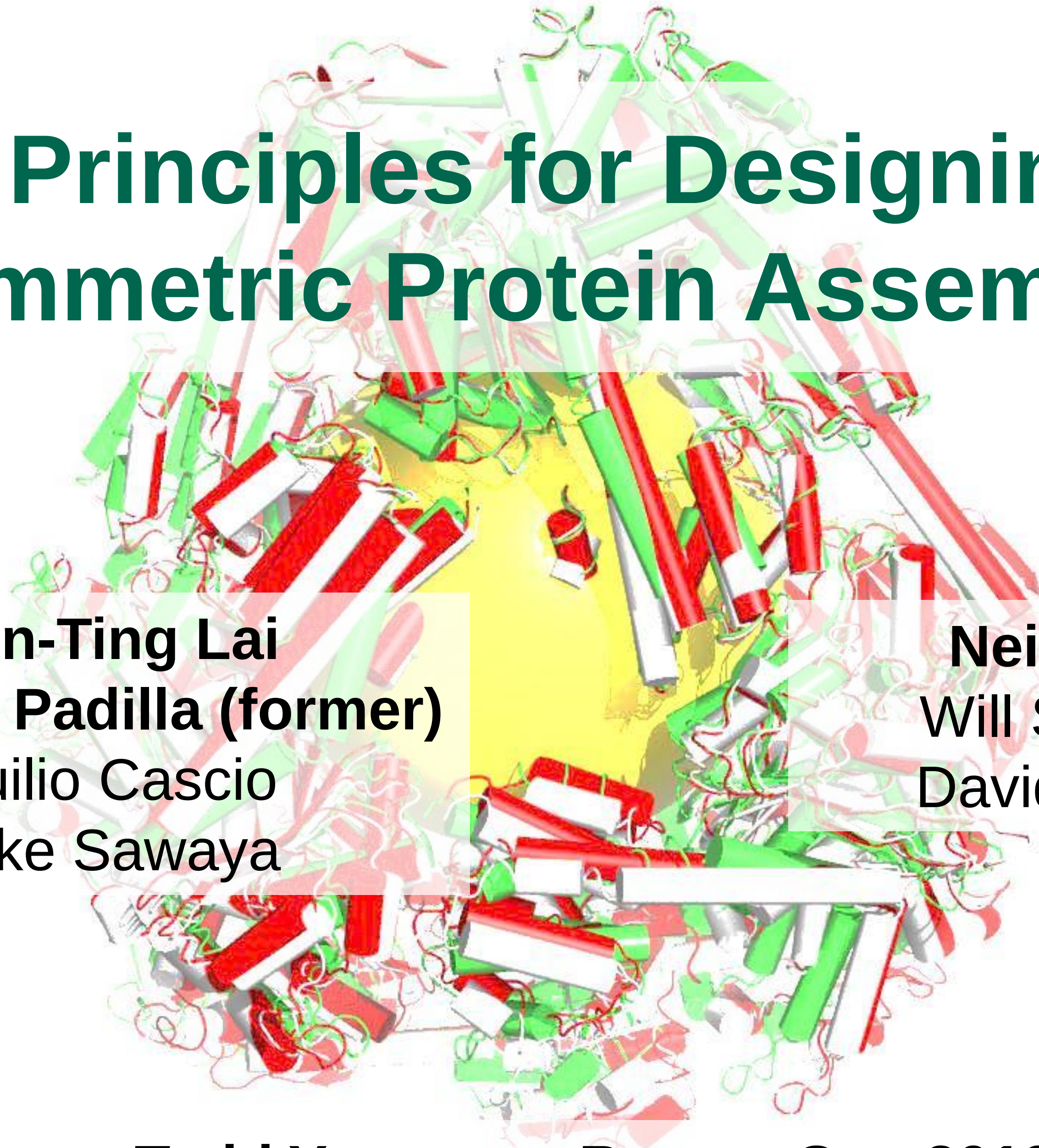
Principles for Designing Symmetric Protein Assemblies

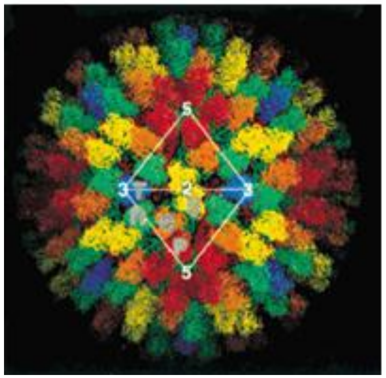


Yen-Ting Lai
Jennifer Padilla (former)
Duilio Cascio
Mike Sawaya

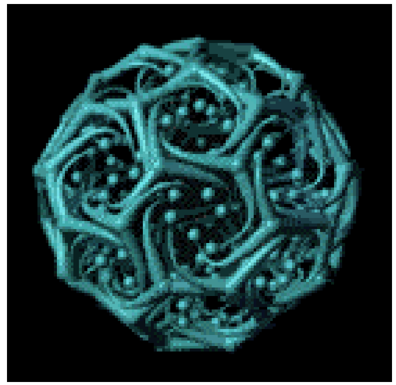
Neil King
Will Sheffler
David Baker

Todd Yeates -- RosettaCon 2012

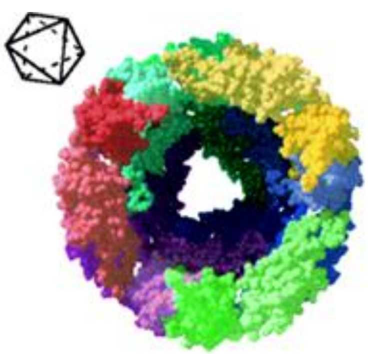
Biological Protein Assemblies: An inspiration and a challenge



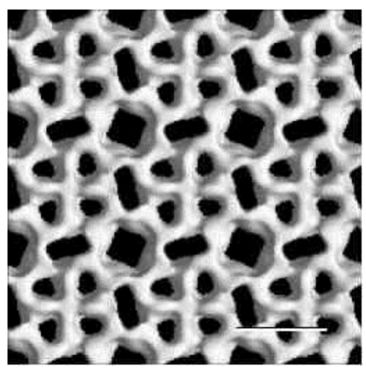
virus capsid



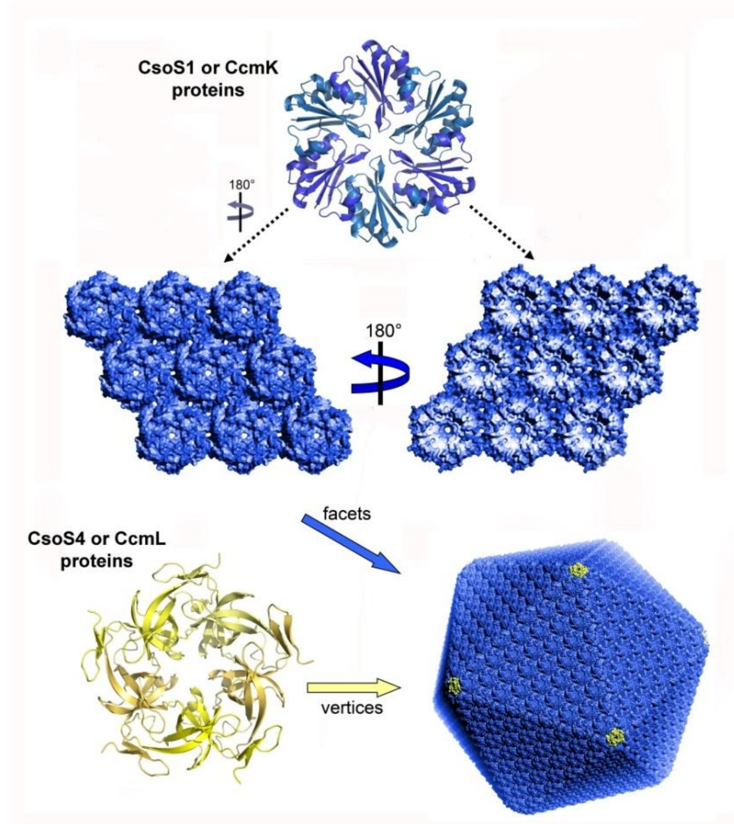
clathrin



Heat shock protein



bacterial S-layers



bacterial microcompartments

Design assembly successes trail those using DNA

Symmetry-Centric Approaches to Designing Protein Assemblies

Biological Protein <u>Assemblies</u>

- **Nearly always symmetric**
- **Predicted as early as 1956 by Crick and Watson**
- **Repetitive symmetric assemblies require <u>fewer distinct contact or interface types</u>**

NATURE March 10, 1956 Vol. 177

STRUCTURE OF SMALL VIRUSES

molecular level, a structure of a definite size and shape has to be built up from smaller units; namely, that the packing arrangements are likely to be repeated again and again—and hence that the sub-units are likely to be related by symmetry elements.

Table 1. The Three Possible Cubic Point Groups for a Spherical Virus

Crystallographic description	No. and type of rotation axes present	No. of asymmetric units	Platonic solid with these symmetry elements
23	3 2-fold 4 3-fold	12	Tetrahedron
432	6 2-fold 4 3-fold 3 4-fold	24	Cube Octahedron
532	15 2-fold 10 3-fold 6 5-fold	60	Dodecahedron Icosahedron

The number of sub-units will be the same as, or a multiple of, the number of asymmetric units

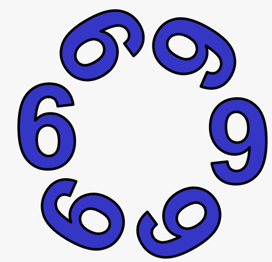
F. H. C. Crick
J. D. Watson*
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory,
Cambridge
Jan. 23.

Number of Distinct Contact Types as a Central Idea

- **Limited Outcomes with Only a Single Distinct Contact Type:**

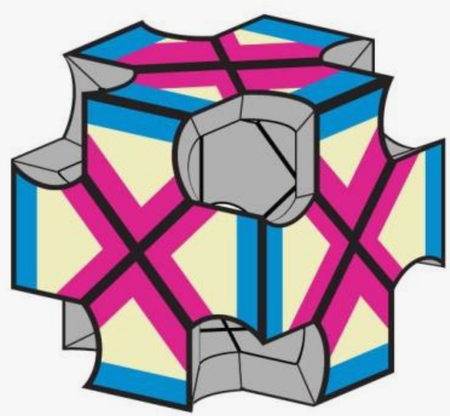
...66666...

linear or helical filaments



cyclic rings

- **Richer Outcomes Using >1 Contact Type:**



Key Design Questions and a Connection to Group Theory

- What kinds of higher symmetries should be targeted for design, and how?
- How many contacts and in what geometries are required for various symmetries?
  - Equivalent to a (incompletely solved) problem in group theory: What is the fewest number of elements of a (potentially infinite) group from which the group can be generated?

