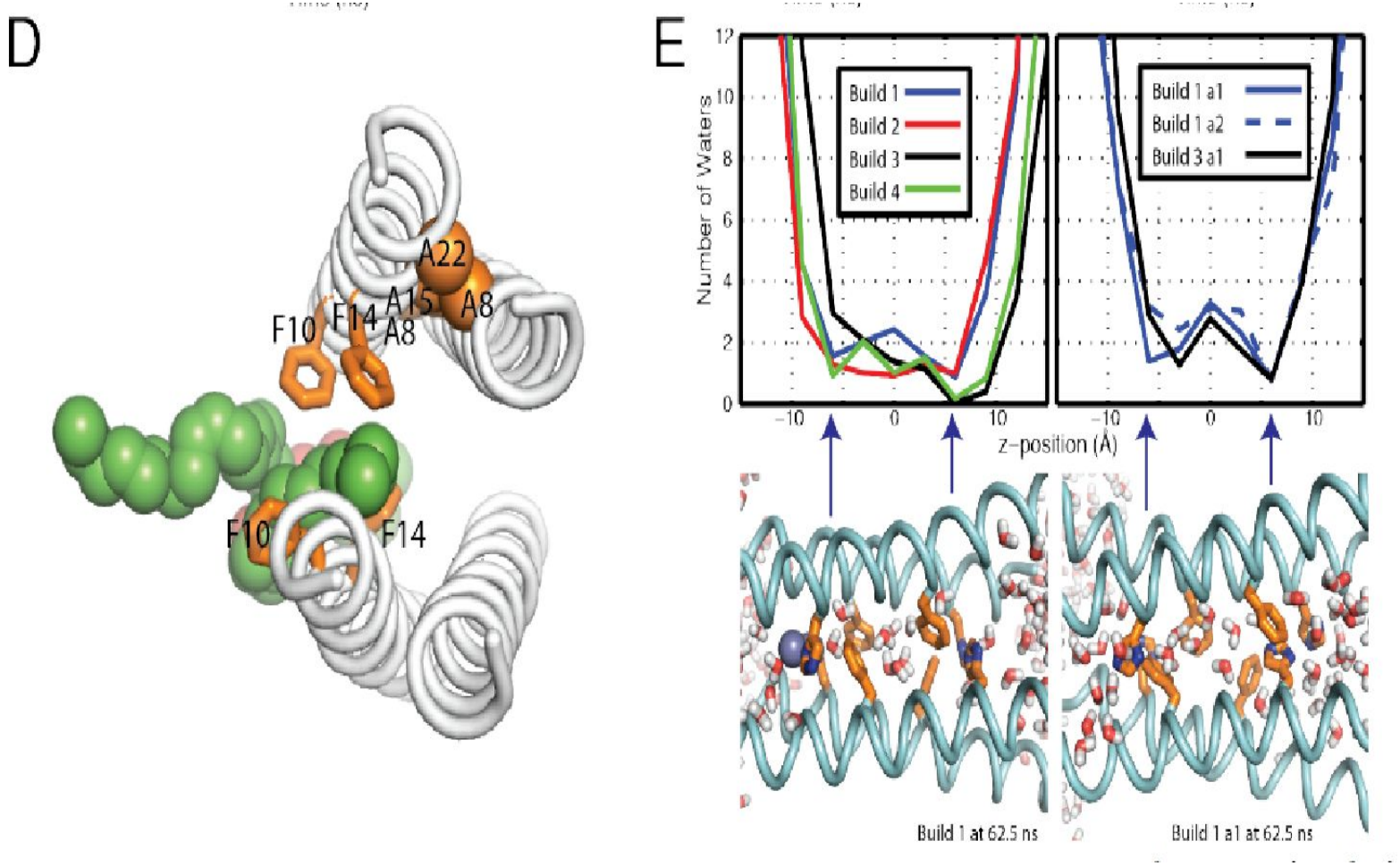


- 10-14 water molecules within lumen of Rocker
- Phe residues near Zn^{2+} binding site create dry zones (blue arrows)

