



Integrated SBIO: everything but the kitchen sink

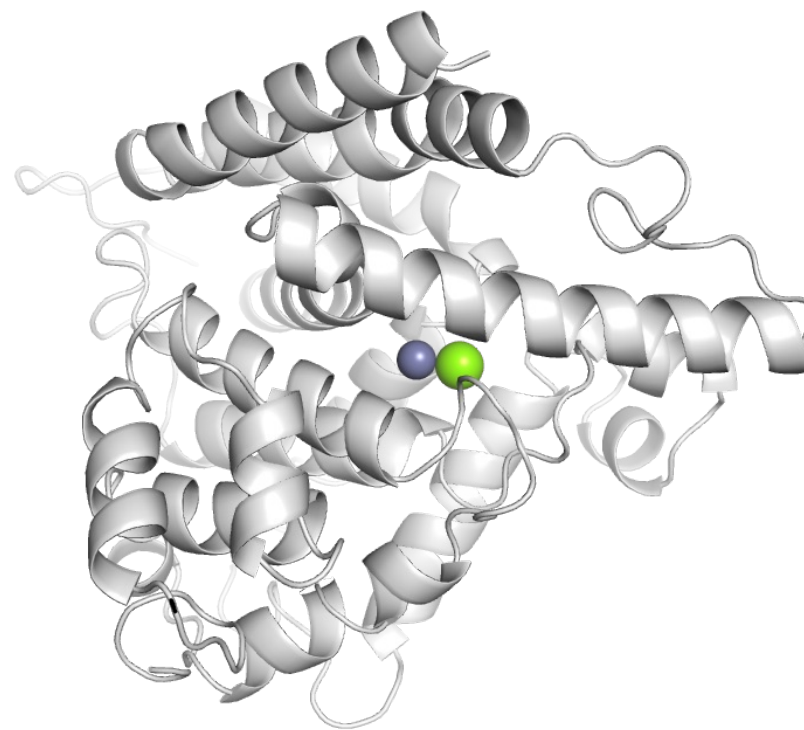
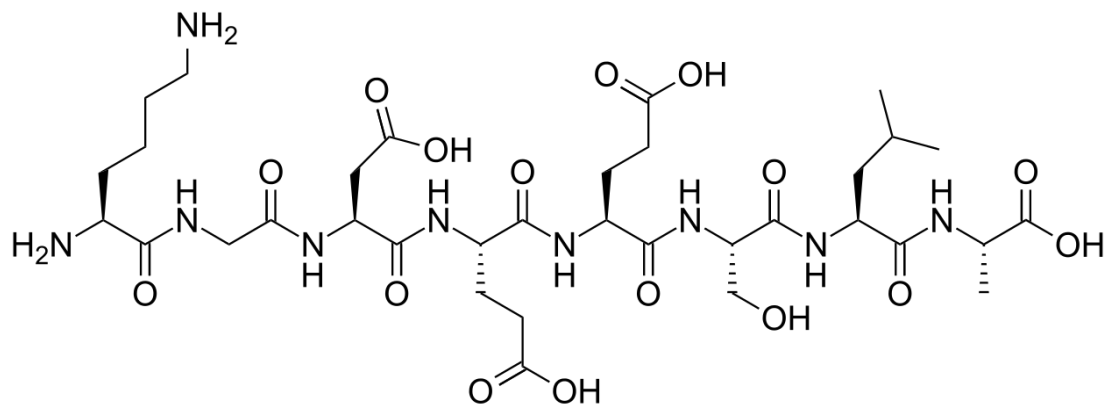


AiChemist + AIDD
04 March 2024
John O'Donnell

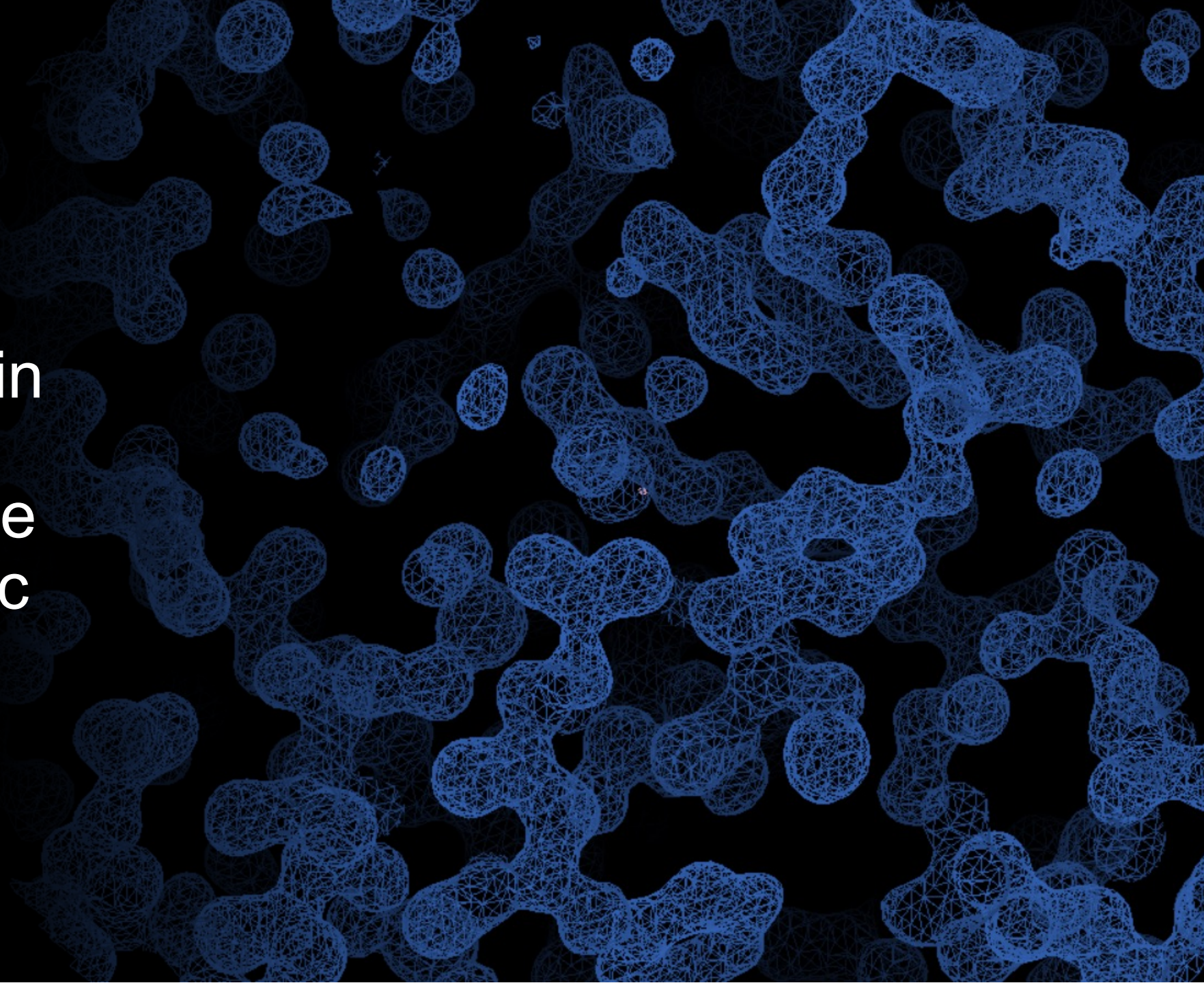


We experimentally determine three-dimensional structures at the atomic level

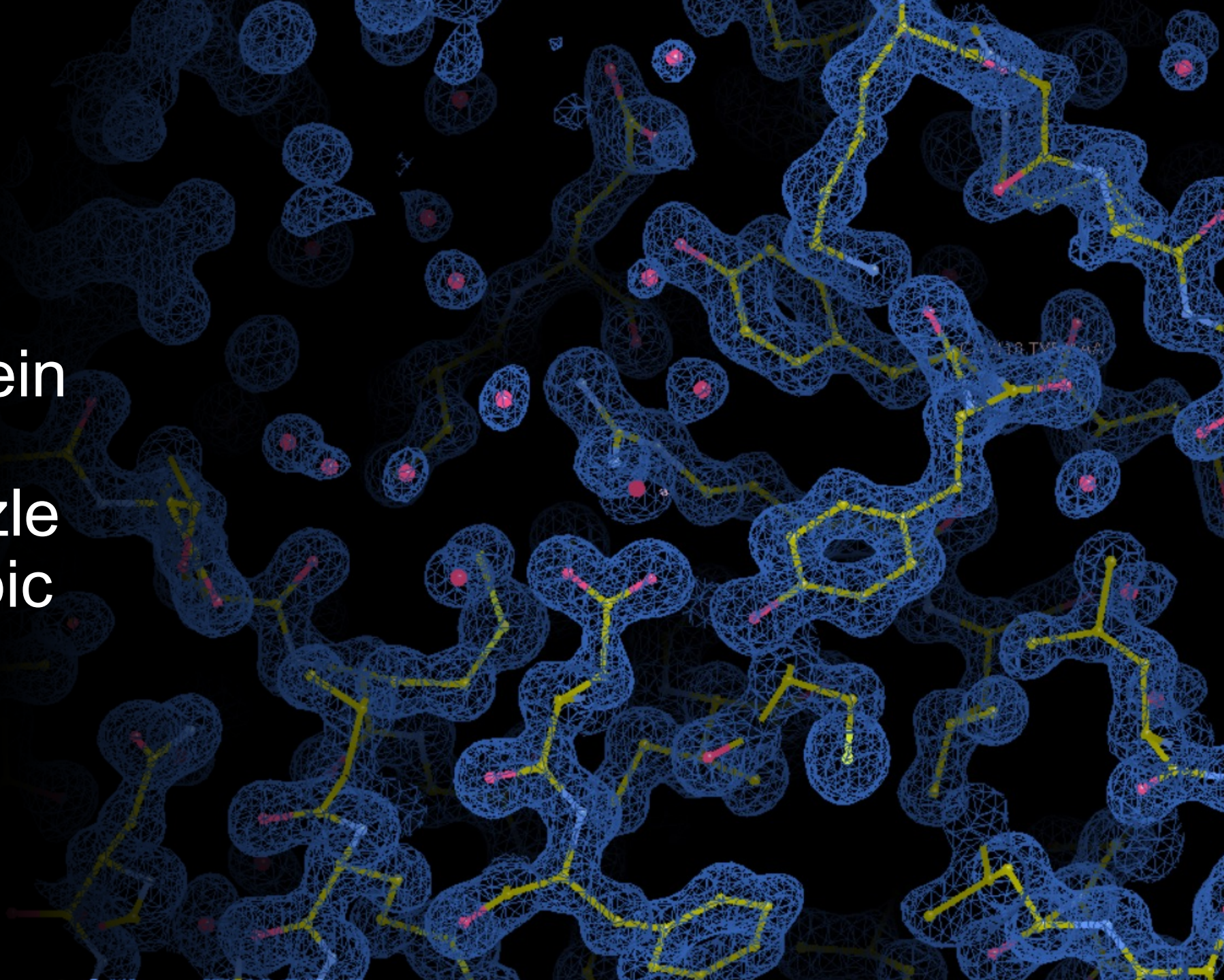
Biological macromolecules are the machines that biology relies on for proper function. As a team of structural biologists, we elucidate the atomic 'blueprints' of these targets to understand their function and how therapeutics we develop can modulate their activities



Building a protein
structure is an
enormous puzzle
on a microscopic
level



Building a protein structure is an enormous puzzle on a microscopic level





Defining molecular mechanisms with structural biology

Structural data informs project work from atoms to disease mechanism

Our mission



Understand complex target biology

- // Decipher molecular mechanism to tailor lead finding strategy
- // Discover druggable sites
- // Mode-of-action studies elucidate how a therapeutic is modulating biology

Drive lead understanding to guide optimization

- // Define binding modes of tool and competitor compounds and enable fast-track computational approaches
- // Structurally guide lead optimization

Finding and understanding leads

Modifying and improving leads

Target Evaluation

D1

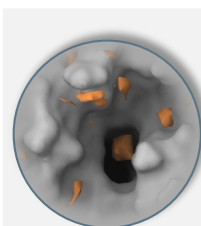
Lead Generation and Lead Evaluation

D2

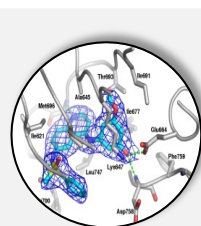
Lead Optimization

PCC

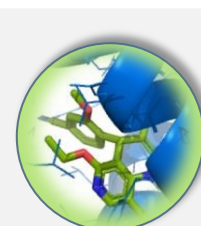
Development



Druggability assessments and mechanistic understanding

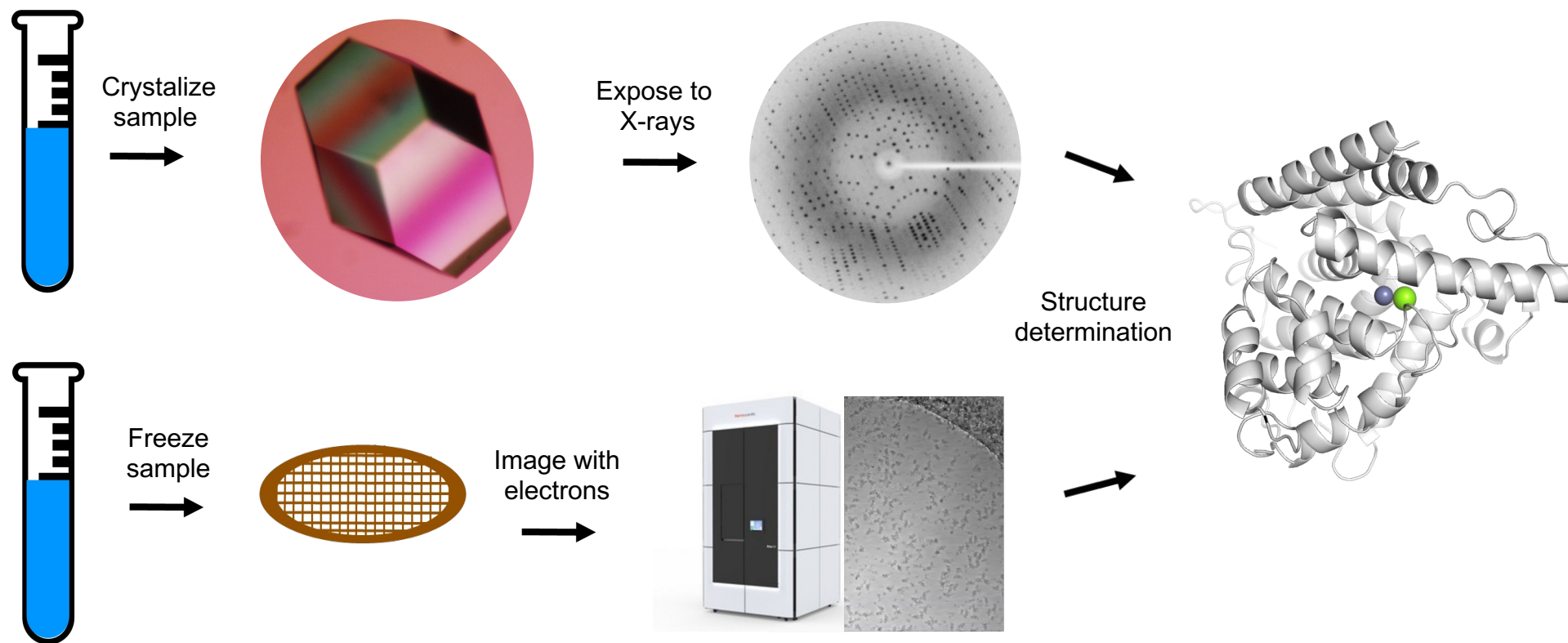


Structurally enable lead finding strategy
e.g. virtual screening, fragment screening, high throughput screening