A Connection Between Design and Group Theory

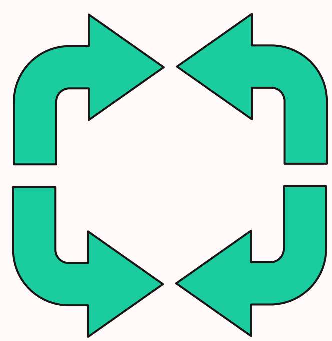
Number of designed contacts required **=** **Minimum number of elements required to 'generate' the group**

Examples:

{1,*i*,-1, -*i*}

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\}$$

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$



A Connection Between Design and Group Theory

Number of designed contacts required = Minimum number of elements required to ‘generate’ the group

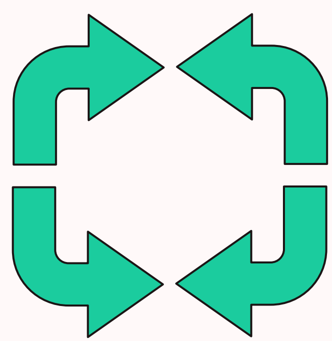
Examples:

{1,*i*,-1, -*i*}

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\}$$

one generator
one contact

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$



A Connection Between Design and Group Theory

Number of designed contacts required **=** **Minimum number of elements required to ‘generate’ the group**

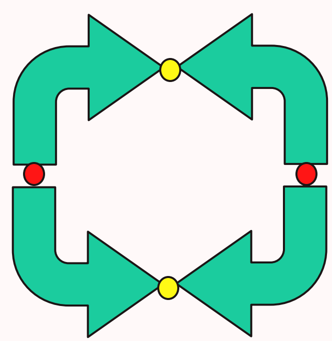
Examples:

{1,*i*,-1, -*i*}

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\}$$

one generator
one contact

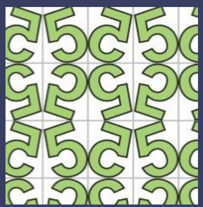
$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

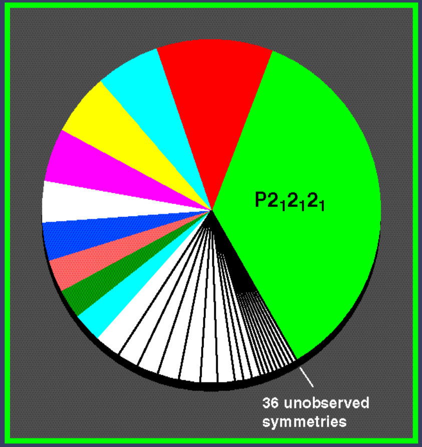


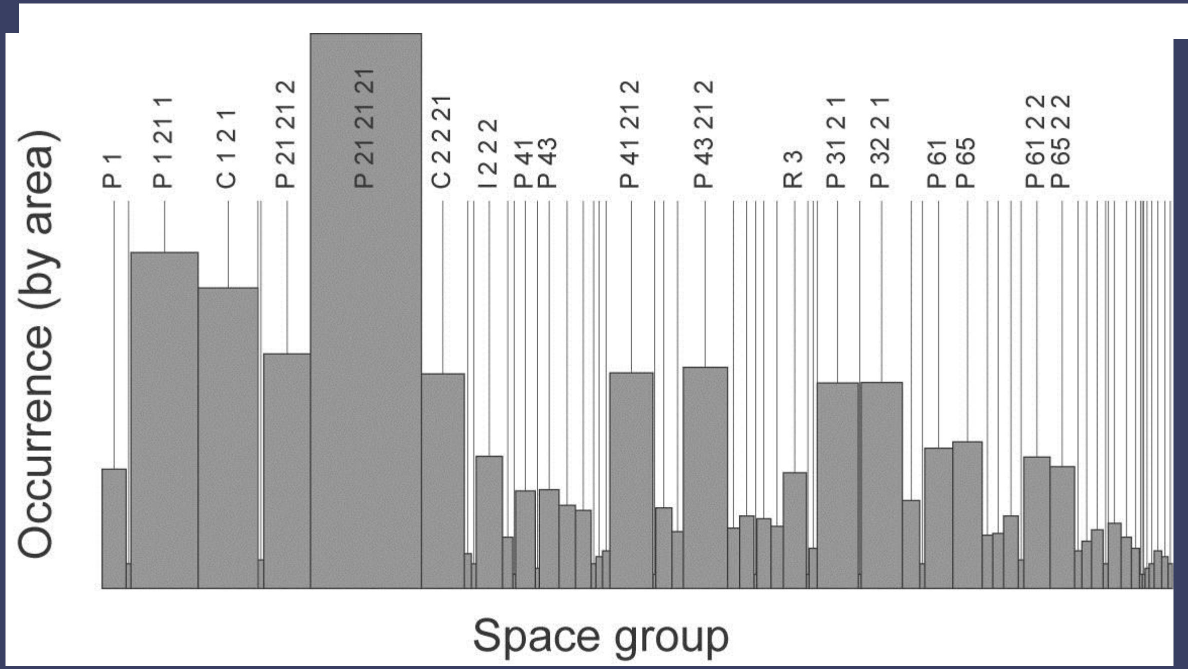
2 generators
2 contacts

A Brief Diversion:
The space group preference problem







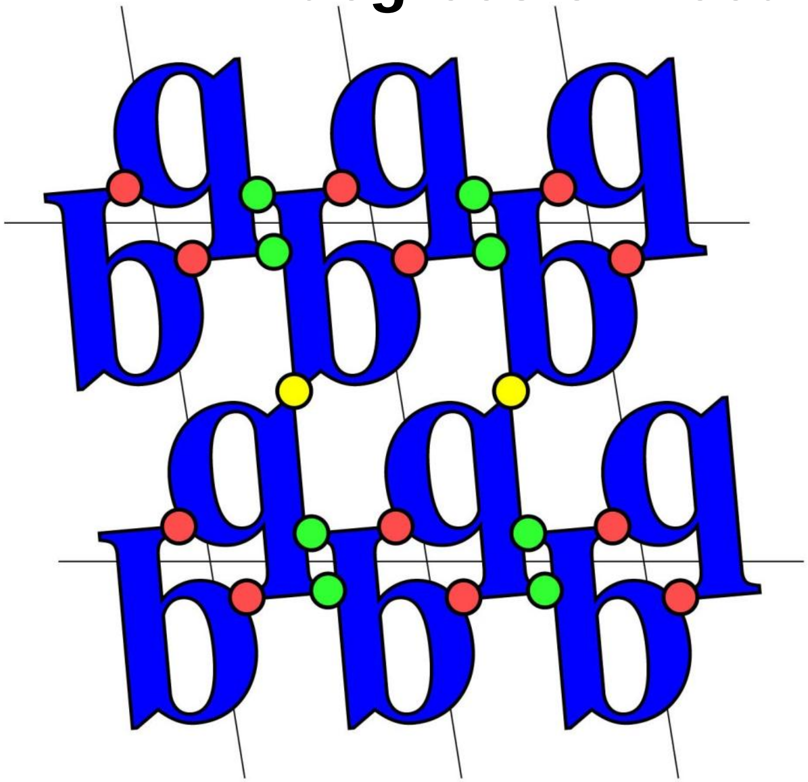


Top 1: ~33%
Bottom 55: 20%

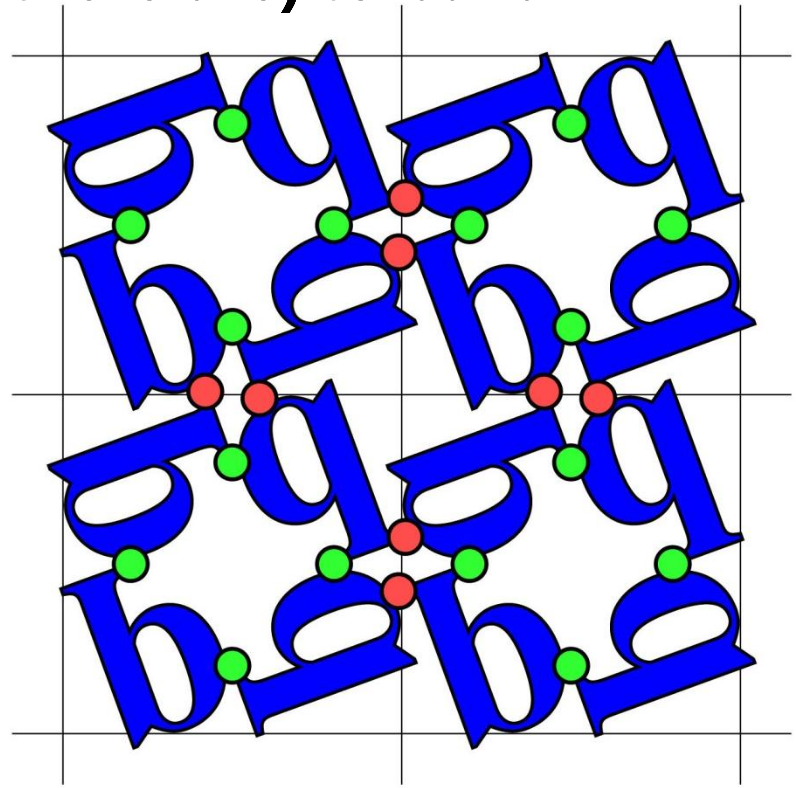
One of the most puzzling (and overlooked) problems in structural biology. The differences in probability span more than 2 orders of magnitude, yet there are no obvious energetic explanations.

Different Crystal Space Group Symmetries Have a Different Minimum Contact Number, C

This number relates to how easy it is (how many degrees of freedom there are) to build



p2, C=3



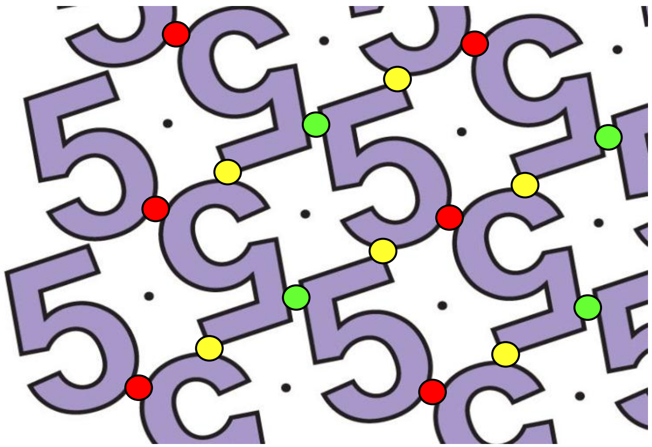
p4, C=2

C is a property of the mathematical group, not the molecule

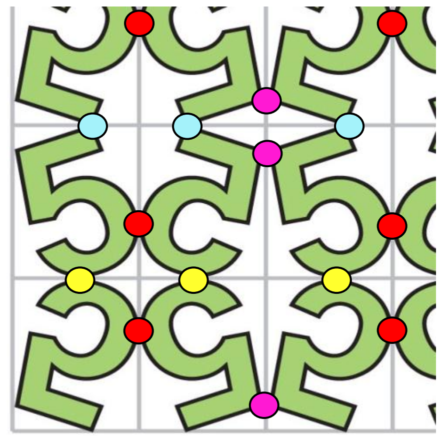
Wukovitz & Yeates, *Nat. Struct. Biol.* **2**, 1062 (1995)

Different Crystal Space Group Symmetries Have a Different Minimum Contact Number, C