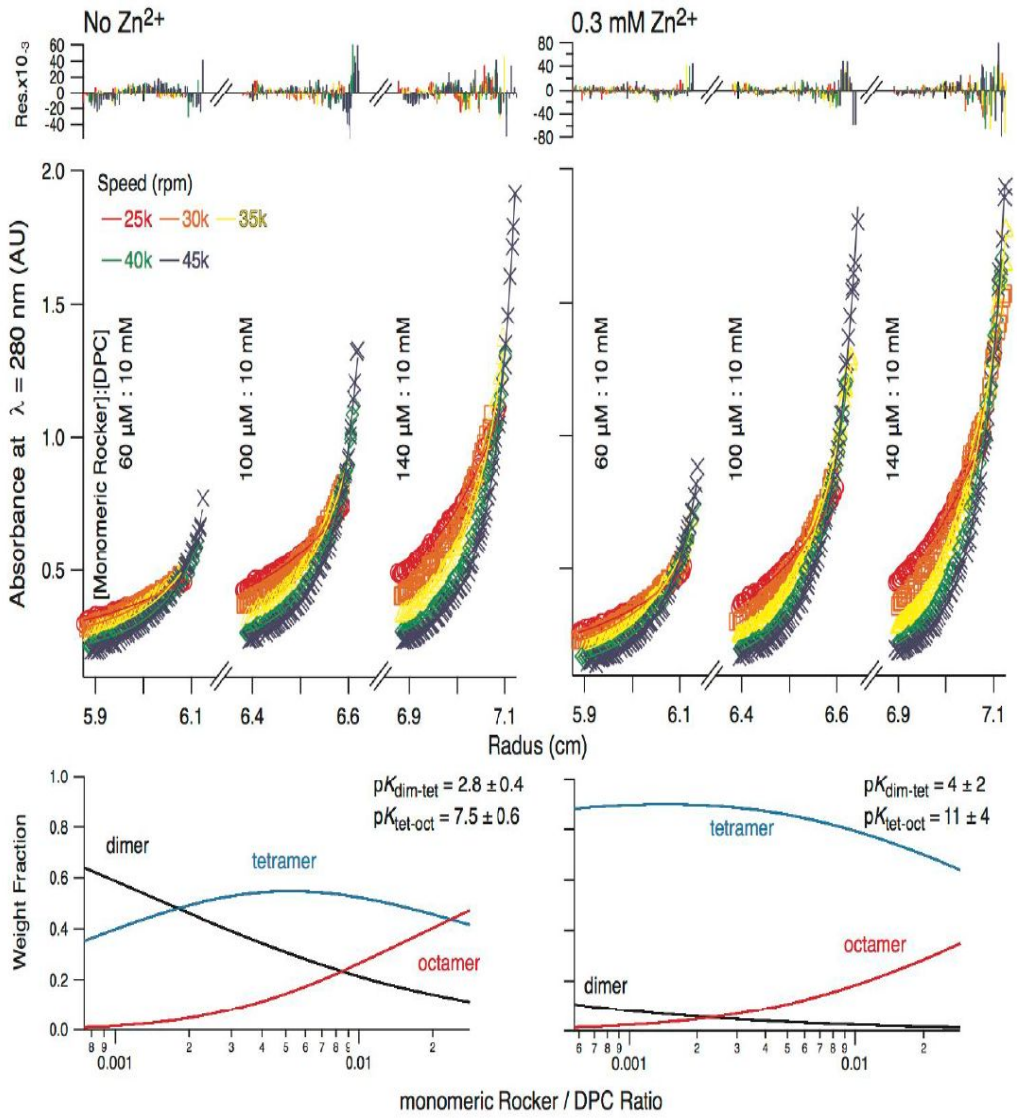
Additional Slide – MD simulations



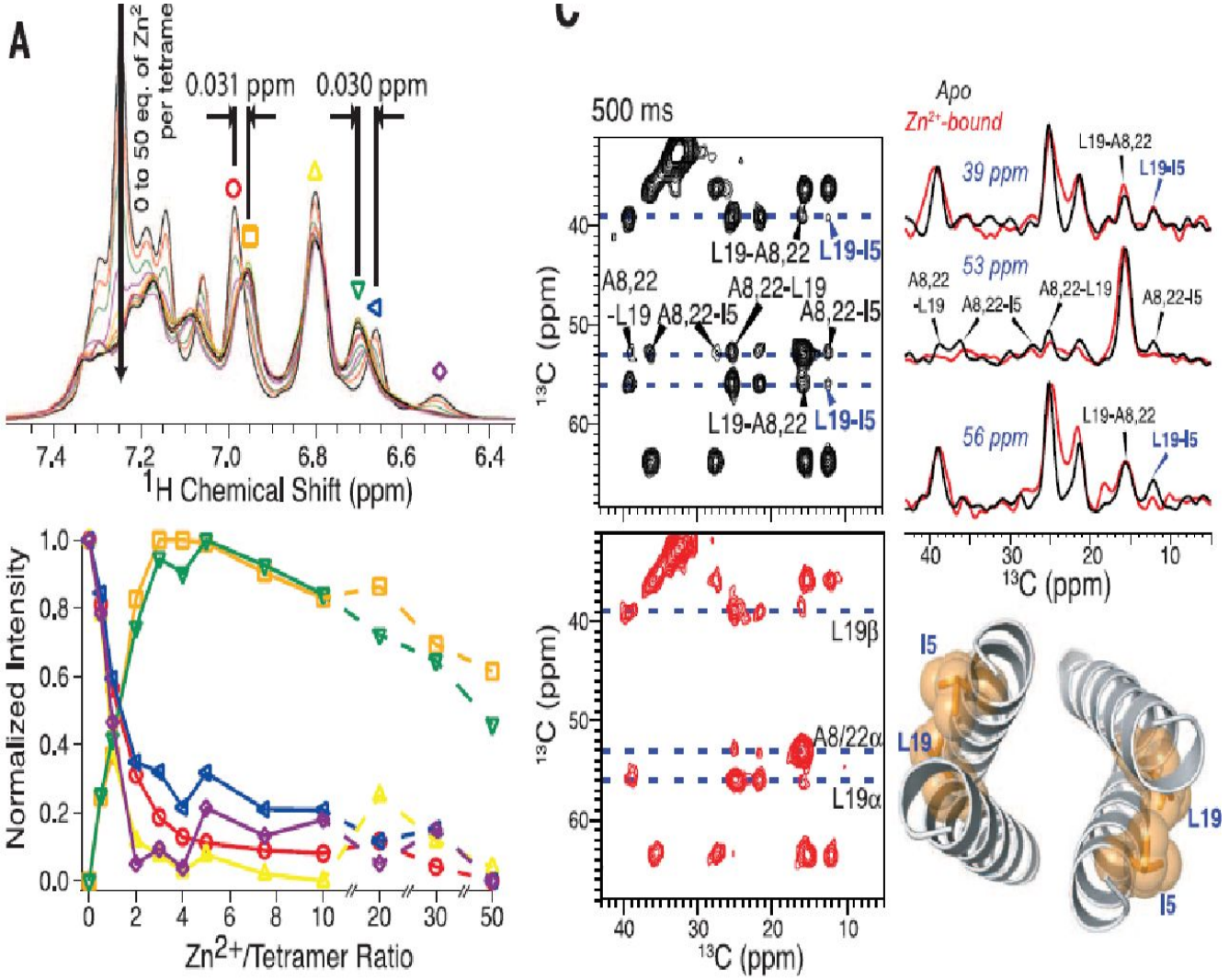
Additional Slide – MD simulations



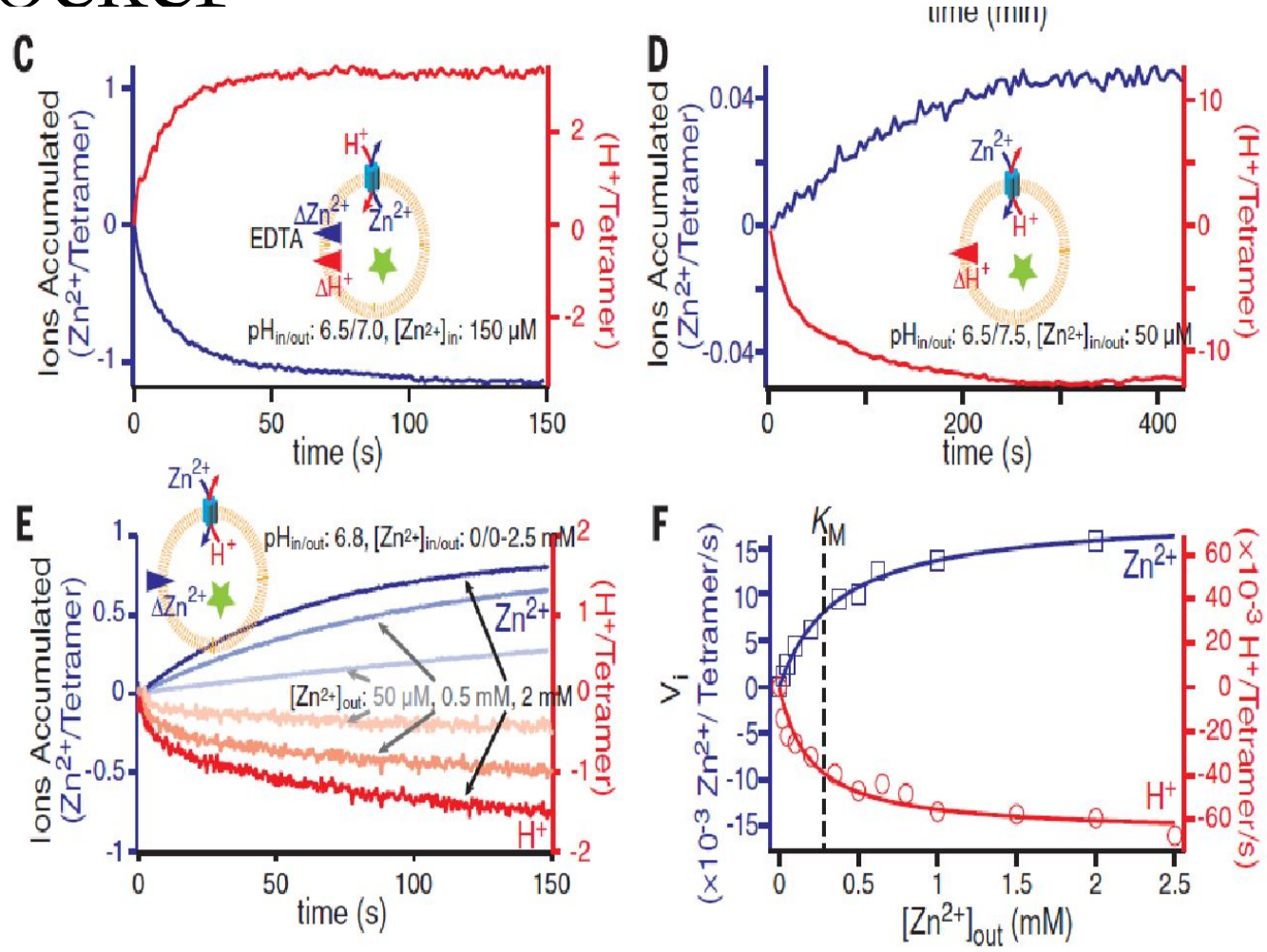