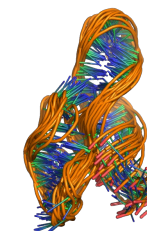
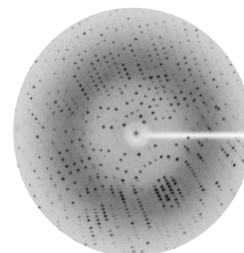
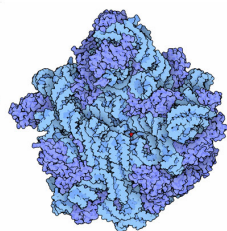
Binding mode determination to enable structure-based optimization

We use X-rays and electrons for structure determination



Overview of structural biology techniques

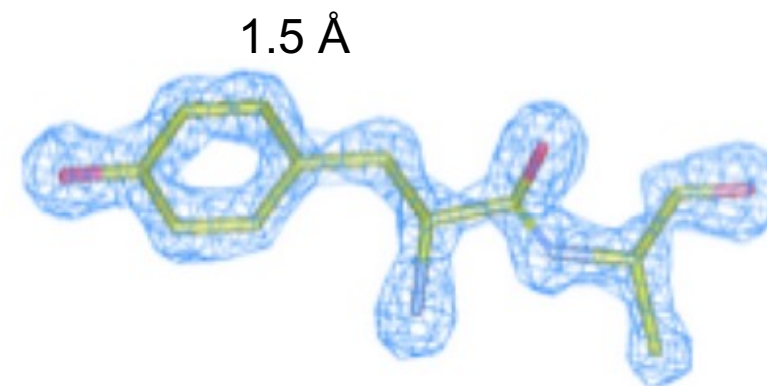
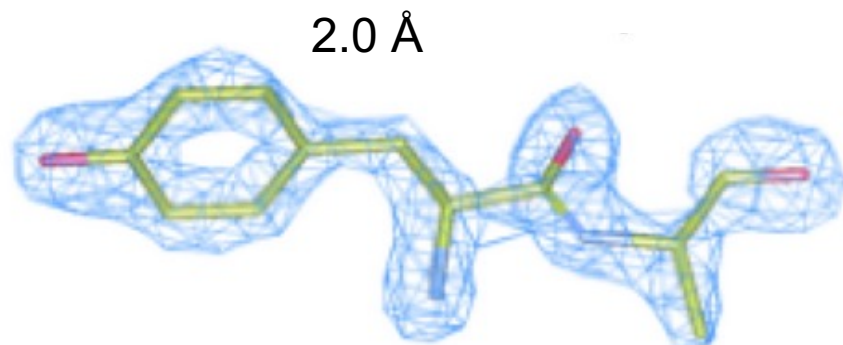
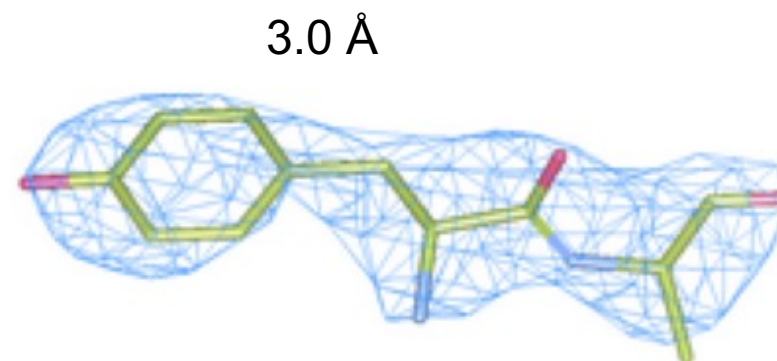
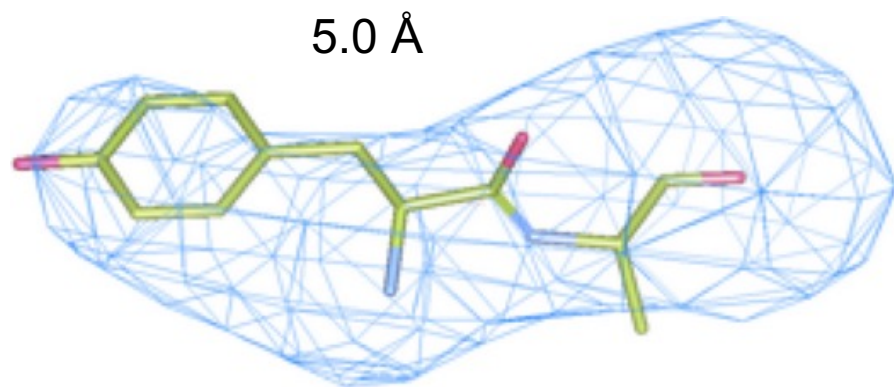
Comparing cryo-EM, X-ray crystallography, and NMR



	Cryo-EM	X-ray crystallography	NMR
Sample types	<ul style="list-style-type: none"> // Membrane proteins // Large complex proteins // Soluble proteins 	<ul style="list-style-type: none"> // Crystallizable samples // Largely limited to soluble proteins 	<ul style="list-style-type: none"> // Proteins with MW <50 kDa
Advantages	<ul style="list-style-type: none"> // Only requires small sample size // Structures are obtained in native state 	<ul style="list-style-type: none"> // High resolution // Broad MW range // Established technique // Moderate throughput // Routinely resolve small molecules and water networks 	<ul style="list-style-type: none"> // Obtains 3D structures in solution // Information about dynamics // Suited for RNAs
Current limitations	<ul style="list-style-type: none"> // Proteins with molecular weights >100 kDa are most feasible // Routine resolutions are not as high as X-ray crystal structures // Costly, but getting cheaper 	<ul style="list-style-type: none"> // Crystallization can be difficult or not possible // Results are in static crystalline state // Diffraction can be difficult 	<ul style="list-style-type: none"> // Needs high purity sample // Has a difficult computational simulation // Sample must be isotopically labeled
Sample amounts	<ul style="list-style-type: none"> // Nanograms to micrograms 	<ul style="list-style-type: none"> // Micrograms to milligrams 	<ul style="list-style-type: none"> // Micrograms to milligrams



Resolution and why it is important

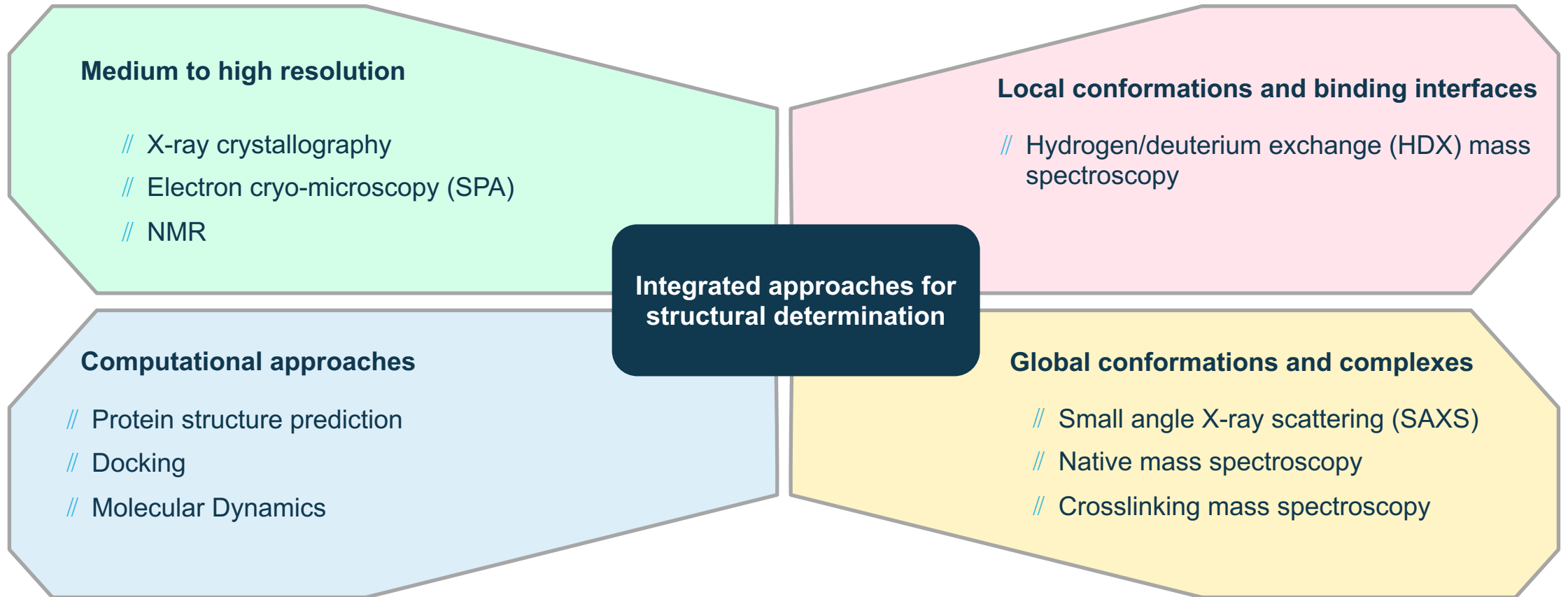


Kuster DJ, Liu C, Fang Z, Ponder JW, Marshall GR (2015) High-Resolution Crystal Structures of Protein Helices Reconciled with Three-Centered Hydrogen Bonds and Multipole Electrostatics. PLOS ONE 10(4): e0123146. <https://doi.org/10.1371/journal.pone.0123146>



Structural biology is not 'one size fits all'

Integrated approaches necessary for challenging targets and systems





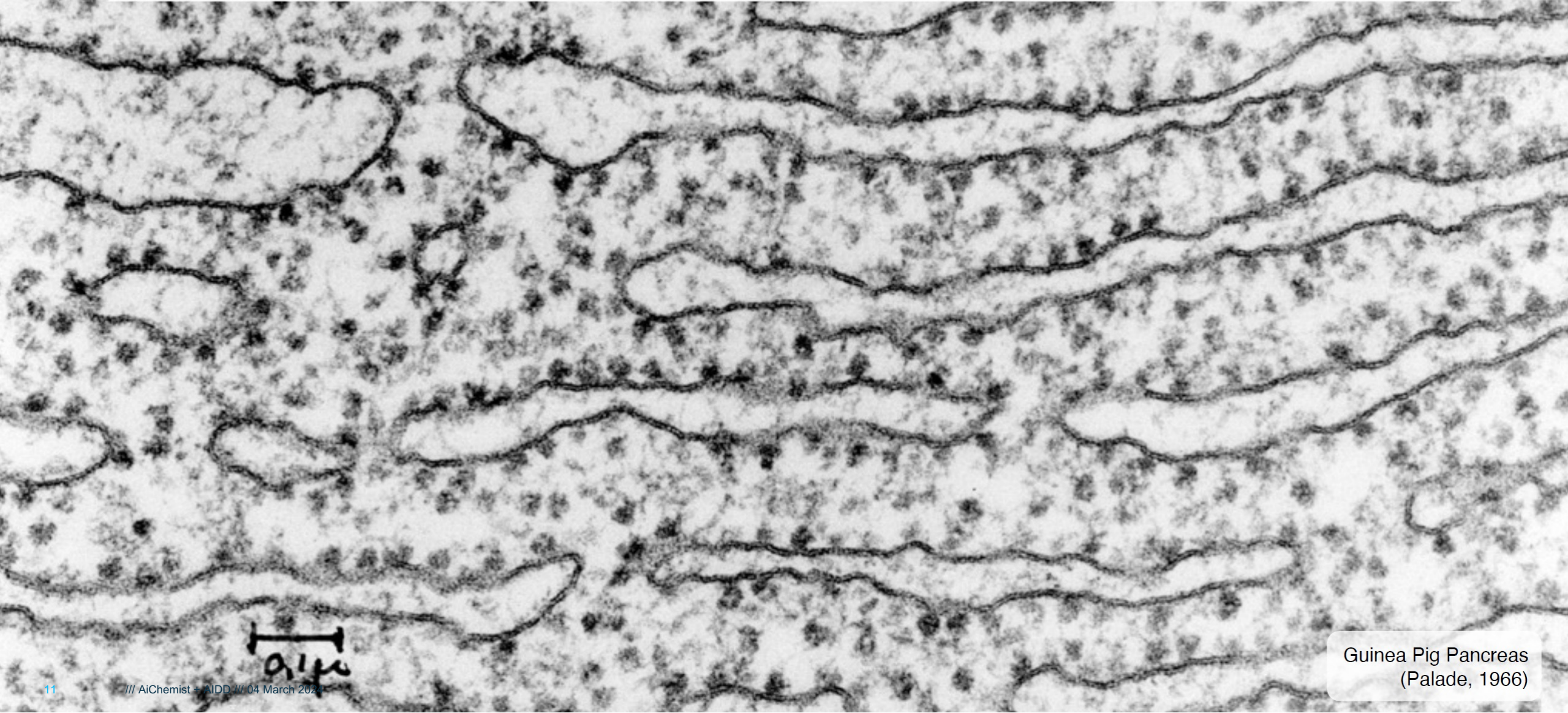
Case study for understanding protein structure-function

Highlight integrating SBIO techniques

- // Structure-function analysis of transmembrane domain insertase called the ER membrane protein complex (EMC)
- // CryoEM delivered a moderate resolution EM map at 6 angstrom, but not good enough to build a structure de novo.
- // Implemented numerous integrated structural biology techniques in addition to cell biology and biophysics.
 - // Integrated structure:
 - // CryoEM and crystallography
 - // Co-evolution to determine spatially linked residues
 - // Cell biology experiments to determine topology and #of TMDs in proteins
 - // Deep learning protein prediction methods (trROSETTA / AF2)
 - // Molecular dynamics flexible fitting with Flex-EM, Namdinator, and ISOLDE; normal mode analysis
 - // Biophysics:
 - // SEC-MALS
 - // Microscale thermophoresis
 - // nanoDSF
 - // Non-natural amino acid incorporation (BPA photocrosslinker) into recombinant protein and in vitro translations
 - // Cell biology:
 - // Cellular site specific photo-crosslinking
 - // Membrane protein expression in HEK freestyle cells
 - // Ratiometric FACS assay for tracking membrane protein biogenesis



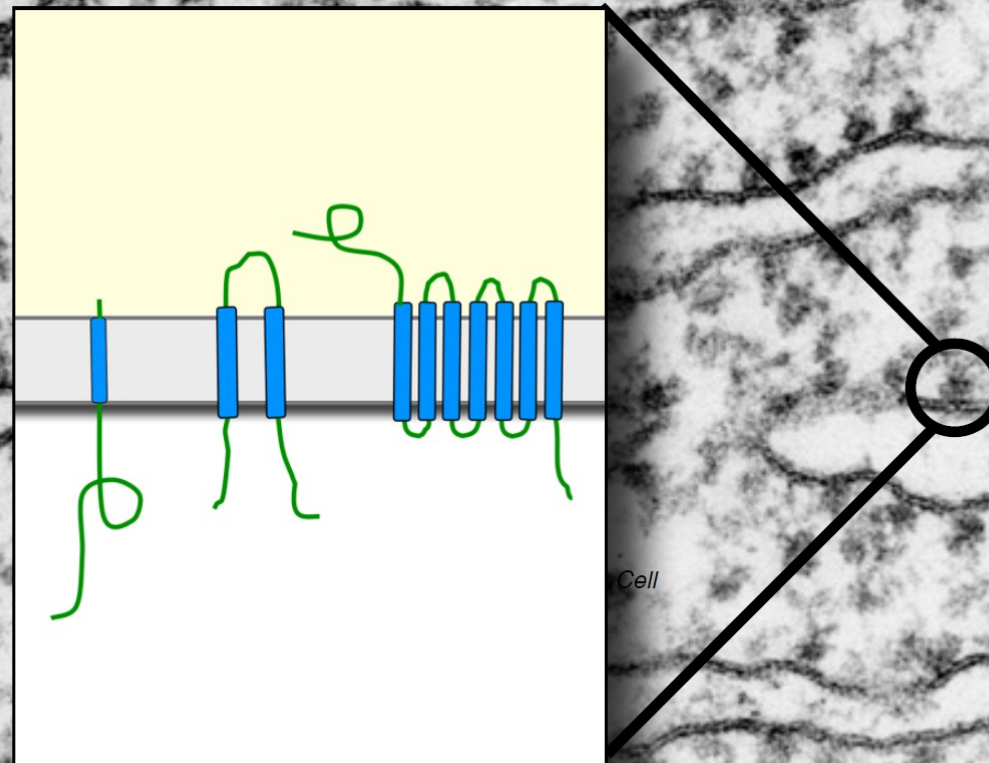
Membrane protein insertion is ubiquitous and essential