This number relates to how easy it is (how many degrees of freedom there are) to build



p2, C=3
Note independence from shape



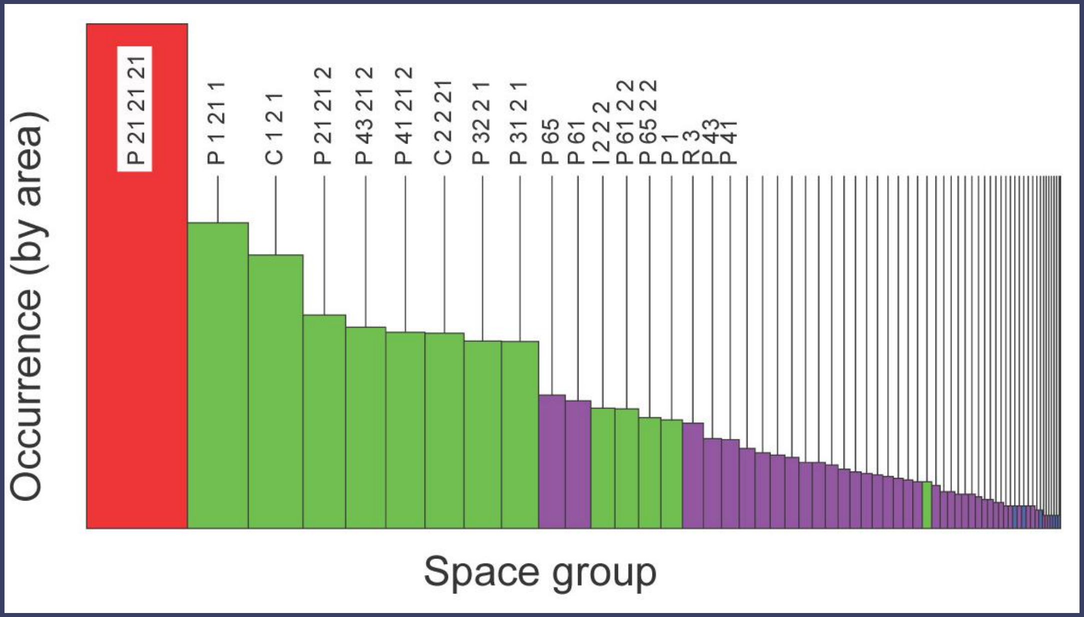
p2mm, C=4

When the values of C for the 65 biological space groups were enumerated, they provided a powerful explanation for observed space group preferences.

Wukovitz & Yeates, *Nat. Struct. Biol.* **2**, 1062 (1995)

Agreement between the dimensionality for forming different space groups and their observed frequencies

- The 65 possible space group symmetries fall into 4 categories of increasing likelihood: D= 4, 5, 6, 7   (factor of ~8 for each increment in D)

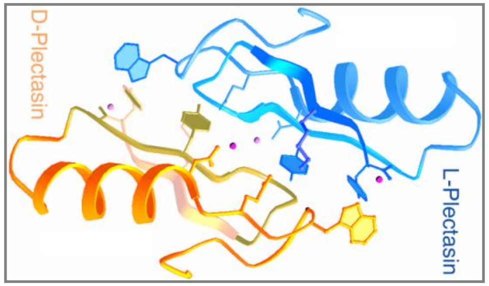


D=7
D=6
D=5
D=4

- Only one space group, $P2_12_12_1$, which dominates in macromolecular crystals, has D=7 !!
- A dimensionality analysis explains most of the observed phenomenon.

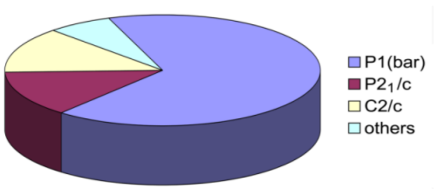
Wukovitz & Yeates, *Nat. Struct. Biol.* **2**, 1062 (1995)

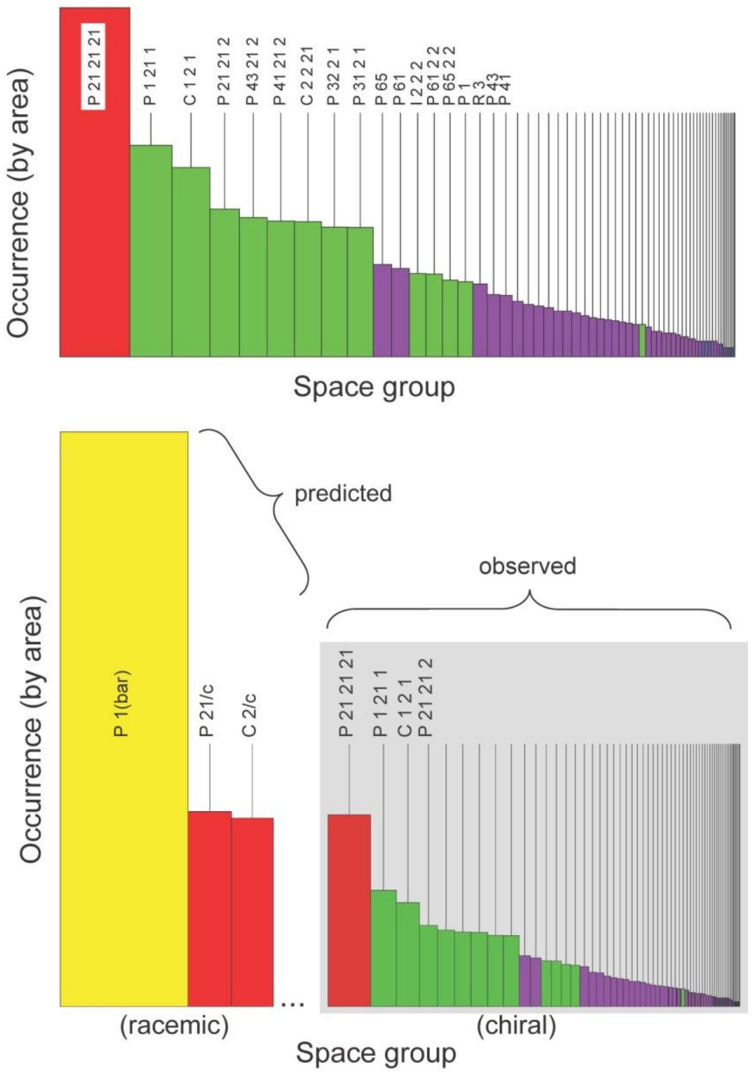
Extending the Theory: Mirror image proteins provide a potentially powerful solution to the protein crystallization problem



Predictions from theory

- **Proteins will crystallize much more easily if they can be prepared as a racemic mixture; this requires chemical synthesis of the mirror image protein (i.e. from D-amino acids)**
- **P1(bar) will dominate for racemic crystallization of proteins; this highly specific prediction provides a powerful test of the theoretical ideas**

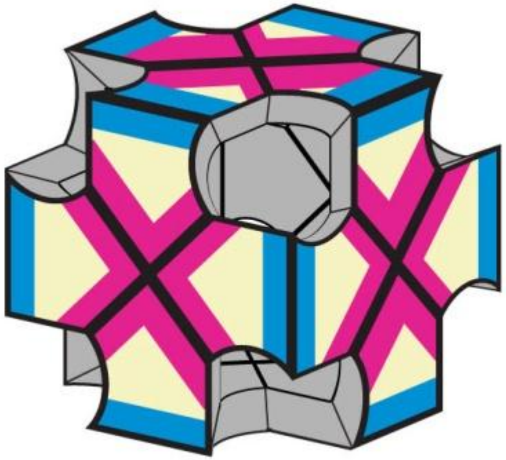




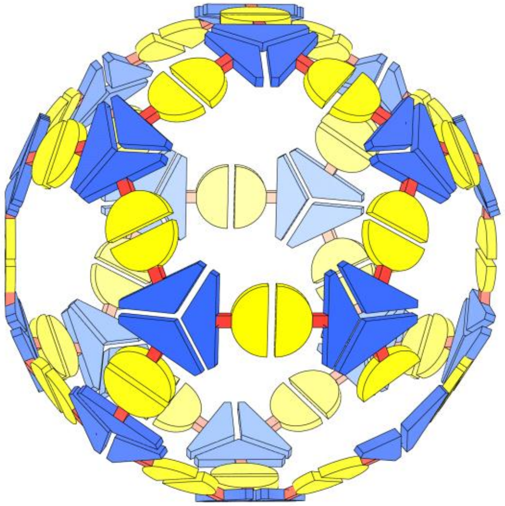
Yeates and Kent (2012). Annu. Rev. Biophys

Returning to the Problem of Designed Assembly:

A remarkable number of highly symmetric groups can be generated with just two operators (or contact types)!



cubic symmetry octahedral symmetry formed only by two-fold and four-fold symmetric contacts



icosahedral symmetry formed only by two-fold and three-fold symmetric contacts

Design Rules (2-fold + 3-fold)

Symmetry	Construction	Geometry of symmetry elements	
		Cages and shells	
T	Dimer-Trimer	**54.7°**,	Intersecting
O	Dimer-Trimer	**35.3°**,	Intersecting
I	Dimer-Trimer	**20.9°**,	Intersecting
		Double-layer rings	
D_n	Dimer-Dimer	**180°/n**,	Intersecting
		Two-dimensional layers	
p6	Dimer-Trimer	**0°**,	Non-intersecting
p321	Dimer-Trimer	**90°**,	Non-intersecting
p3	Trimer-Trimer	**0°**,	Non-intersecting
		Three-dimensional crystals	
$I2_13$	Dimer-Trimer	**54.7°**,	Non-intersecting
$P4_132$ or **$P4_332$**	Dimer-Trimer	**35.3°**,	Non-intersecting
P23	Trimer-Trimer	**70.5°**,	Non-intersecting
		Helical filaments	
Helical	Dimer-Dimer	**any angle**,	Non-intersecting
		Tubes of indefinite length	
Tubular	Dimer-Dimer-Dimer	**N, N, N,** each intersecting the cylinder axis perpendicularly	

cages

2-D layers

3-D crystals

filaments and rods

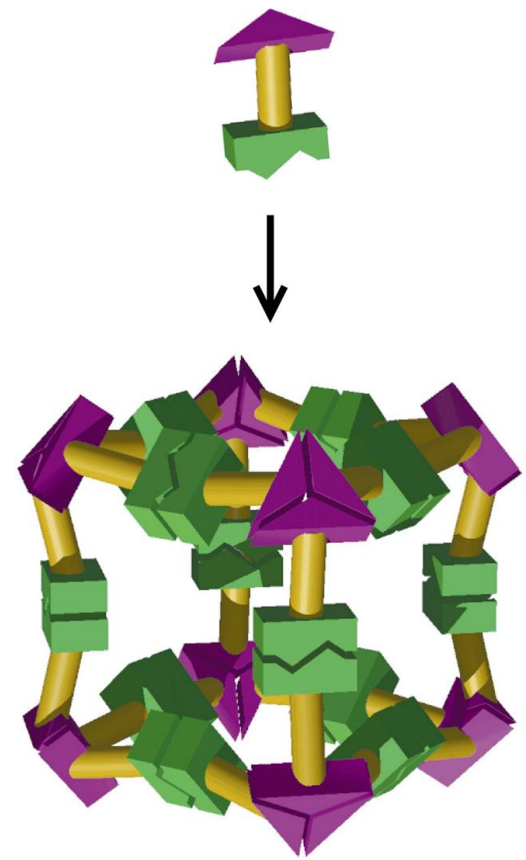
Tetrahedral, T
2-fold & 3-fold Intersecting at 54.7°