Additional Slide – Analytical Ultrac



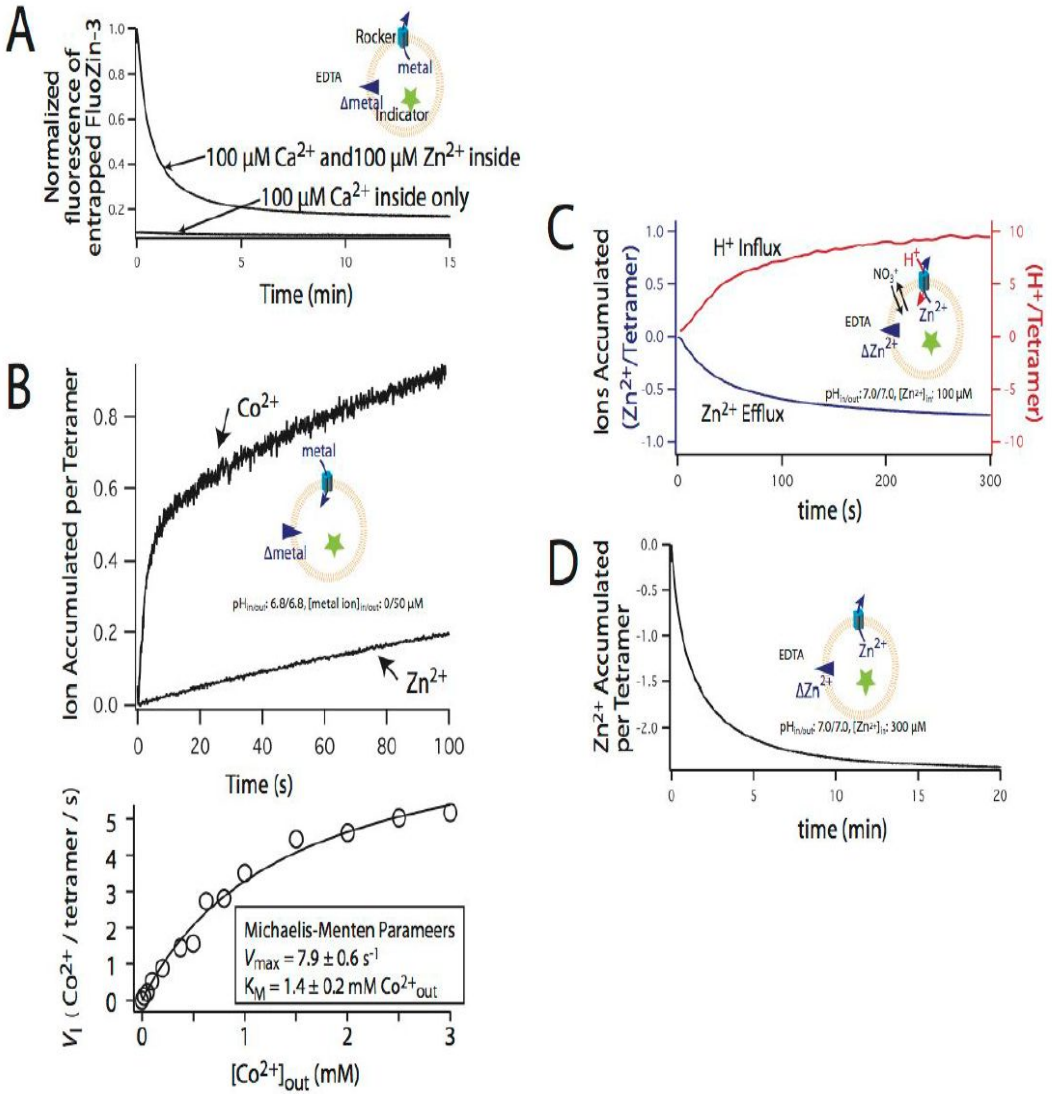
Additional Slides – SSNMR/SDNMR ^{1}H and ^{13}C



Additional Slides – Kinetics of Rocker



Additional Slide – Cobalt Kinetics



Design Process

- Design backbone
  - Ab initio
  - From crystal structures
- Select rotamers
  - To minimize energy or select desired functions
- Filter designs computationally
- Confirm results experimentally

Structure Confirmation Methods

- Nuclear magnetic resonance spectroscopy (NMR)
- X-ray crystallography
- Electron microscopy (EM, cryoEM)
- Electrophoresis
- Size-exclusion chromatography
- Small-angle light scattering