Guinea Pig Pancreas
(Palade, 1966)



Membrane protein insertion is ubiquitous and essential

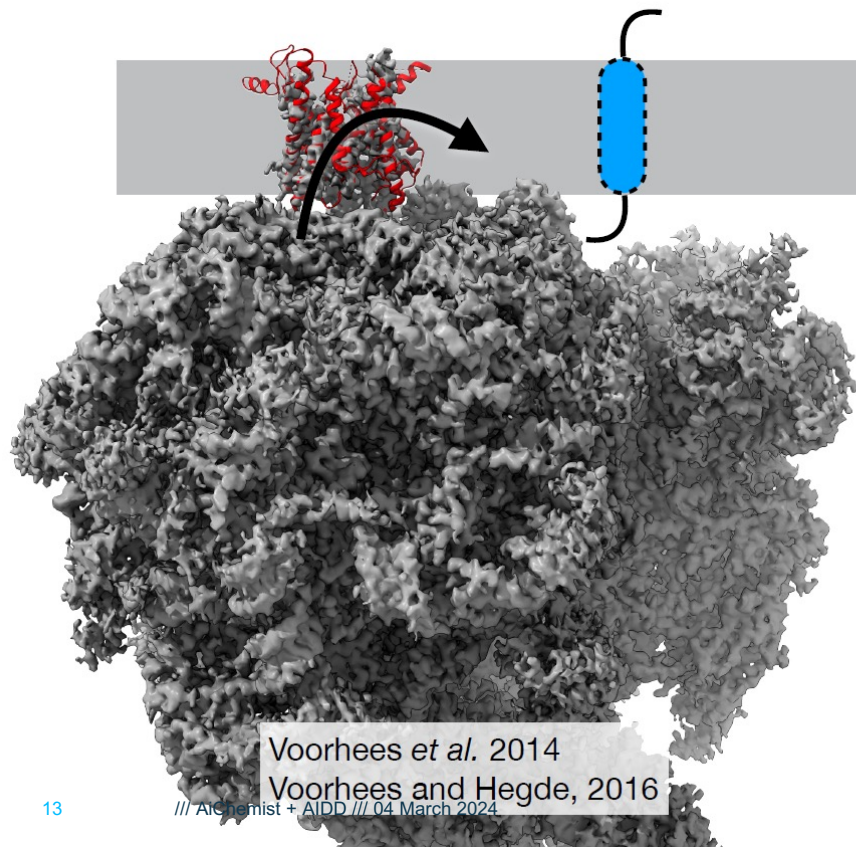


0.1 μm

Guinea Pig Pancreas
(Palade, 1966)

Insertion machineries

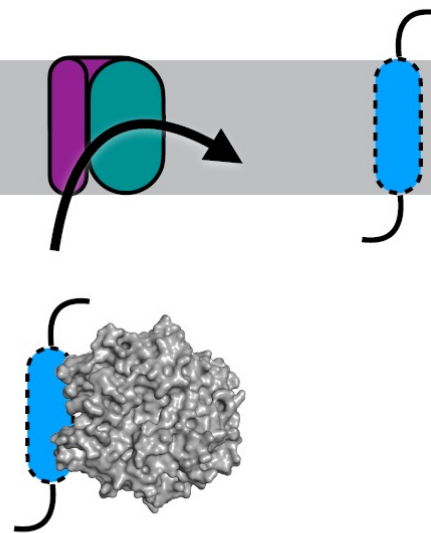
Sec-translocon



Voorhees *et al.* 2014
Voorhees and Hegde, 2016

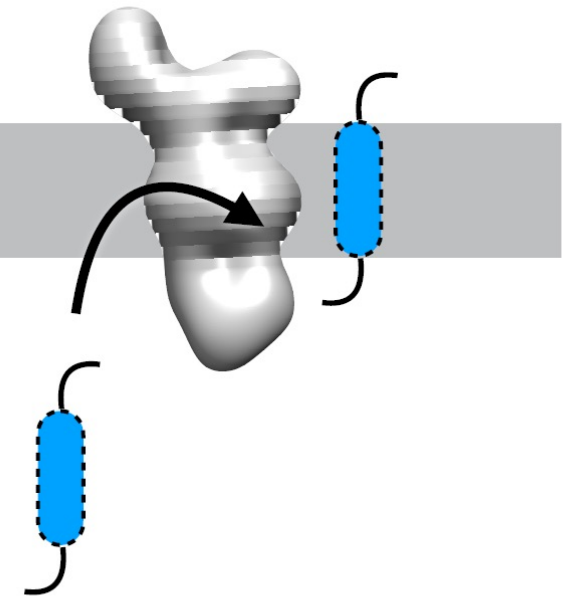
/// AIChemist + AIDD /// 04 March 2024

TRC/GET pathway



Mariappan *et al.* 2011
Mateja *et al.* 2015

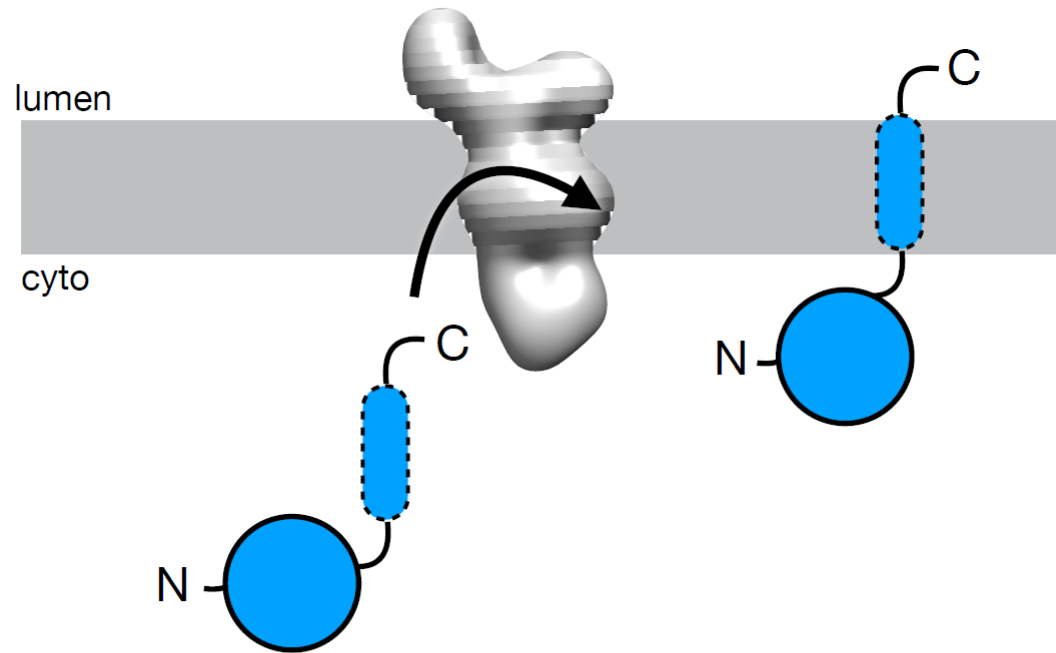
ER membrane protein complex (EMC)



Guna and Hegde, 2017
Chitwood *et al.* 2018

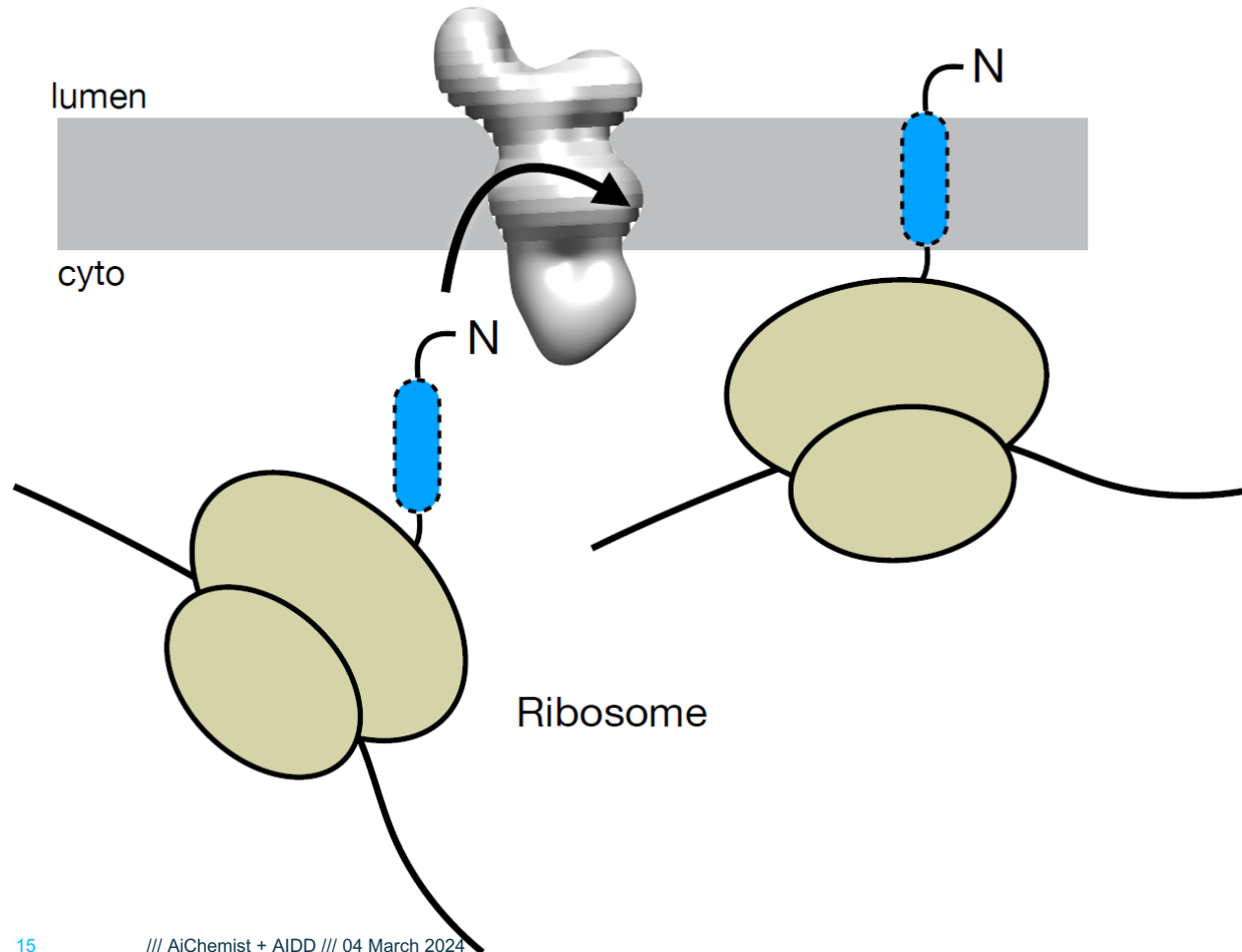


EMC is an insertase for terminal transmembrane domains



- Translocated luminal domain is very short
- Either topology is okay
- Energy-independent

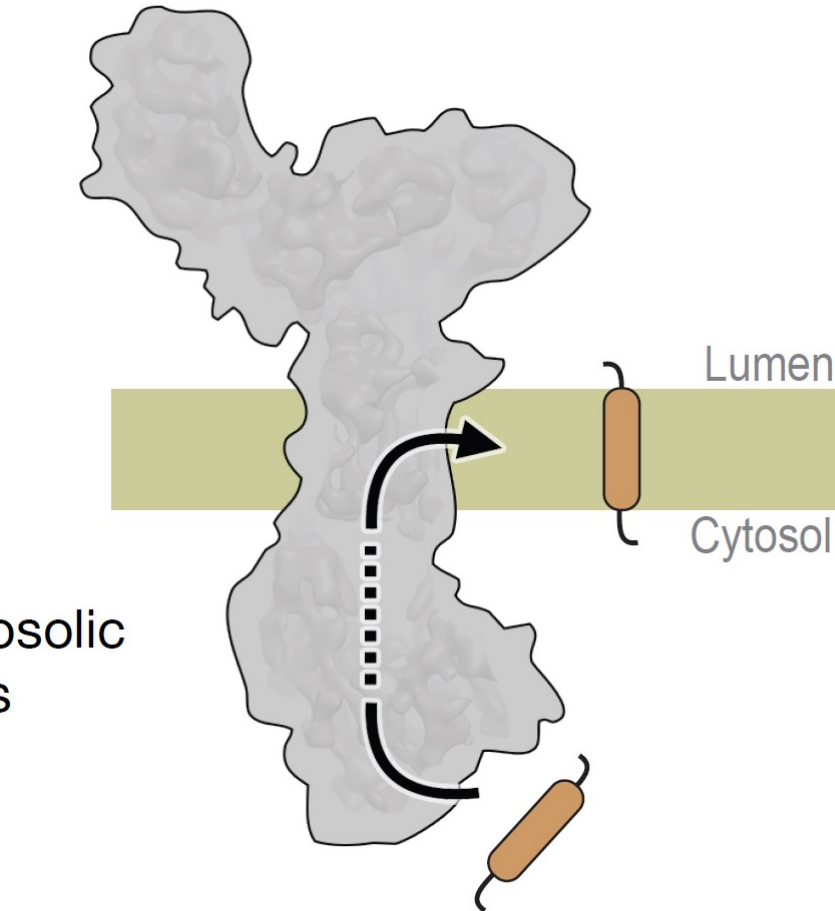
EMC is an insertase for terminal transmembrane domains



- Translocated luminal domain is very short
- Either topology is okay
- Energy-independent