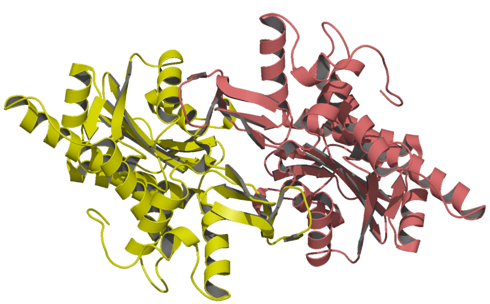
Space group $I2_13$
2-fold & 3-fold non-intersecting at 54.7°

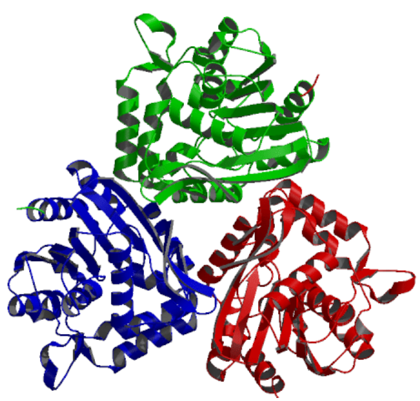
Padilla, Colovos, & Yeates, *PNAS*, **98**, 2217 (2001)

Extension of the Symmetric Contact Idea to a Strategy for Designing Self-Assembling Protein Materials

- **Natural oligomeric (e.g. dimeric and trimeric) proteins can serve as the building blocks**
- **Fusing two such proteins together (e.g. by genetic engineering) provides the two interactions needed for a rich variety of designs**



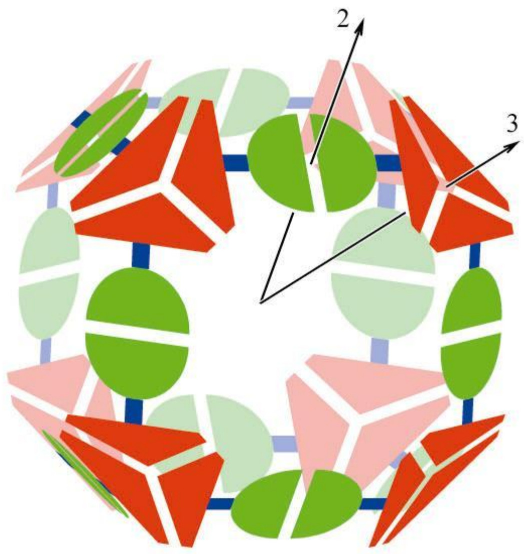


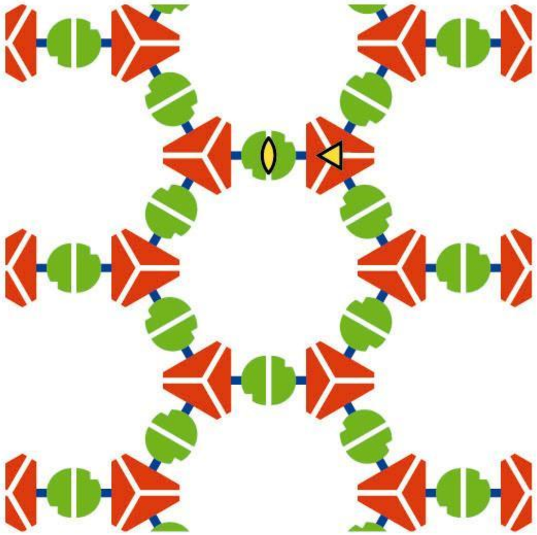


N.B. Sequence design not feasible in 2001.

The Geometry of the Symmetry Axes Dictates the Assembly and <u>Must</u> be Controlled:

Two example outcomes:

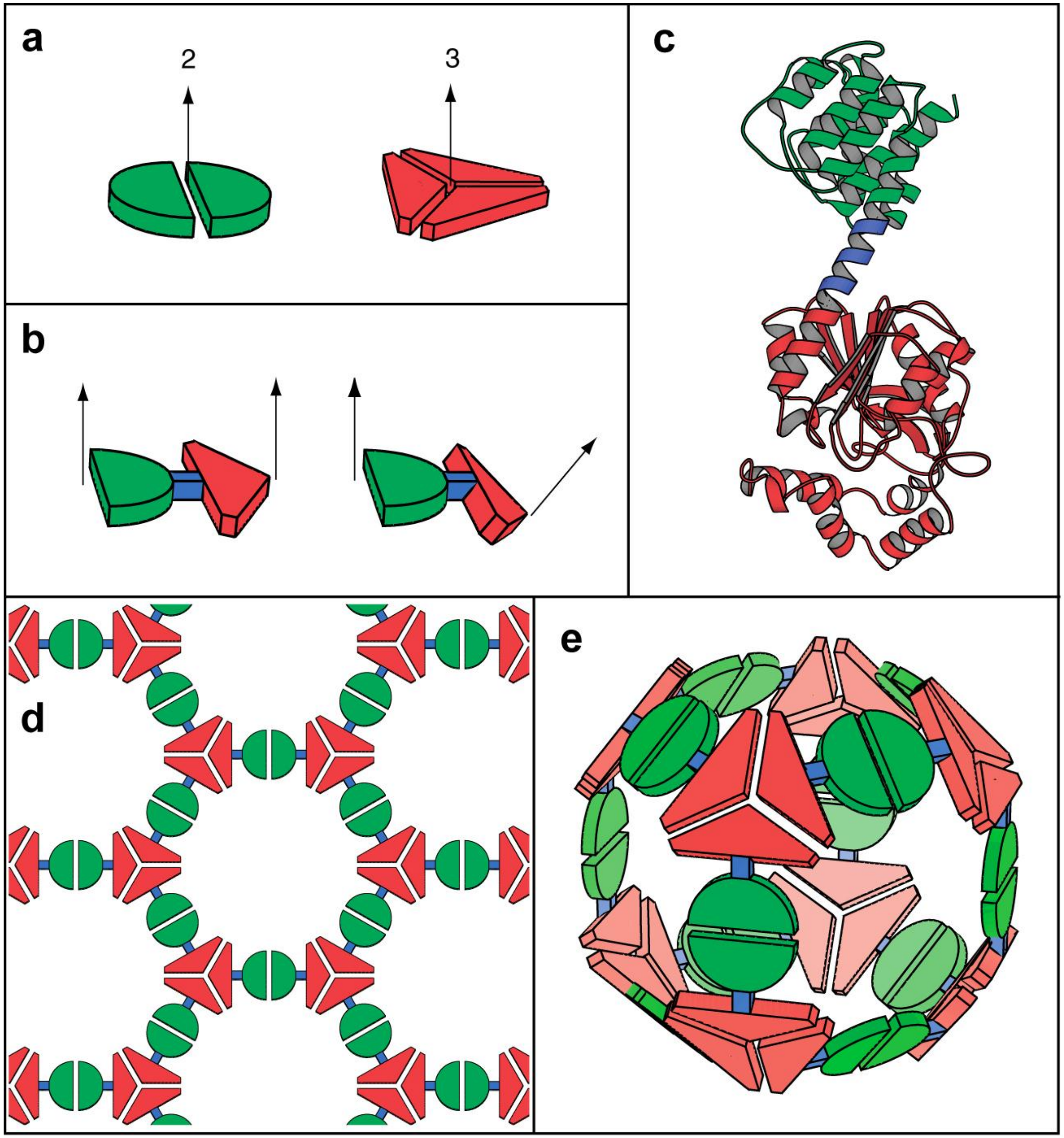




Regularly ordered self-assembly results when the relative orientation of the symmetry elements matches one of the known point, layer, or space groups.

A General Method for Designing Self-Assembling Protein Materials

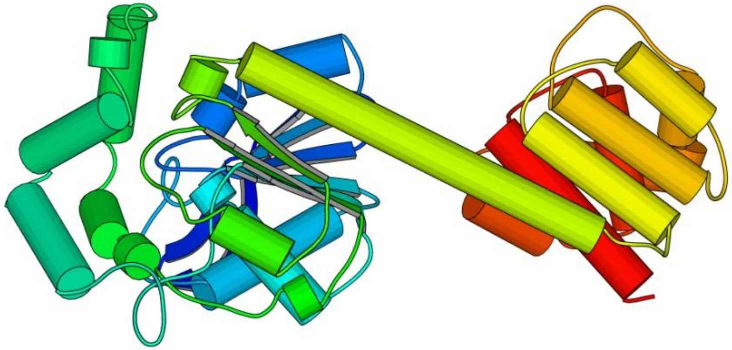
- Fusion of two simple oligomers (e.g. dimer + trimer)
- **Use of a continuous α-helix to dictate geometry**
- Satisfies predictability req., though not freely designable. (combinatorial)

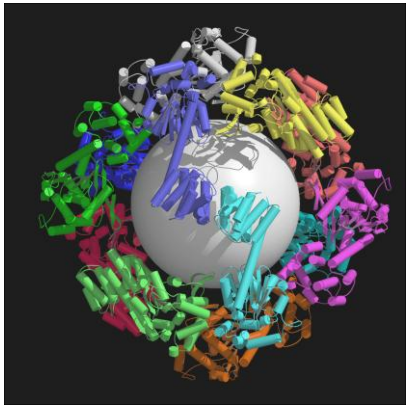


Padilla, Colovos, & Yeates, *PNAS*, **98**, 2217 (2001)

1st Design

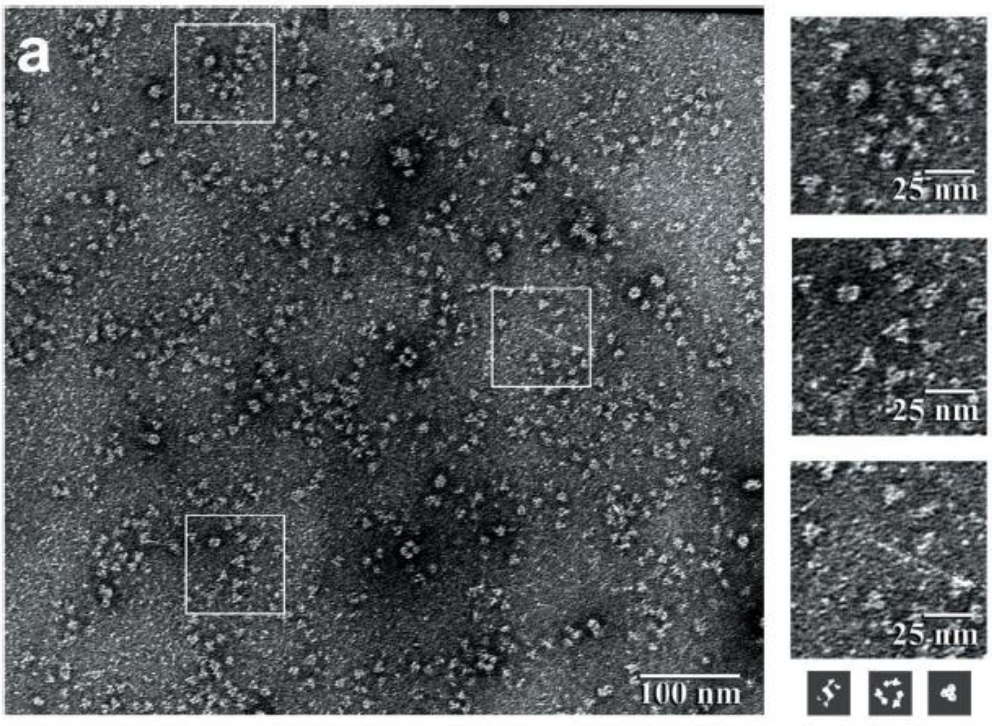
- **Intended architecture: tetrahedral cage (T), 12 subunits, 170 Å diameter, 1/2 MDa**
- **Components:**
  - **bromoperoxidase, 276 aa (trimer)**
  - **a helical linker (9 residues from L9 ribosomal protein)**
  - **influenza M1 coat protein, 150 aa (dimer)**
- **Symmetry element geometry:**
  - **angle between 2-fold and 3-fold: 53.2° [ideal = 54.7°]**
  - **failure to intersect: 2.8Å [ideal = 0.0]**
- **Expressed and purified in soluble form from *E. coli* (48 kD)**