How does EMC facilitate insertion

2. Pathway links EMC's cytosolic and membrane regions

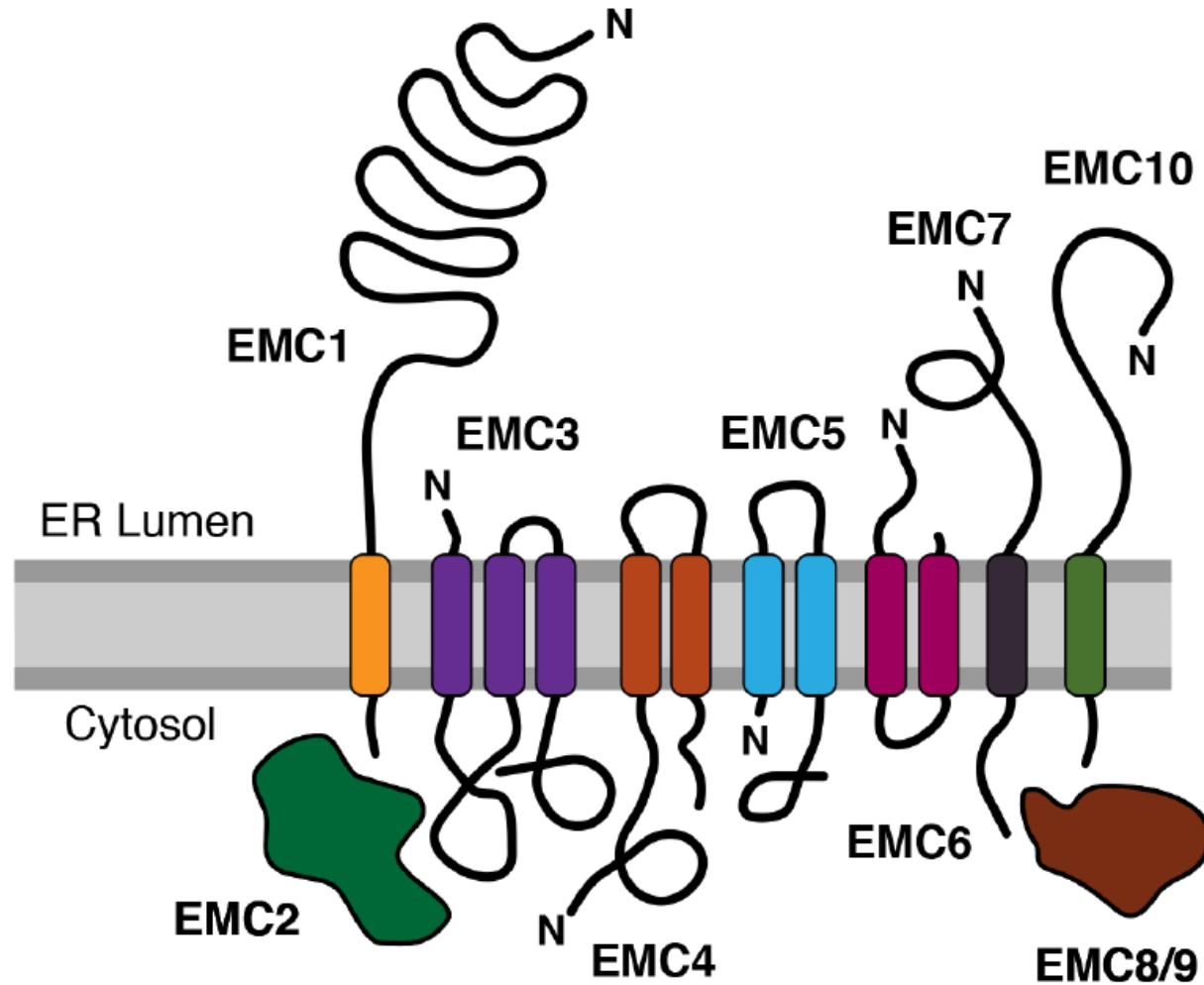


1. Cytosolic EMC subunits bind to client TMDs

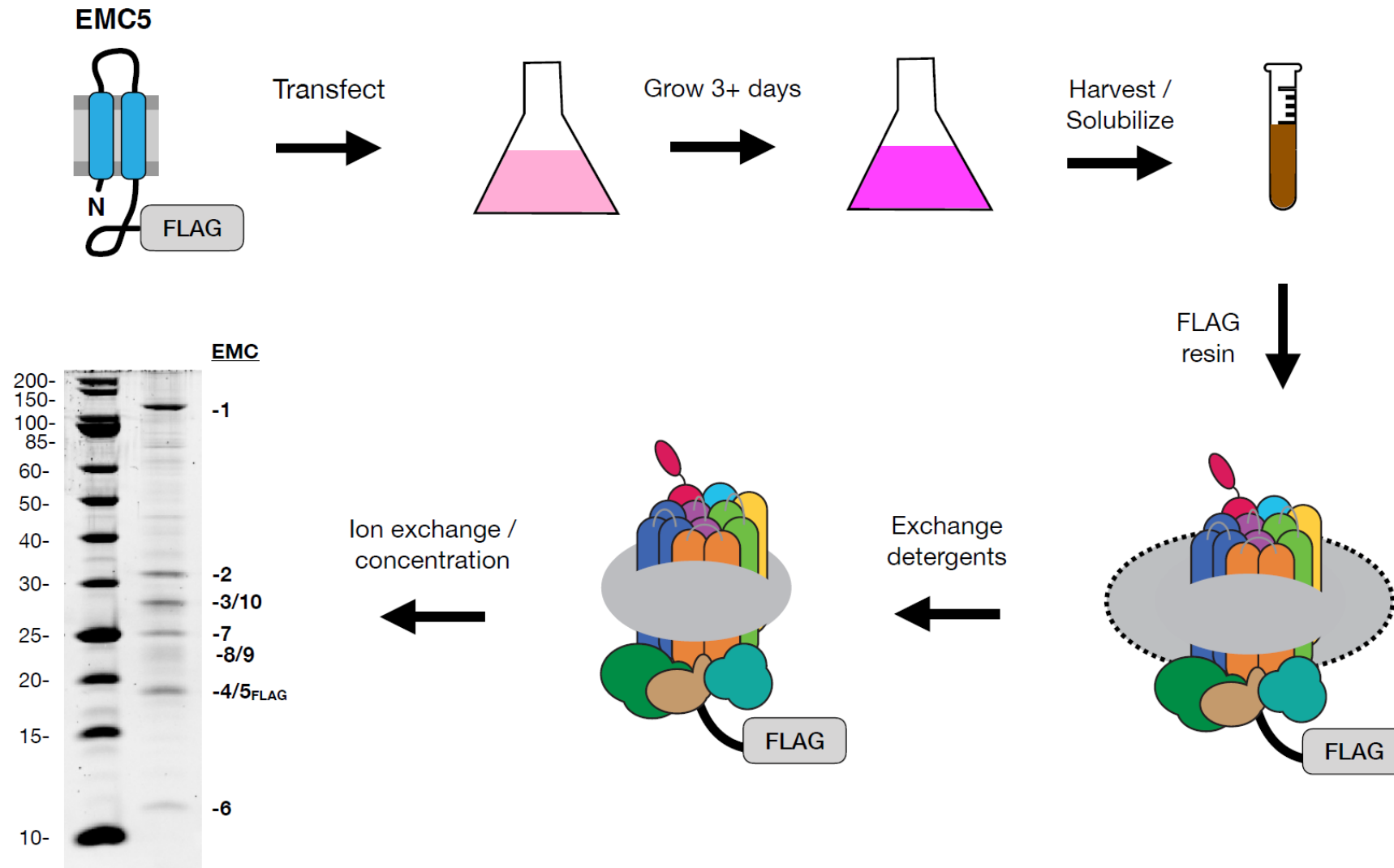
3. TMD architecture promotes insertion



EMC is a multi-subunit integral membrane protein



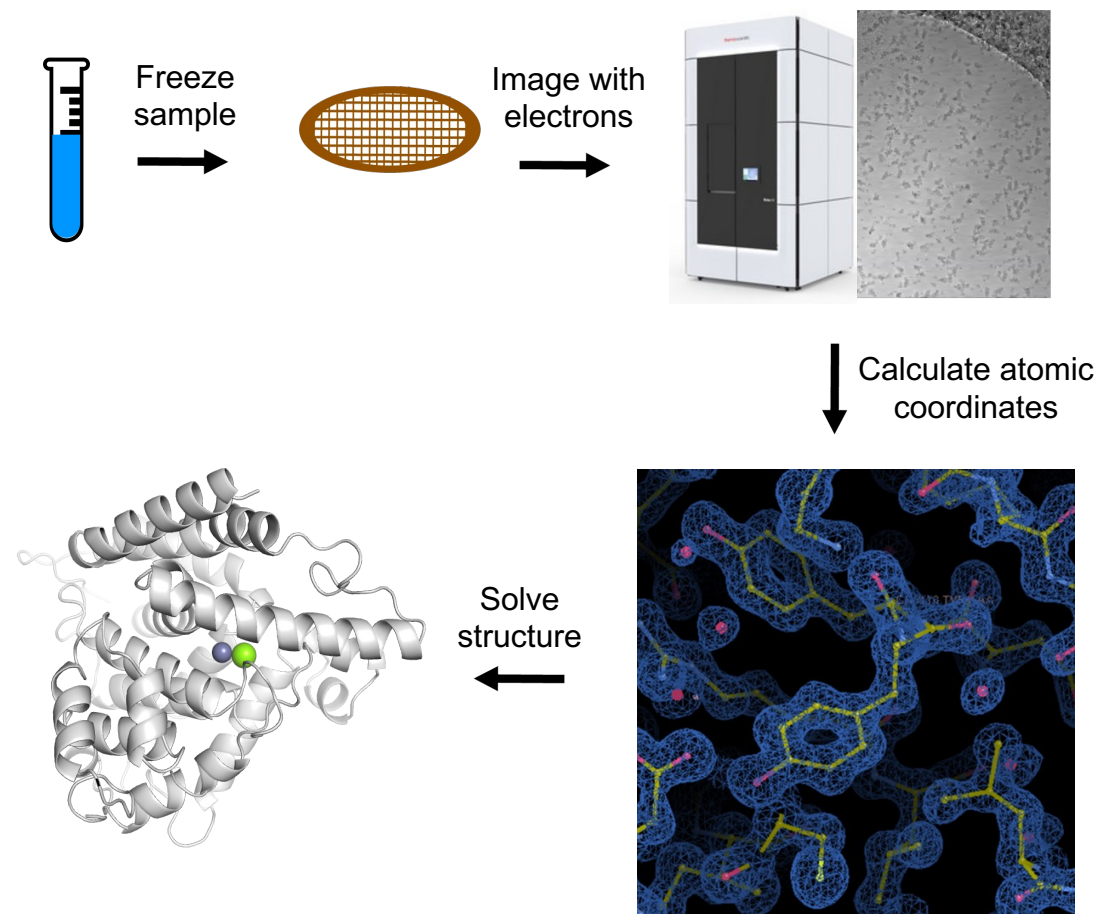
EMC purification strategy



Cryo-EM for 3D structure determination of macromolecules

State-of-the-art in structural biology

- // Cryo-EM is an imaging-based technique that reveals the atomic structures of macromolecules (protein, RNA, DNA)
- // Technical and algorithmic advances has transitioned cryo-EM from niche to method of choice for structural biologists
- // Cryo-EM opens many exciting possibilities because it can be used to study previously intractable questions
 - // For challenging specimens...
 - // Small amounts of biological material
 - // Membrane proteins
 - // Biological assemblies (protein and nucleic acid complex)
 - // Native / full length proteins
 - // Deconvolving conformational dynamics