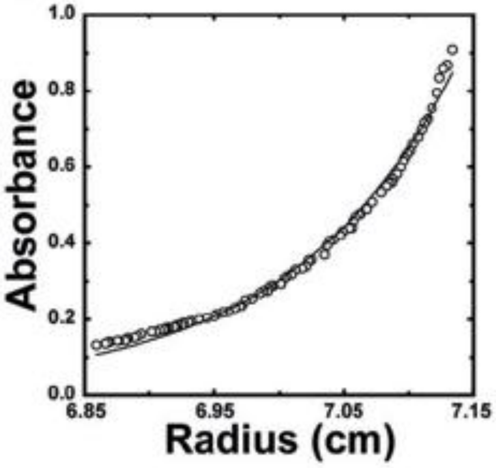


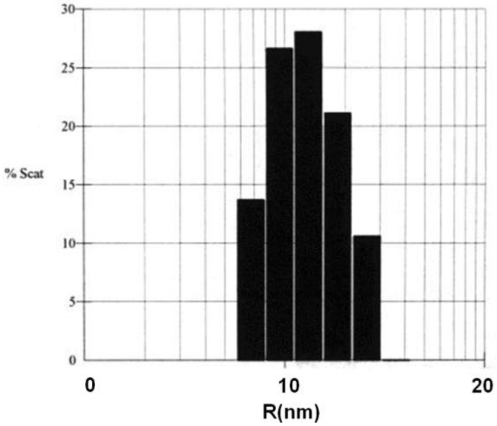


Padilla, Colovos, & Yeates, *PNAS*, **98**, 2217 (2001)

1st Design – A partial success

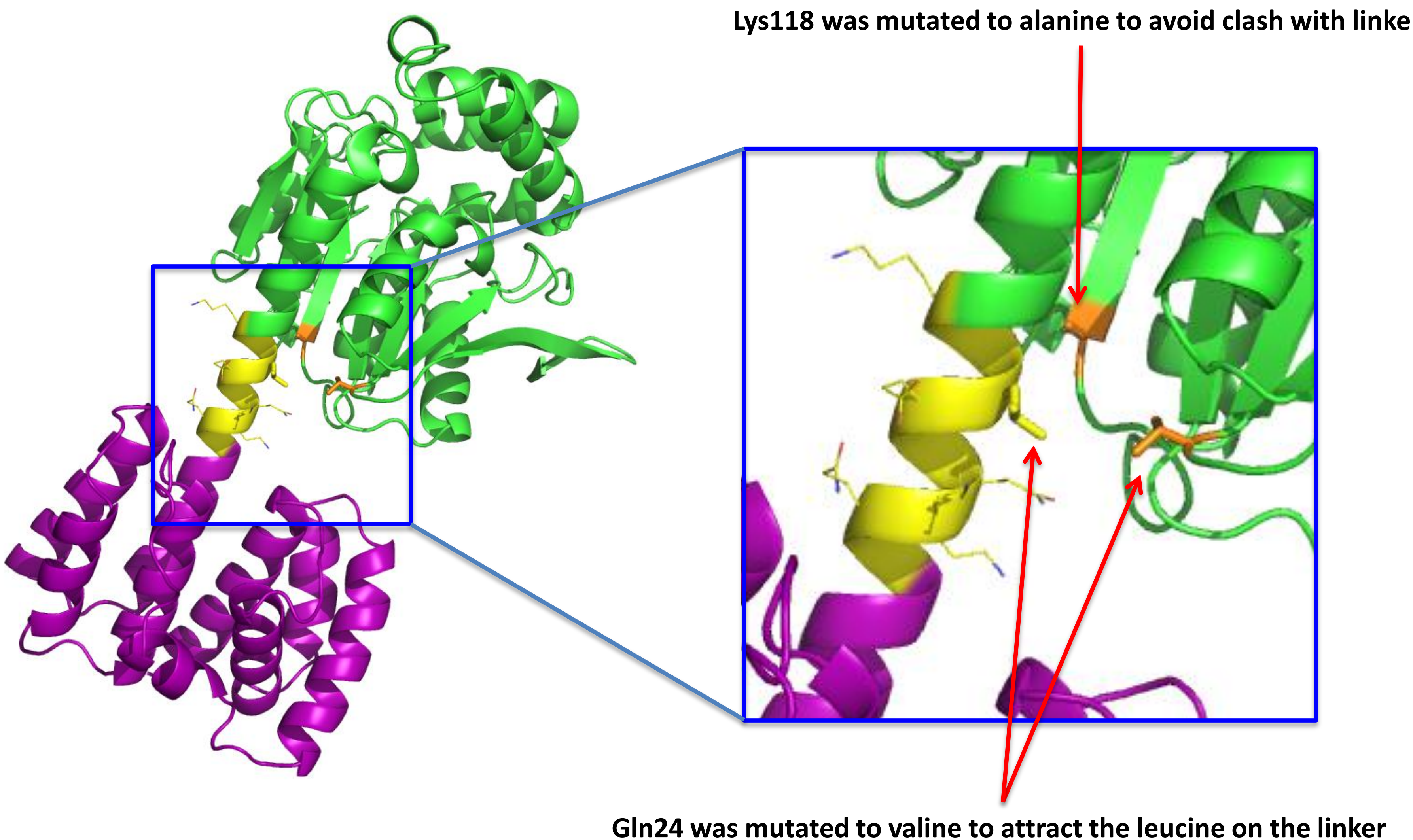




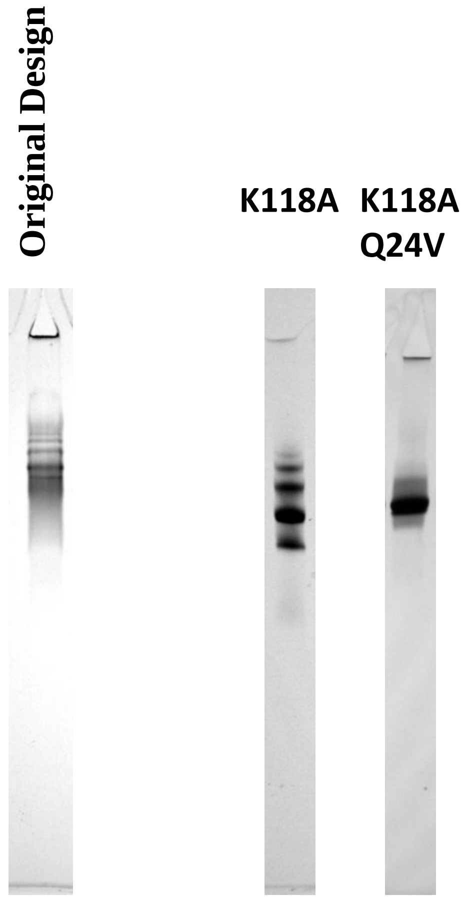


Discrete particles of approximately the right size, but polymorphic. Crystals never obtained!

Original Design Revisited…11 years later:
Two mutations

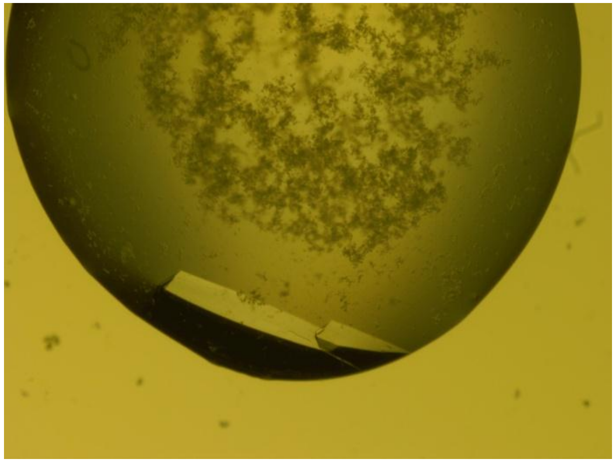


Original Design Revisited…11 years later: Two mutations



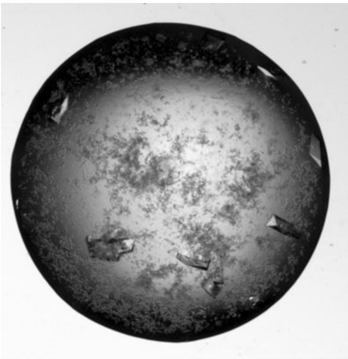
7.5% Native gel
(not on the same gel)

Native gels provide the guide for obtaining homogeneous assemblies, and crystals!

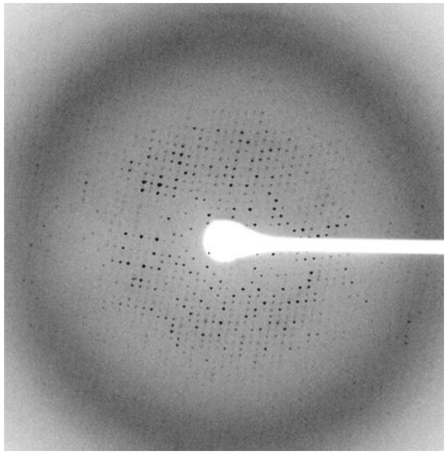


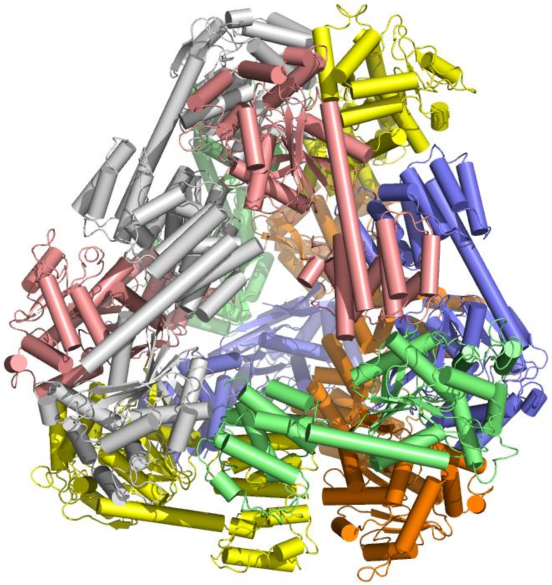
A first atomic structure of a designed protein cage

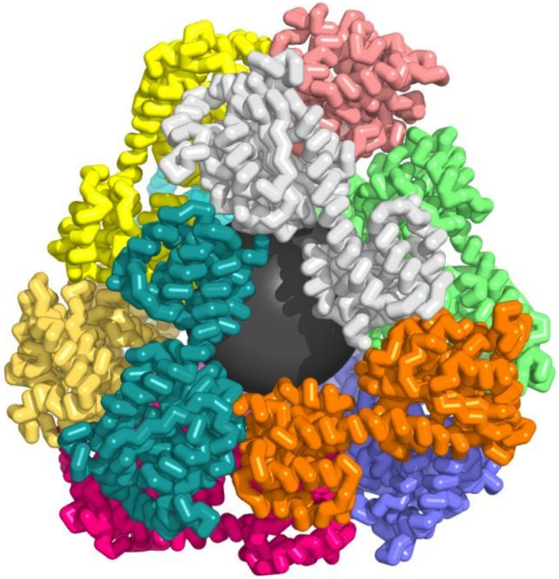
(~11 years after publication of idea and preliminary experiments)



3 Å resolution



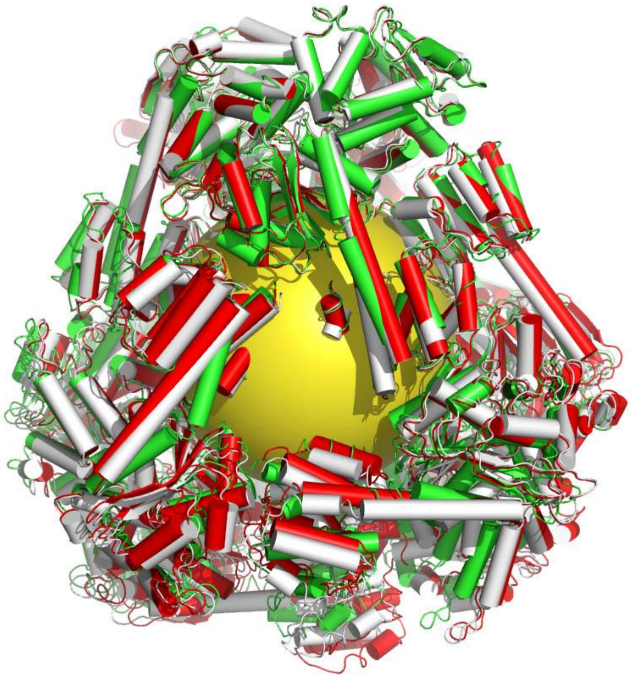


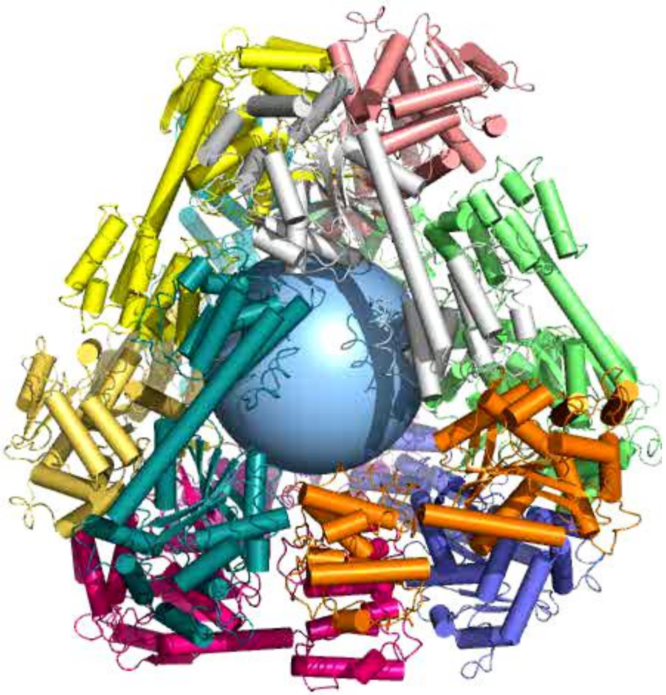


- **12 subunits**
- **Pseudo-tetrahedral symmetry**
- **Partially flattened (crystal packing and weak helical linker)**

Lai, et al. (2012). *Science* **336,** 1129**.**

Three independent cages in two crystal forms



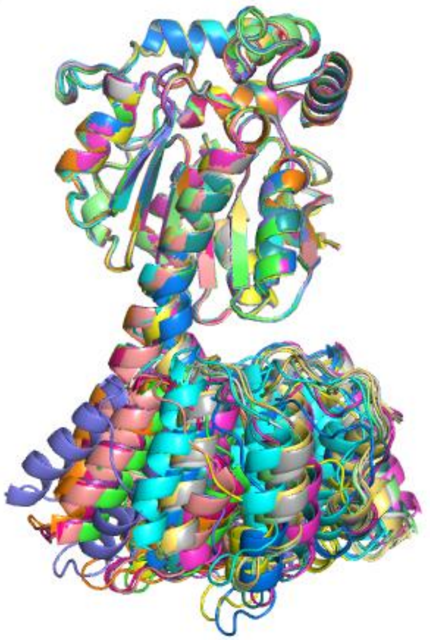


70 Å diameter (hypothetical) inner sphere

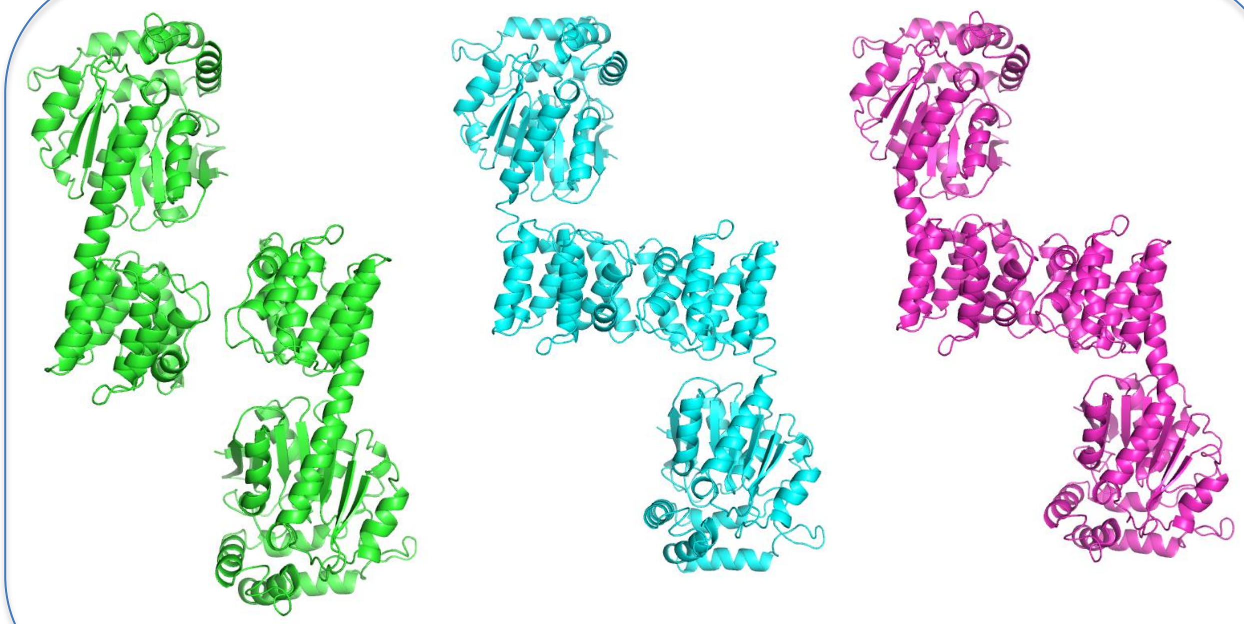
Lai, et al. (2012). *Science* **336,** 1129**.**

Surprisingly large deviations from symmetric design (~8 Å)

Distorted helices



Distorted dimeric interfaces

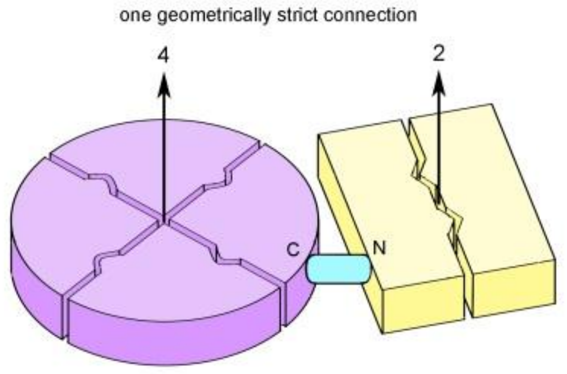


Not surprising in retrospect. Interface polymorphism was revealed after initial design choice. (Luo, et al.)

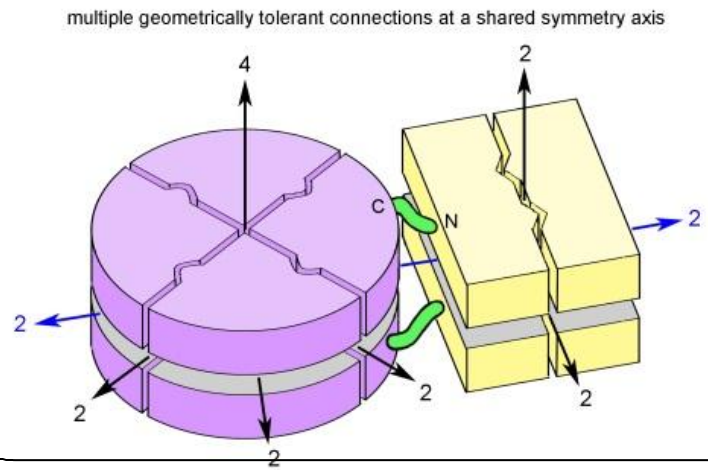
New results by others in symmetry-based design of complex protein assemblies/materials:

A variation on the oligomer fusion strategy

Sinclair JC, Davies KM, Vénien-Bryan C, Noble ME. (2011). *Nat. Nanotechnol.* **6**, 558-62.