AI/ML methods are essential for all stages in cryo-EM

SBGrid Consortium is a great online resource for talks on applications

// Automated data collection

// Scipion

// Image processing

// CryoAssess

// Particle Picking

// Topaz & crYOLO

// 3D-reconstruction

// CryoDRGN

// Model Building

// ModelAngelo

// Map improvement

// DeepEMhancer



SBGrid Consortium

@SBGridTV · 5.9K subscribers · 200 videos

For structural biologists, by structural biologists! >

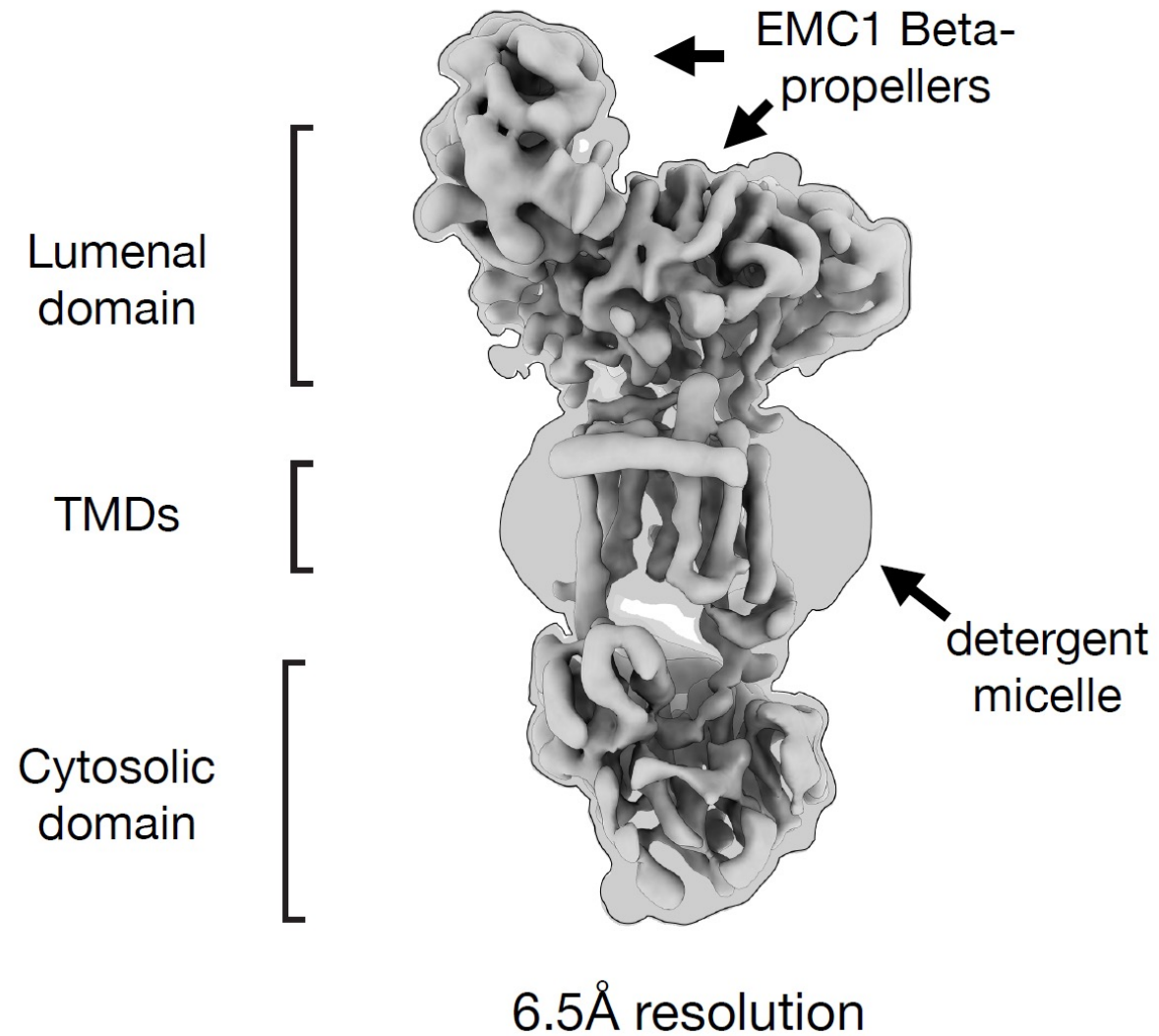
sbgrid.org and 4 more links

Subscribe

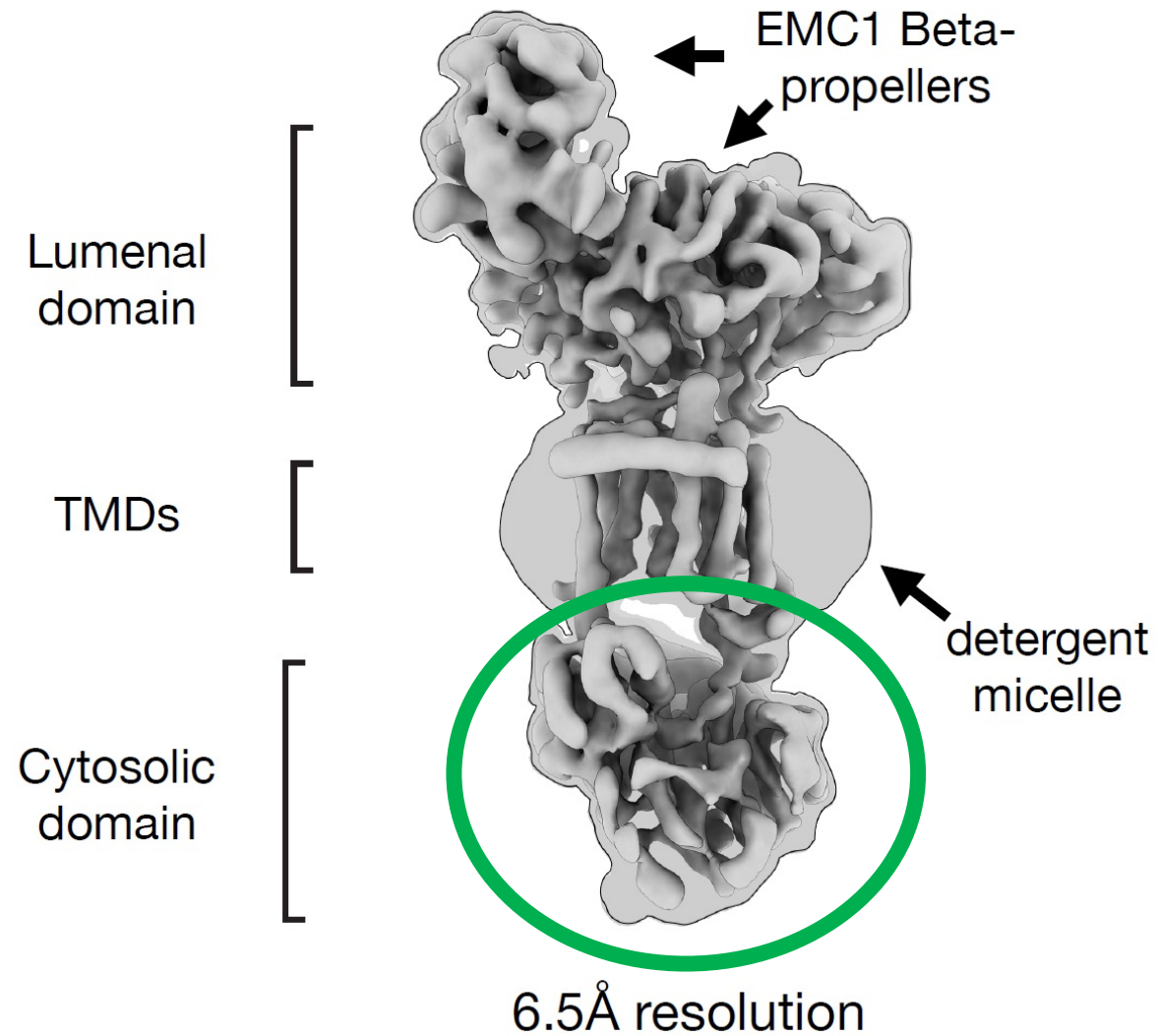
<https://sbgrid.org/>

<https://www.youtube.com/@SBGridTV>

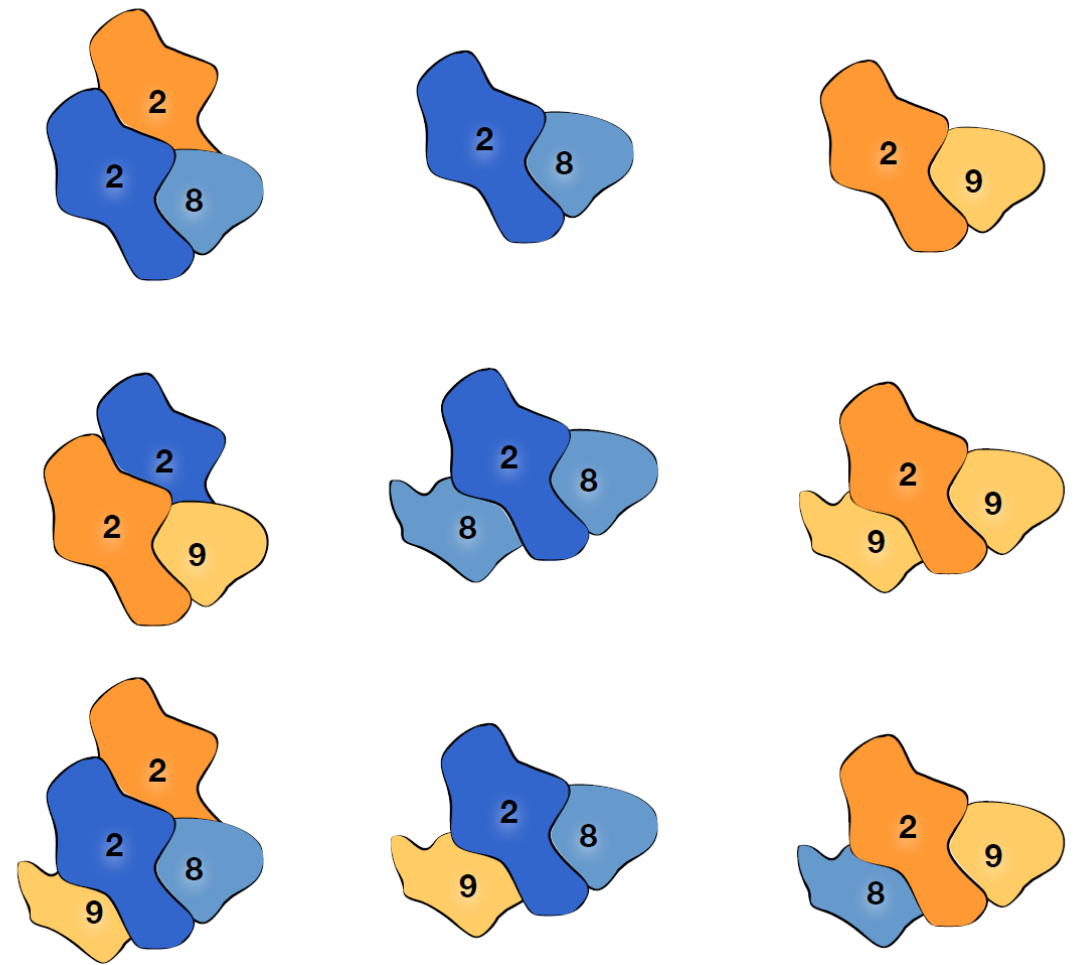
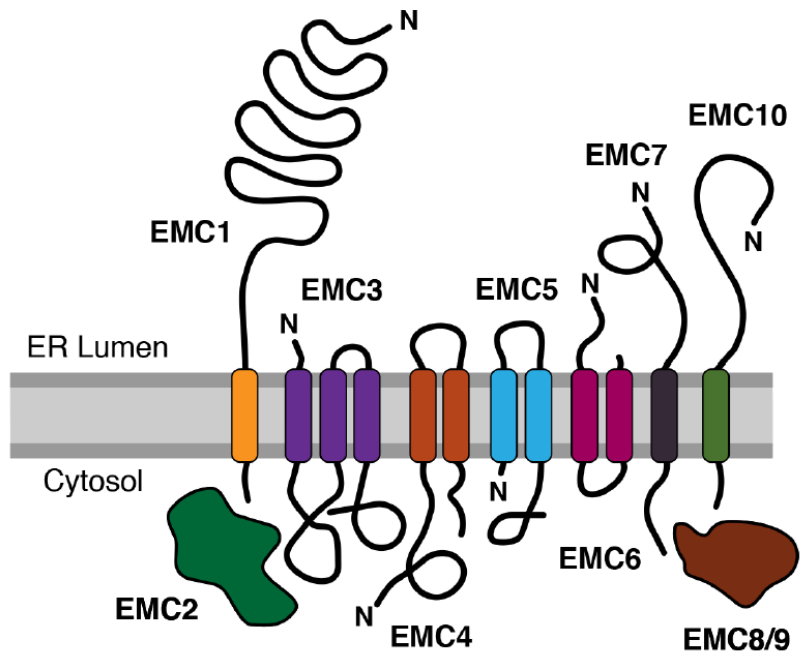
Low resolution cryoEM structure of EMC



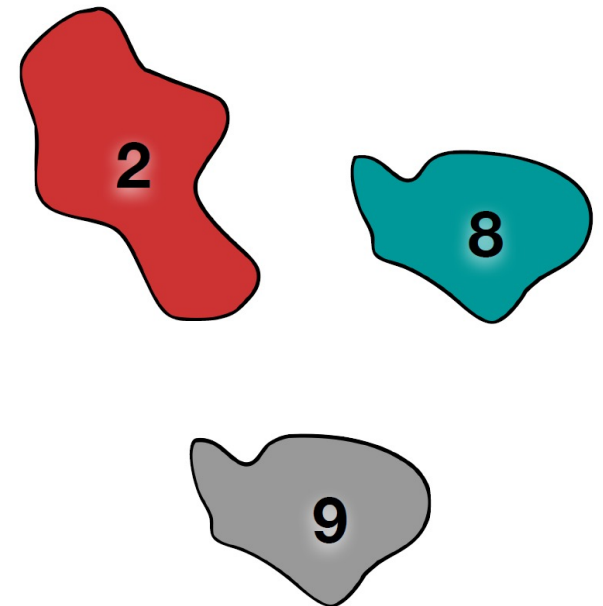
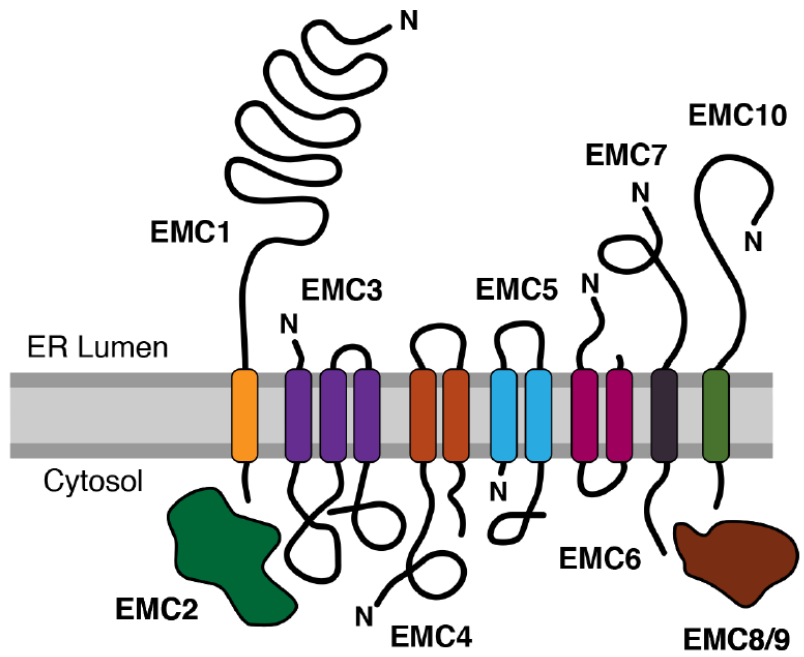
Low resolution cryoEM structure of EMC



What is the composition of the cytoplasmic EMC



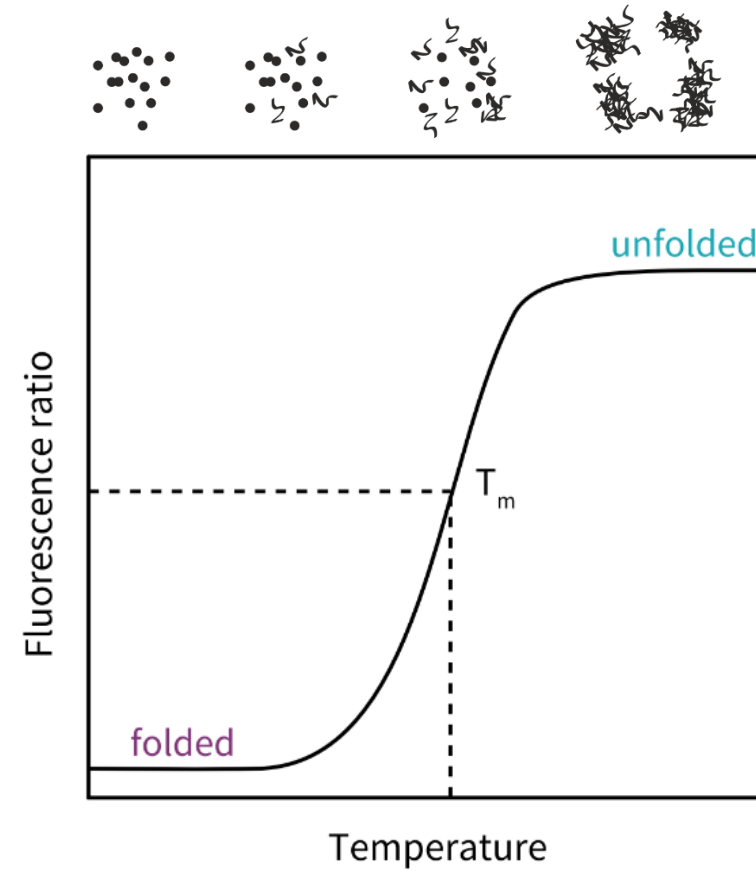
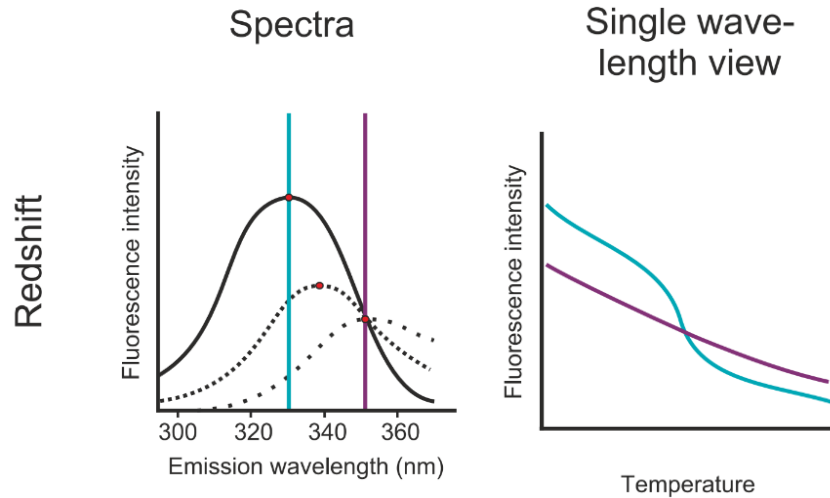
Purification and analysis of individual subunits



Nano Differential Scanning Fluorimetry (nanoDSF)

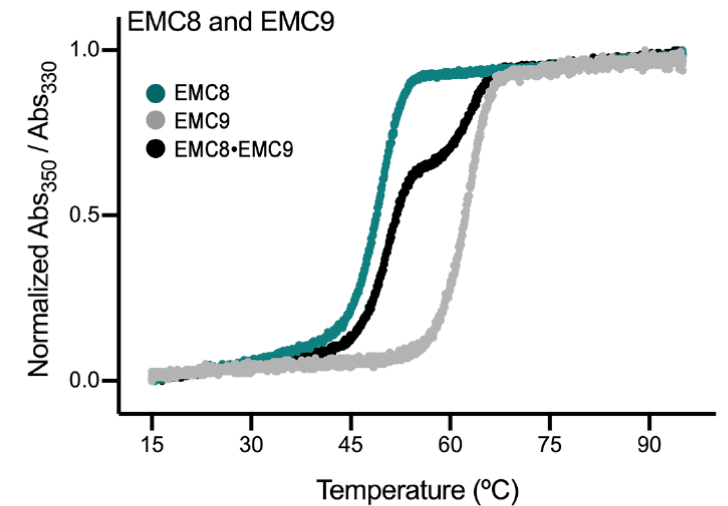
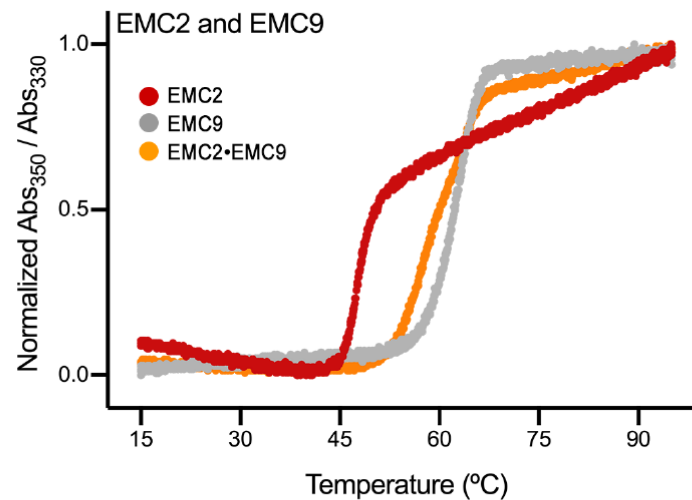
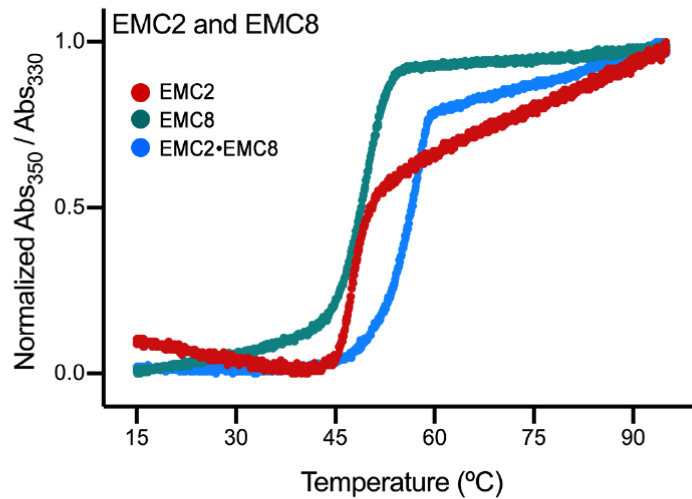
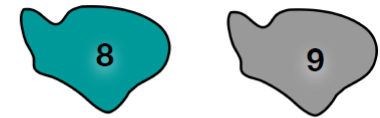
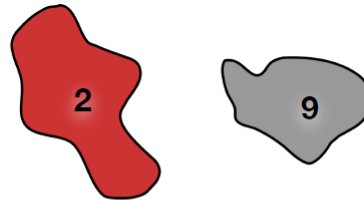
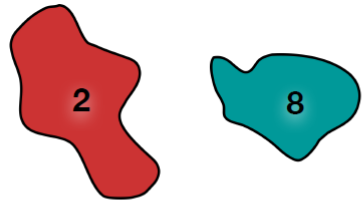
Measure protein unfolding as a function of temperature

Tryptophan fluorescence changes based on environment

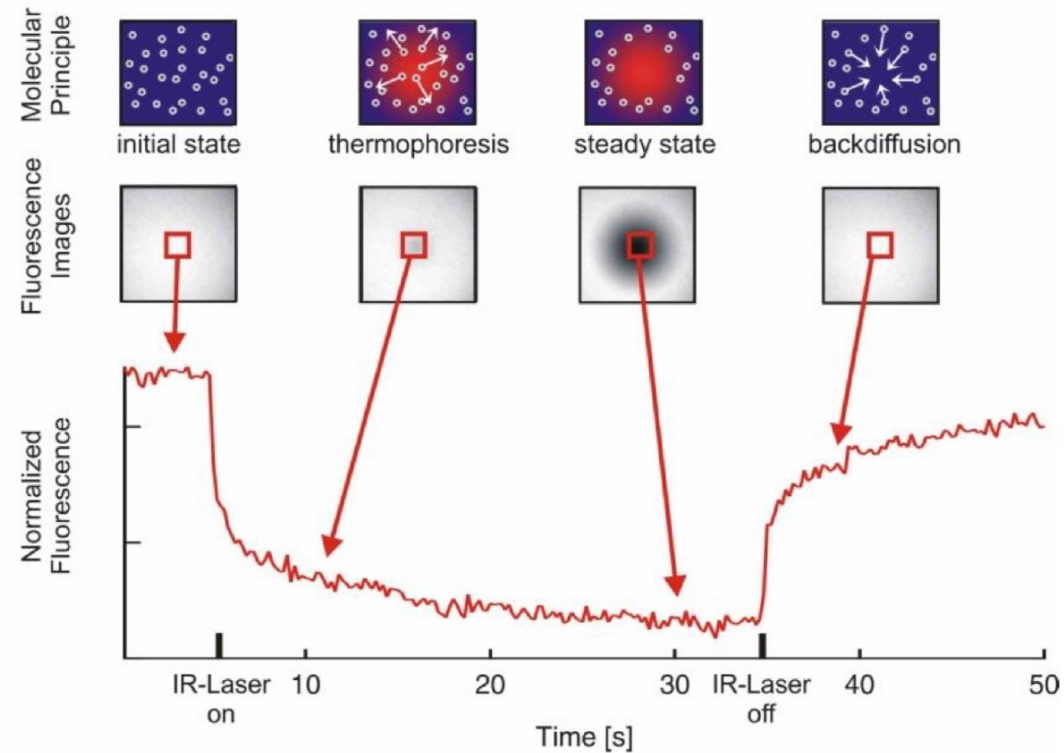
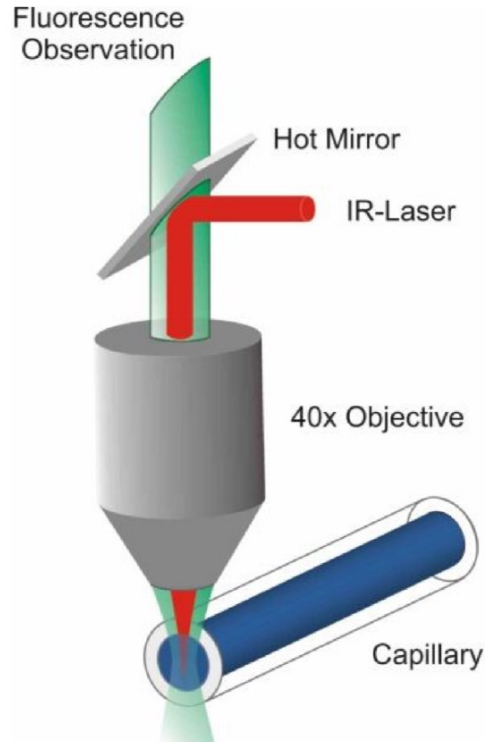




EMC2 increases stability when combined with EMC8 or EMC9

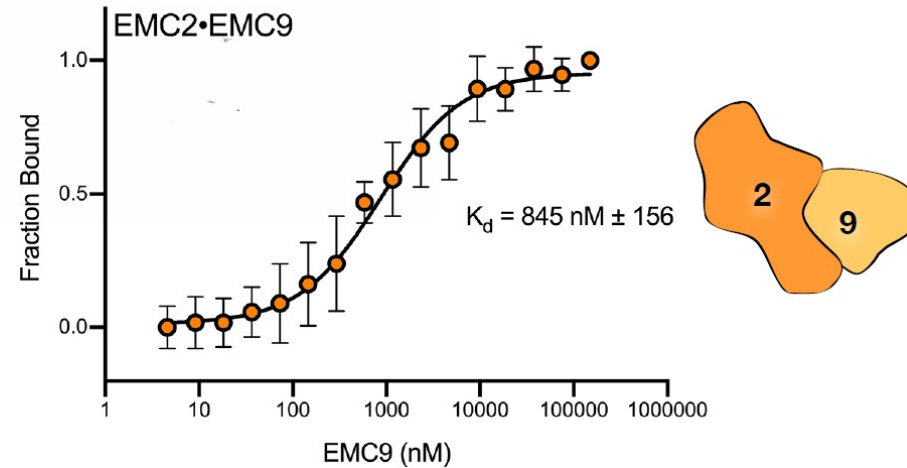
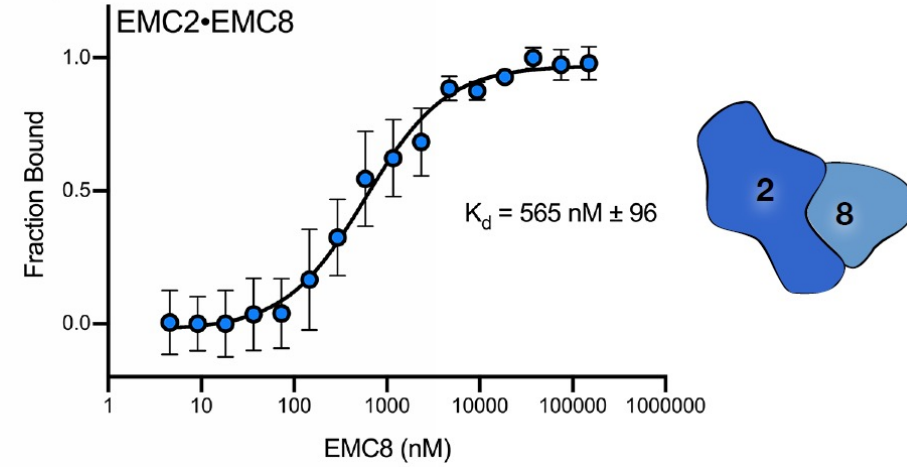
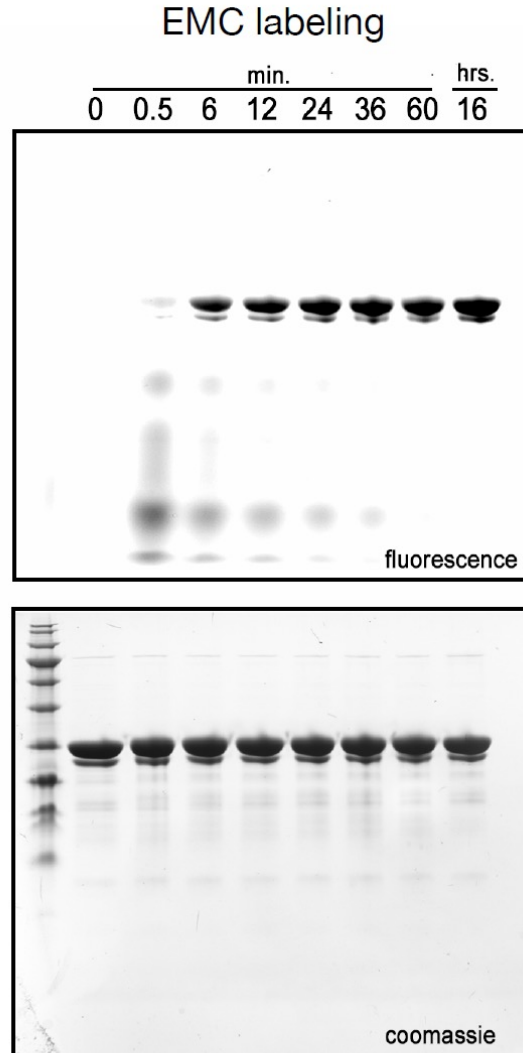
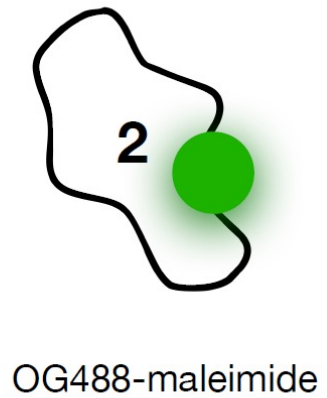


Are stable complexes forming?

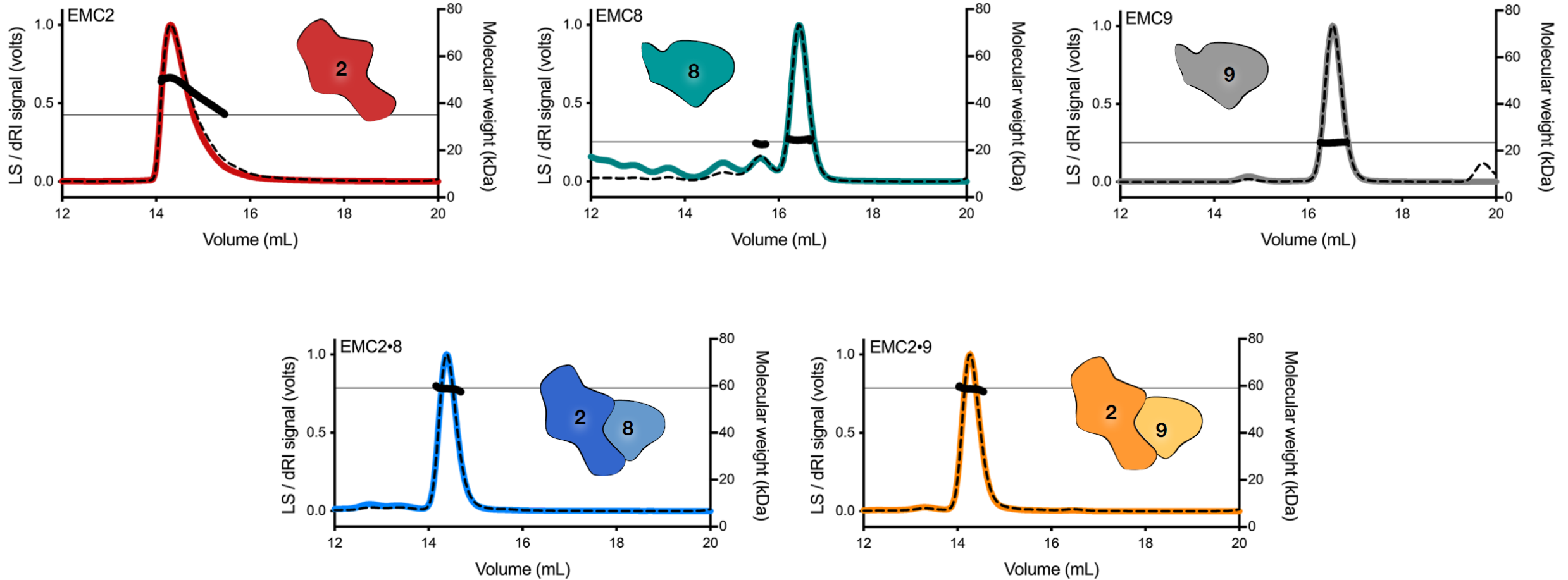


Microscale thermophoresis is based on the detection of a temperature-induced change in fluorescence of a target as a function of the concentration of a non-fluorescent ligand. The observed change in fluorescence is based on two distinct effects.

Are stable complexes forming?



What is the stoichiometry of the complexes?