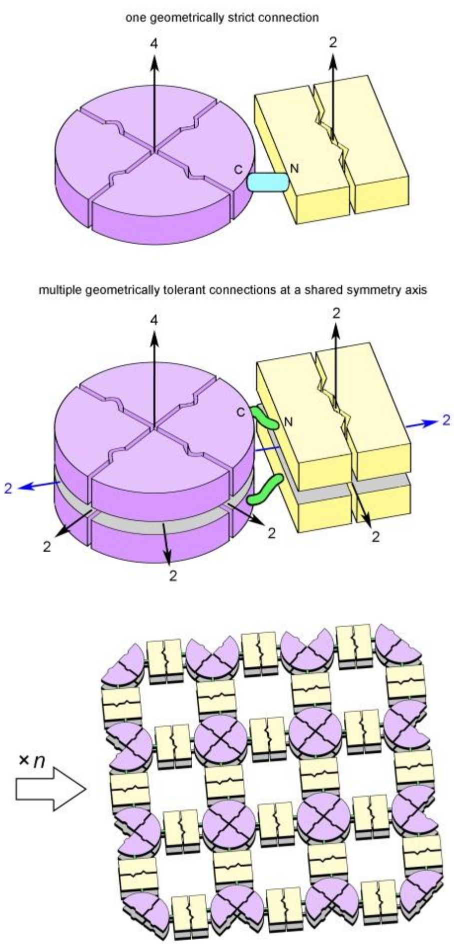
C4+C2



D4+D2 with a shared symmetry axis

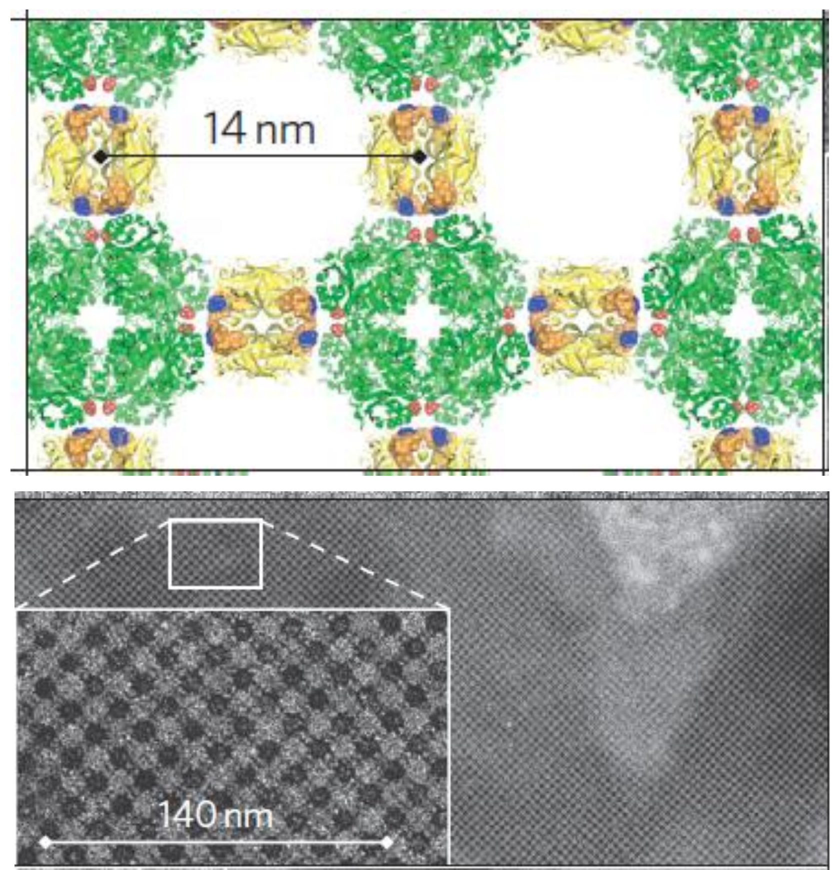
New results by others in symmetry-based design of complex protein assemblies/materials:

A variation on the oligomer fusion strategy



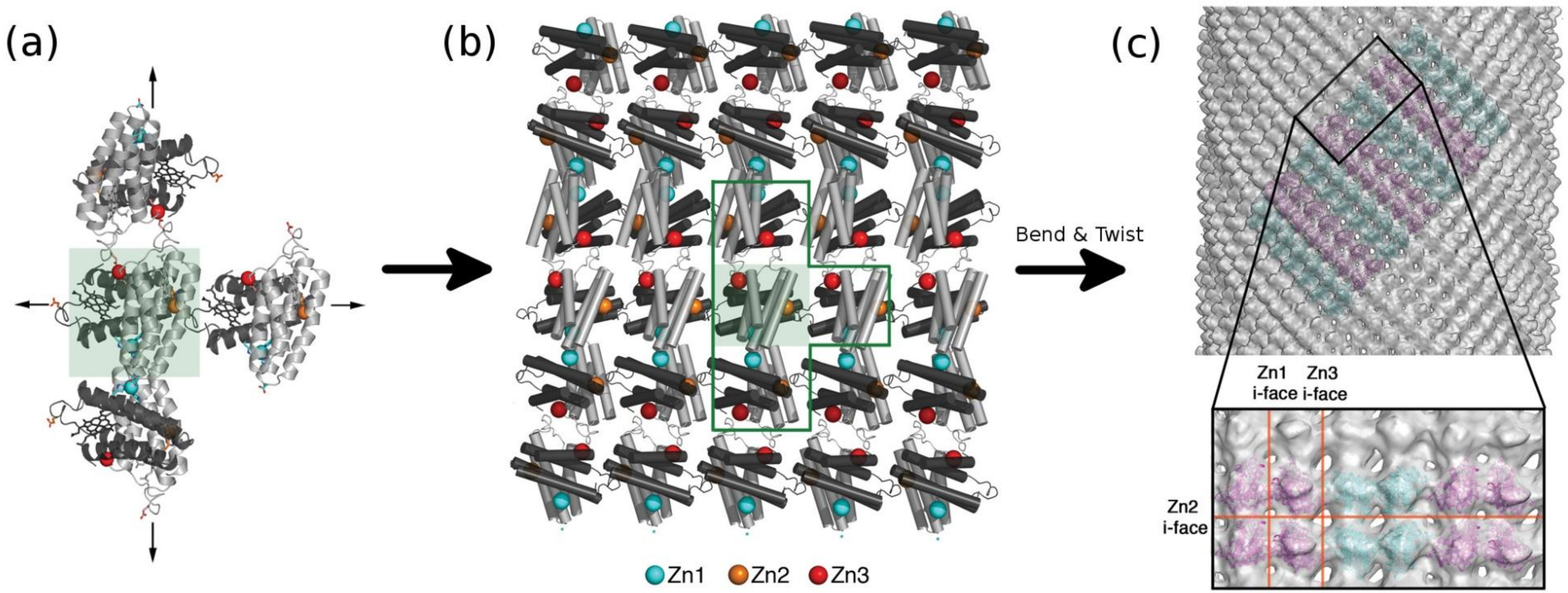
Yeates, TO. (2011) *Nat. Nanotechnol*. **6**, 541-2

Sinclair JC, Davies KM, Vénien-Bryan C, Noble ME. (2011). *Nat. Nanotechnol.* **6**, 558-62.



New results by others in symmetry-based design of complex protein assemblies/materials:

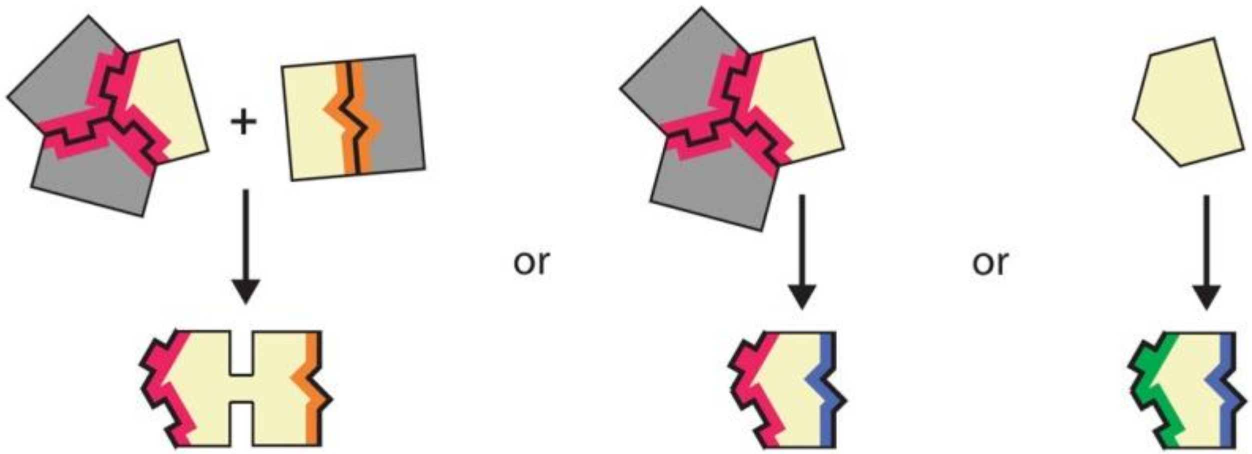
Metal interface design



Brodin, J.D., et al. (2012) Nat Chem 4, 375-382

New results by others in symmetry-based design of complex protein assemblies/materials:

Introducing new interface(s) by sequence design

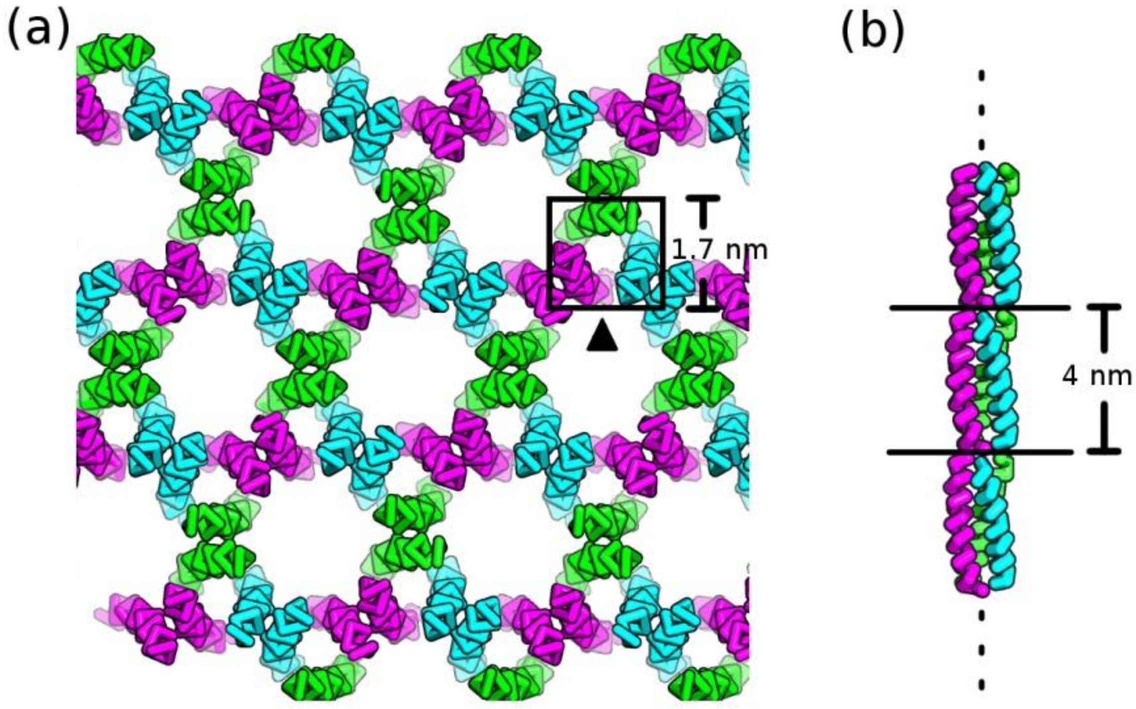


fusion → **design** →

New results by others in symmetry-based design of complex protein assemblies/materials:

Introducing new interface(s) by sequence design

Design of a protein crystal based on coiled coil motifs

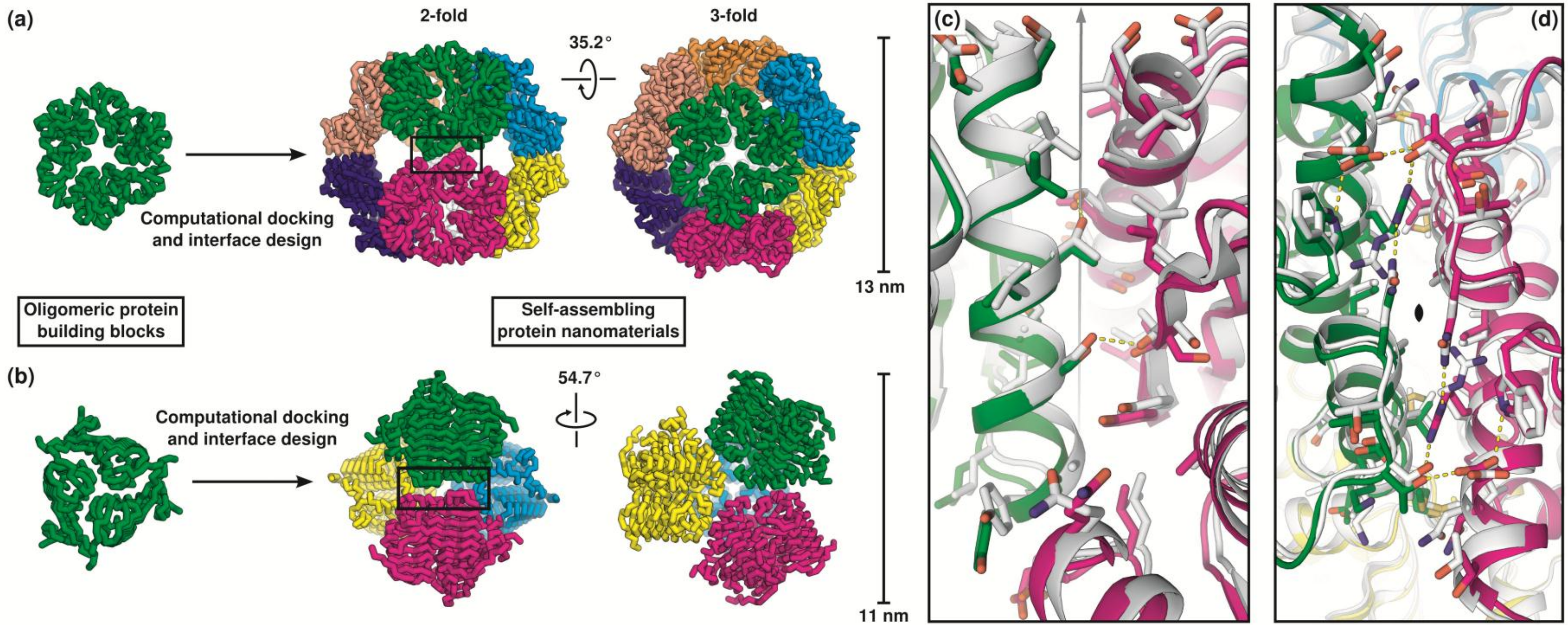


Lanci, C.J., *et al.* (2012) *PNAS* **109**, 7304-7309

New results by others in symmetry-based design of complex protein assemblies/materials:

Introducing new interface(s) by sequence design

Design of protein cages and cubic assemblies based on ‘general’ oligomers and sequence design



King, N.P., et al. (2012). *Science* **336**, 1171-1174.

Future Directions

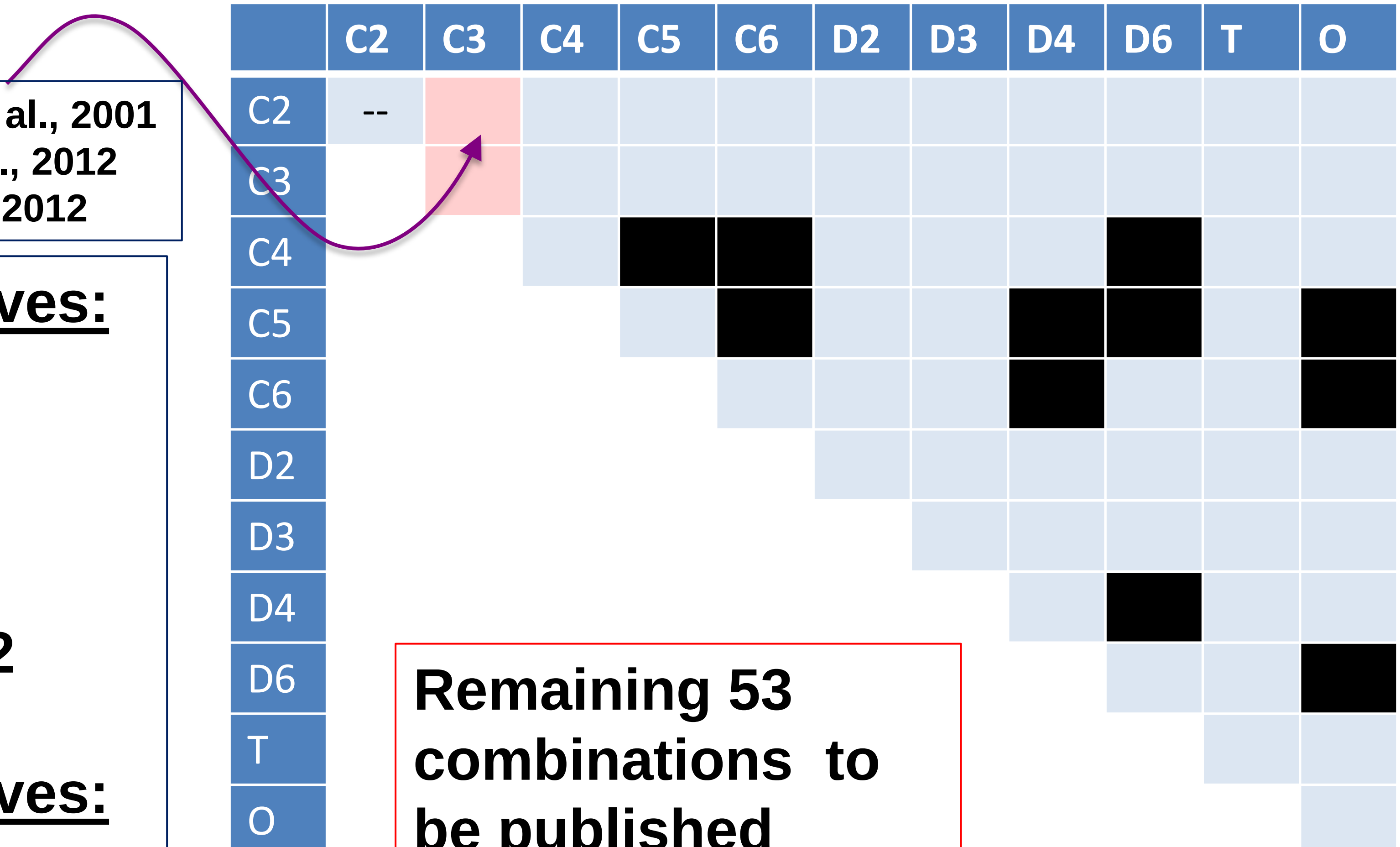
- Algorithm/strategy improvements; success rates remain low
- Theoretical enumeration of complete rules and possible outcomes
- Biomedical and nanotechnology applications
    - display (e.g. vaccine), containment/delivery, bioactive (e.g. enzymatic) solids and surfaces

Possible combinations of two symmetry point groups and their assembly outcomes

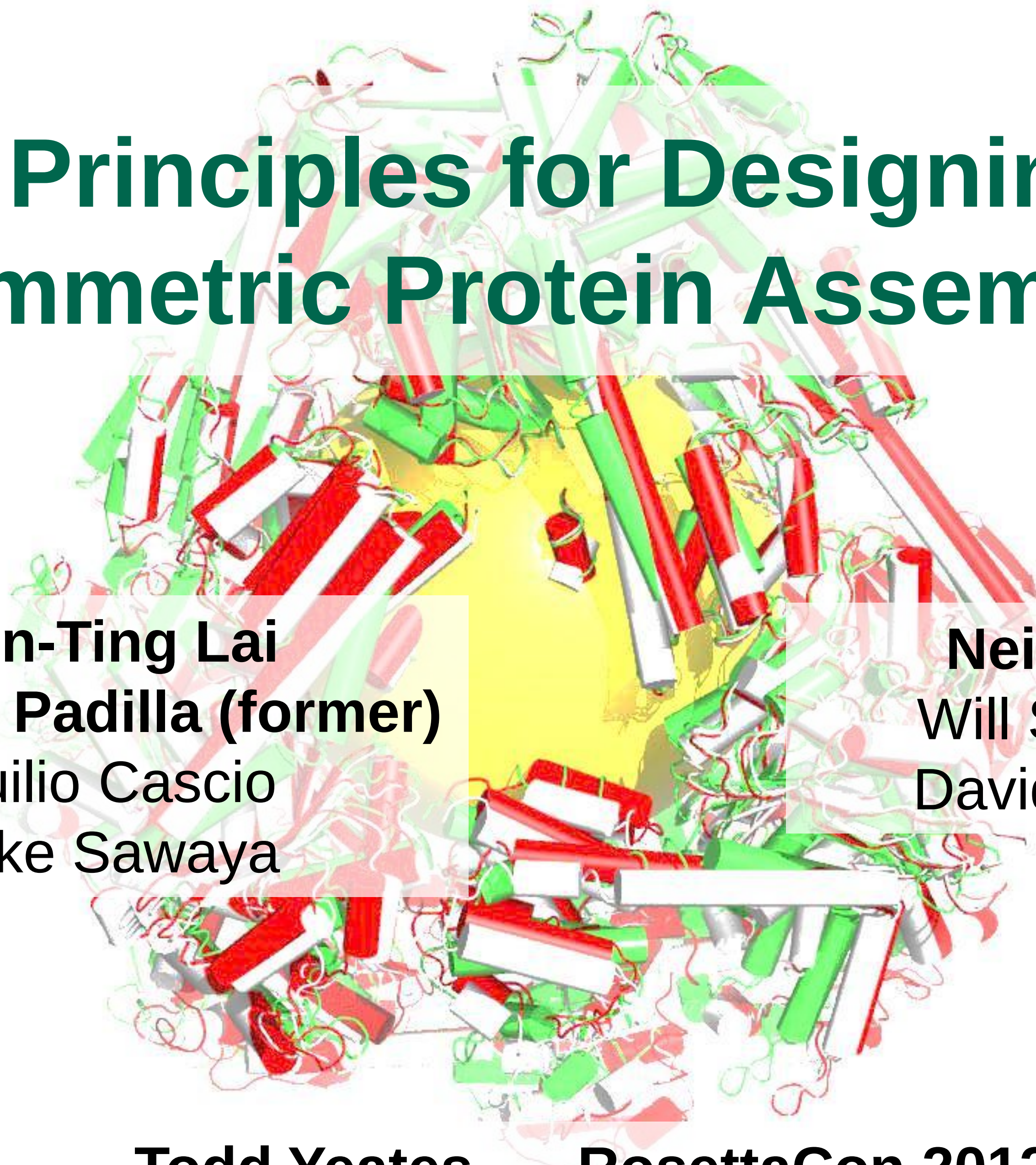
Padilla, et al., 2001
King, et al., 2012
Lai et al.,, 2012

<u>2+3 gives:</u>
T, O, I
p6
p321
$I2_13$
$P4_{1,3}32$

<u>3+3 gives:</u>
p3
P23



Remaining 53 combinations to be published



Principles for Designing Symmetric Protein Assemblies

Yen-Ting Lai
Jennifer Padilla (former)
Duilio Cascio
Mike Sawaya

Neil King
Will Sheffler
David Baker

Todd Yeates -- RosettaCon 2012