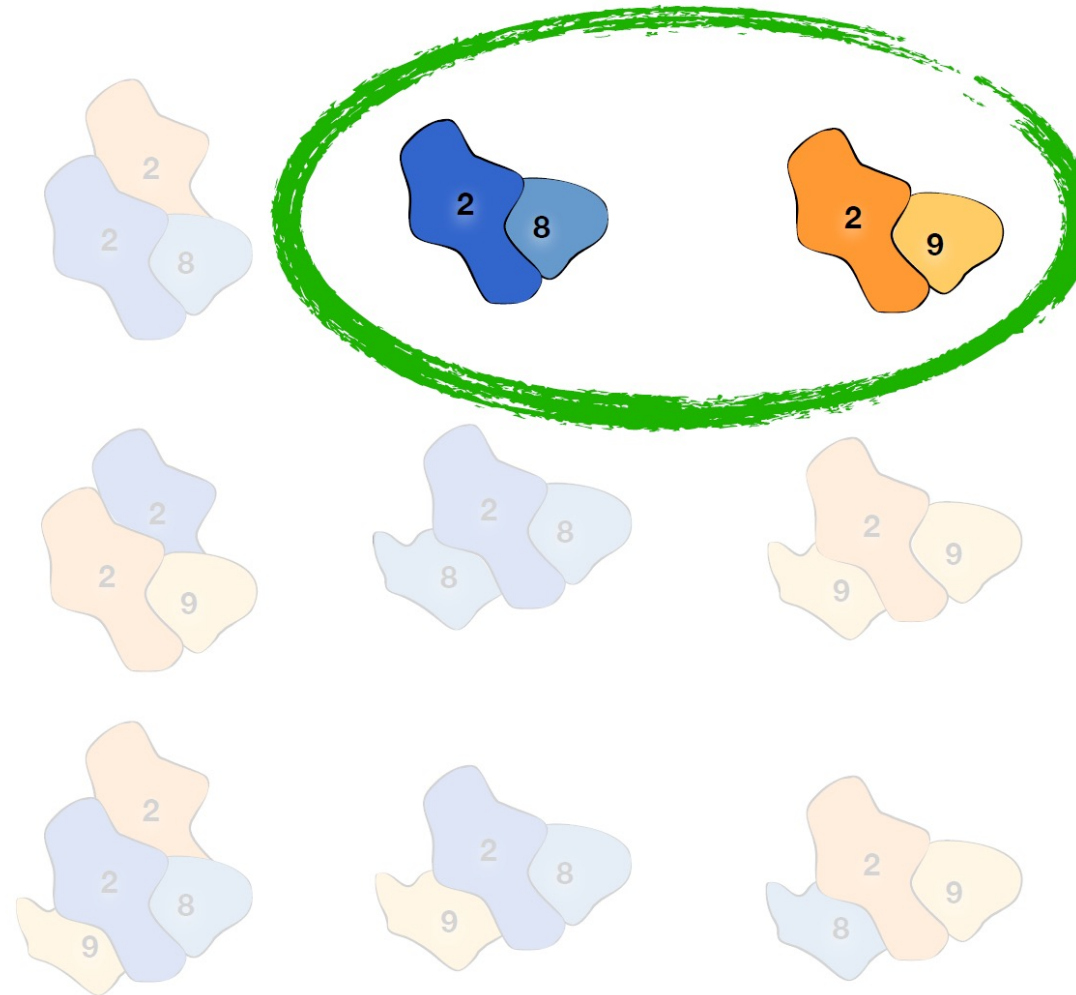


Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) gold standard in MW determination



Complexes consist of EMC2+8 and EMC2+9

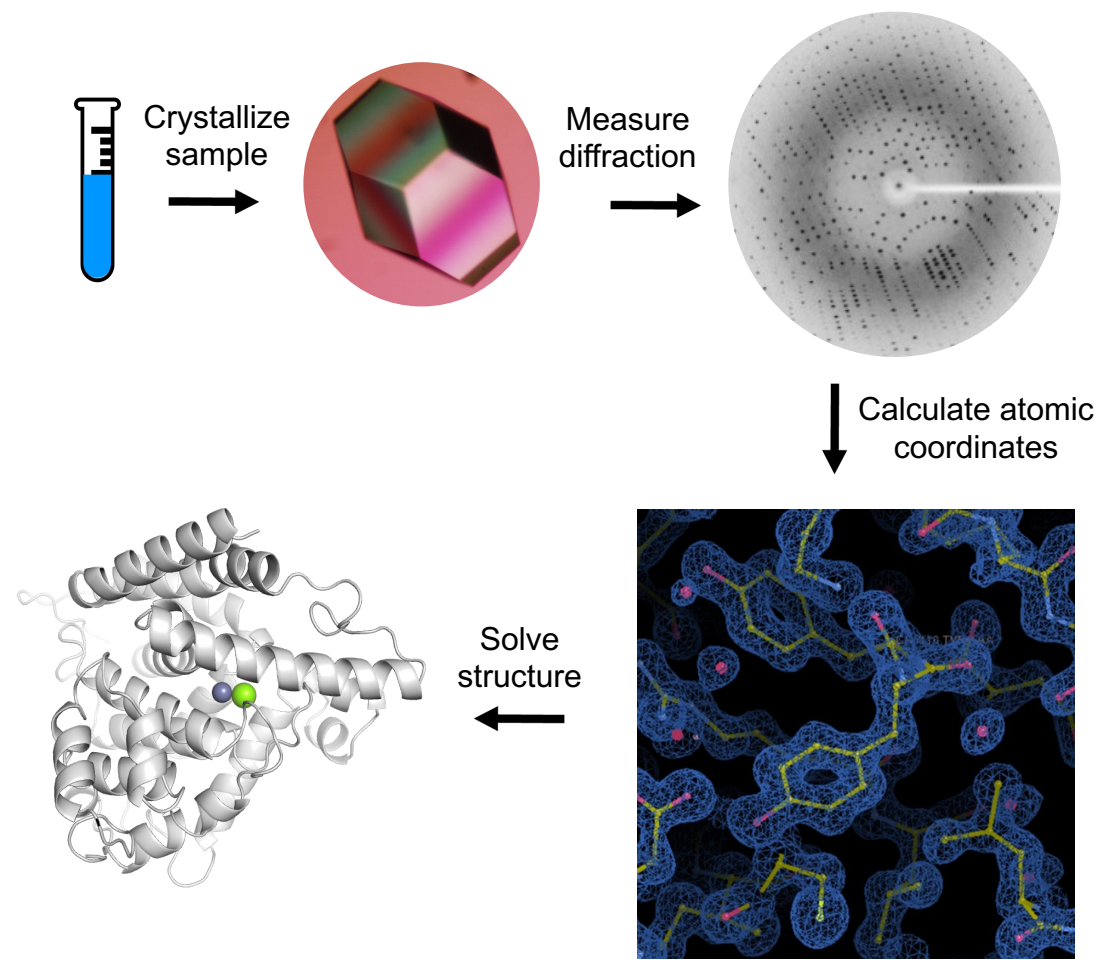
Independently validated with MST and SEC-MALS



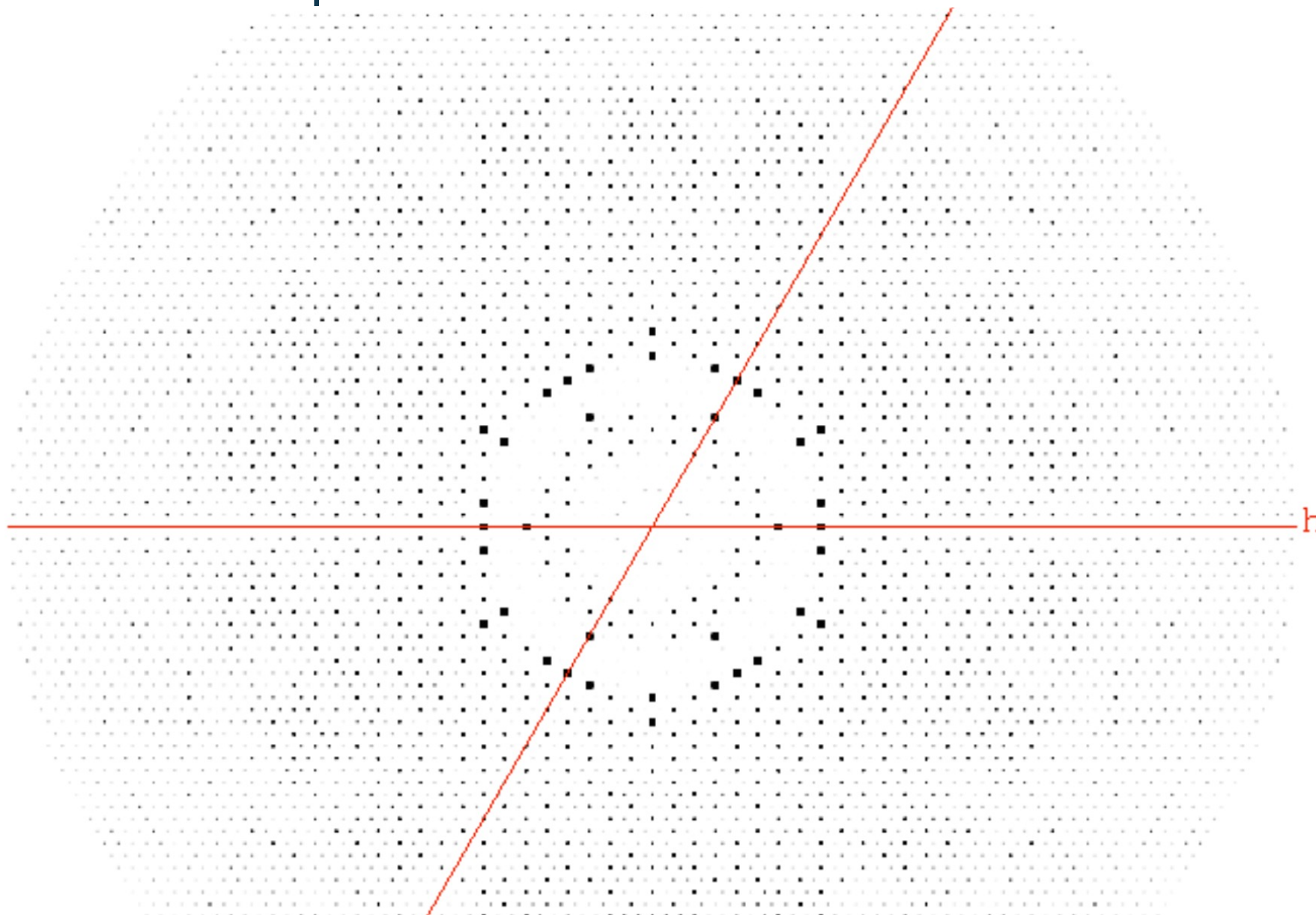
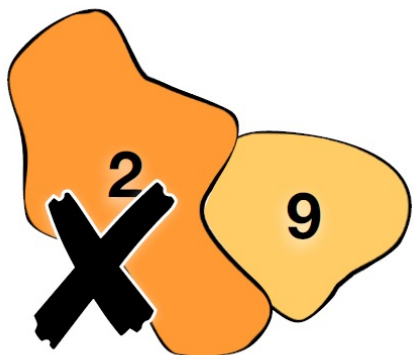
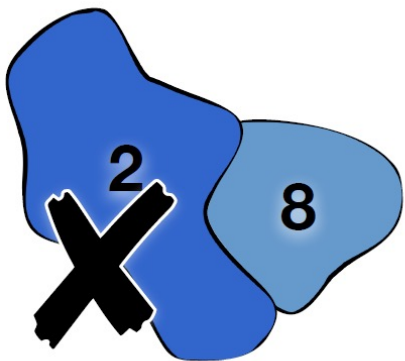
X-ray crystallography for 3D structure determination of macromolecules

Established SBIO technique

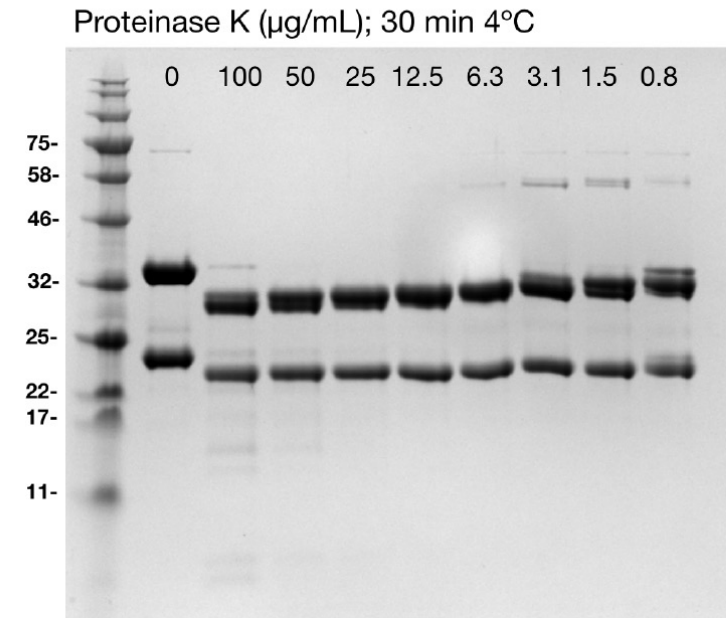
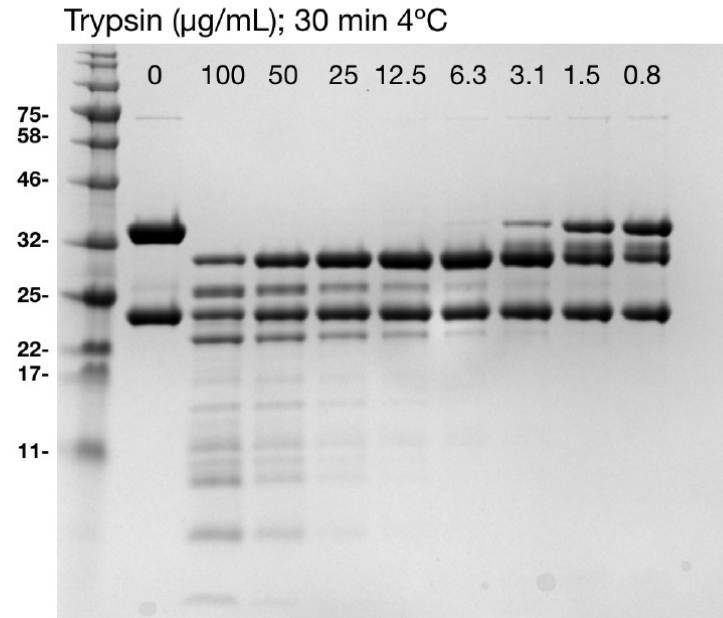
- // X-ray crystallography is a diffraction-based technique that reveals the atomic structures of macromolecules (protein, RNA, DNA)
- // X-ray crystallography has long been the method of choice when determining macromolecular structure.
- // Pros:
 - // High resolution;
 - // High-throughput compared to cryo-EM in robust crystallization systems
 - // Can be coupled to lead discovery with fragment screening
- // Cons:
 - // **Relies on crystallization of protein** (or RNA/DNA)
 - // Requires large amount of sample
 - // Crystallization disfavors proteins with flexibility/dynamics



Crystallography with the complexes



Limited proteolysis to determine crystallization boundaries



>EMC2

MAKVSELYDV TWEEMRDKMRKWRENSRNSEQIVEVGEELINEYASKLGDDEIWI IYEQVMIAALDYGRDDLALFCLQELR
 RQFPGSHRVKRLTGMRFEAMERYDDAIQLYDRILQEDPTNTAARKRKIAIRKAQGKNVEAIRELNEYLEQFVGDQEAWHE
 LAELYINEHDYAKAAFCLLELMMTNPHNHLQCQYAEVKYTQGGLENLELSRKYFAQALKLNNRNMRALFGLYMSASHIA
 SNPKASAKTKKDNMKYASWAASQINRAYQFAGRS **KKETKYSLKAVEDMLETLOITQS**

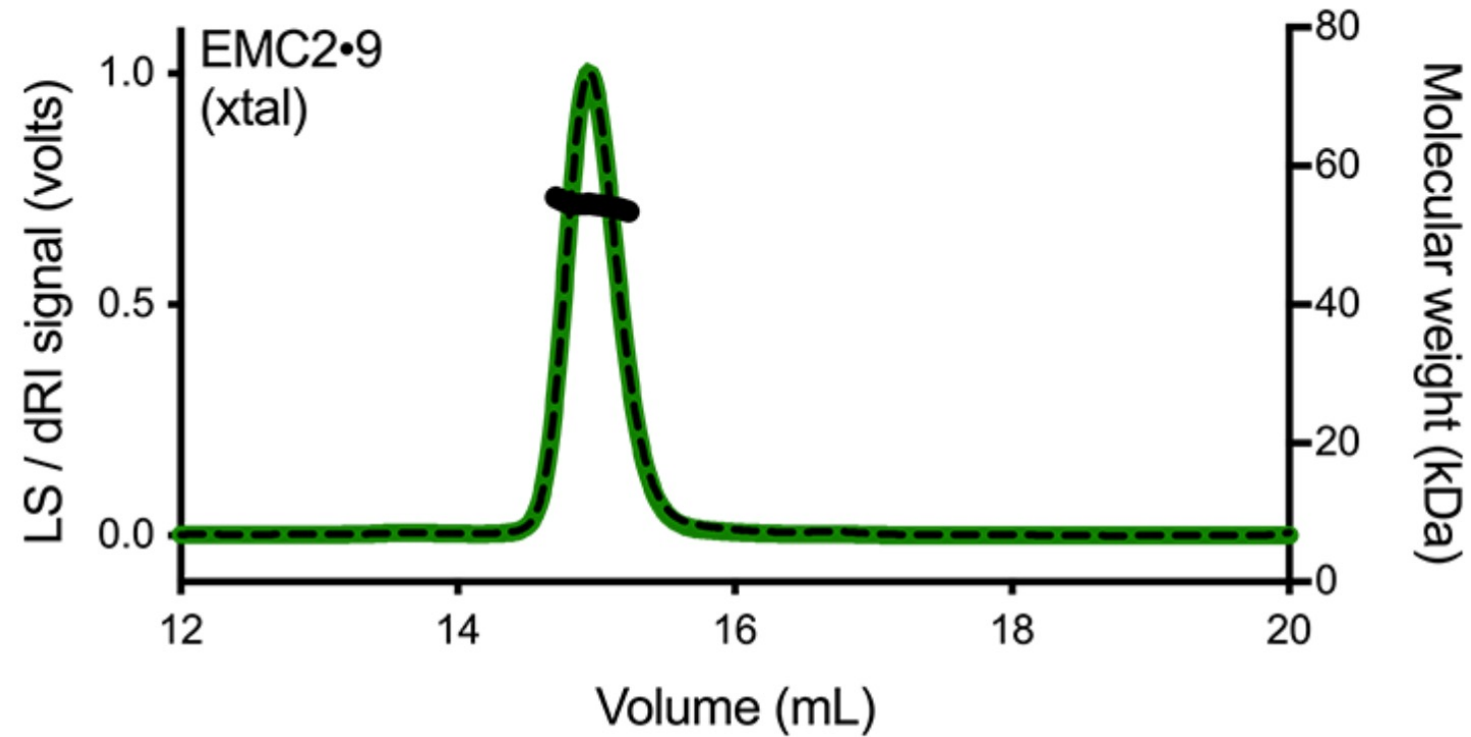
>EMC9

MGEVEI SALAYVKMCLHAARYPHAAVNGLFLAPAPRSGECLCLTDCVPLFHSHLALSVMLEVALNQVDVWGAQAGLVVAG
 YYHANAAVNDQSPGPLALKIAGRIAEFFPDAVLIMLDNQKLVPPQPRVPPVIVLENQGLRWVVPKDKNLVMWRDWEESRQMV
 GALLEDRAHQHLVDFDCHLDDIRQDWTNQRLNTQITQWVG **PTNGNGNA**



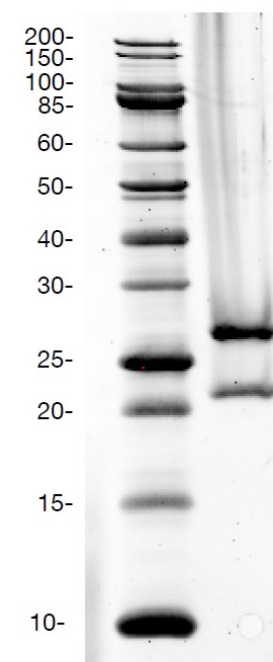
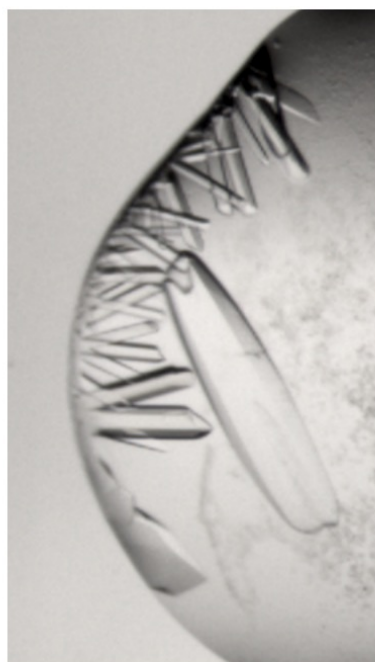
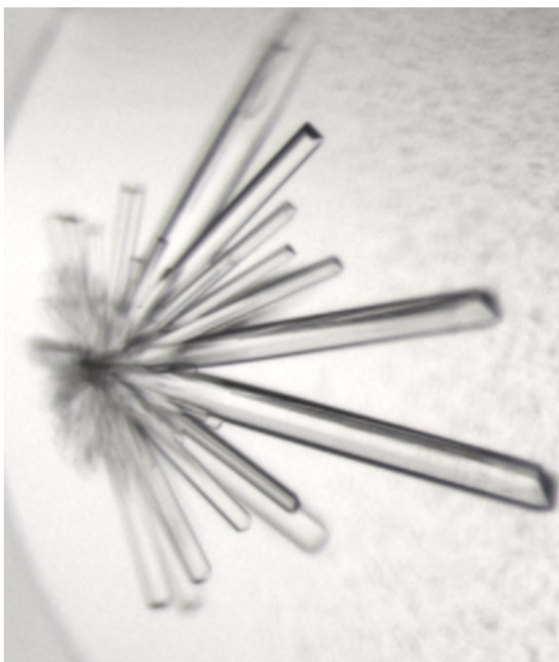
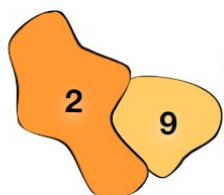
Limited proteolysis to determine crystallization boundaries

SEC-MALS data shows excellent QC



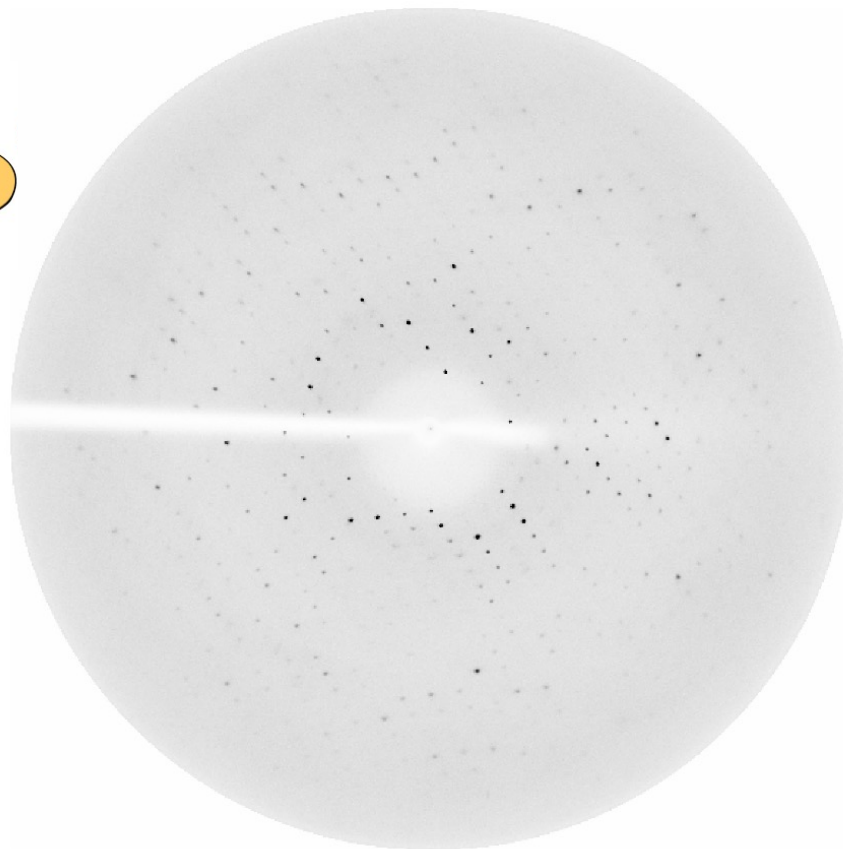
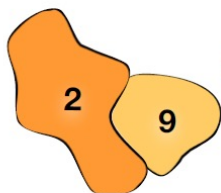


Crystallography with the complexes



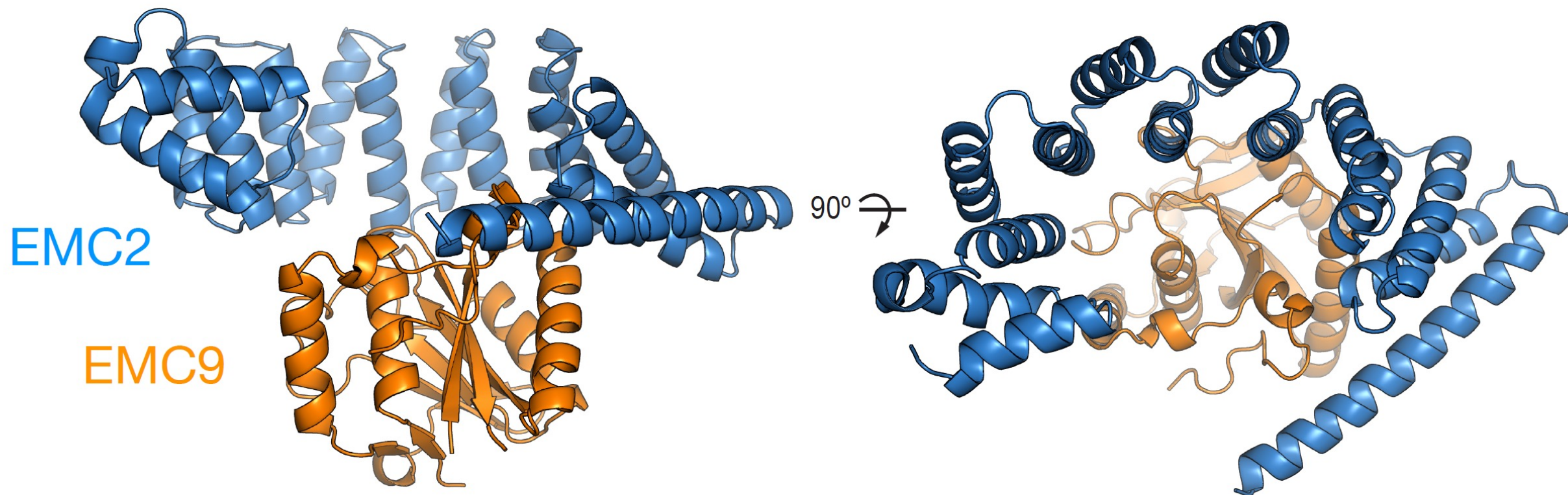


Native data set to 2.8Å



Resolution at 2.2 Å but
with interesting pathology

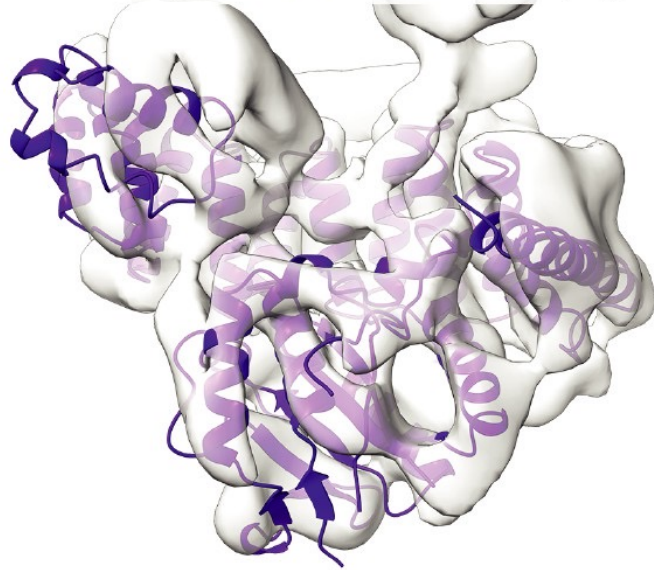
Crystal structure of EMC2+9



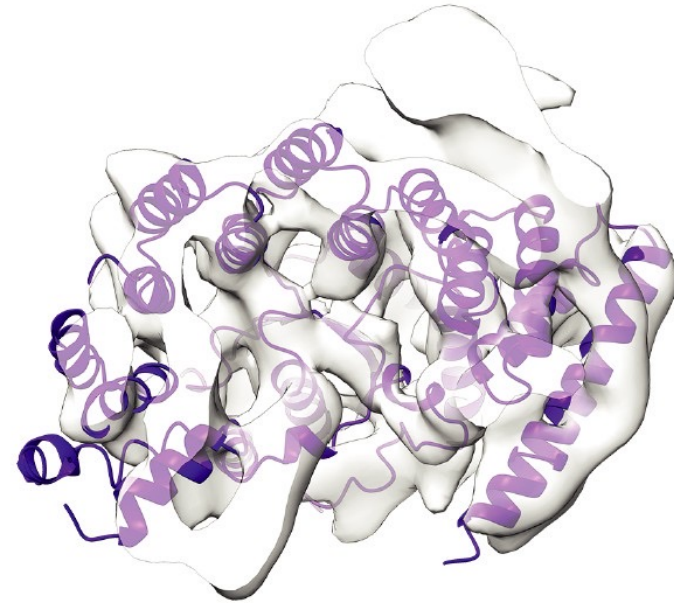
SG: P2₁2₁2₁ at 2.2 Å resolution

Crystal structure and EM density is not a perfect match

○ EM Maps ● EMC2•9 (rigid body)

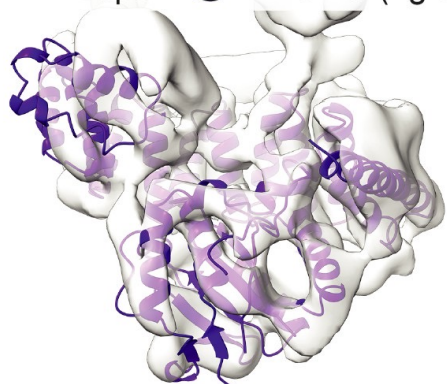


90° ↻

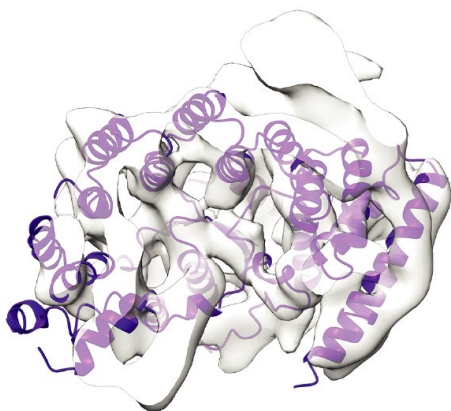


Normal mode analysis as a proxy for dynamics

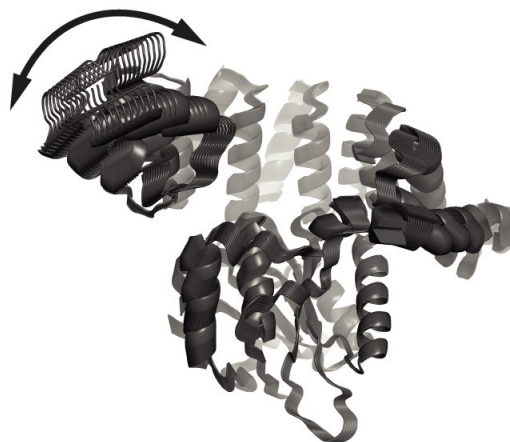
○ EM Maps ● EMC2•9 (rigid body)



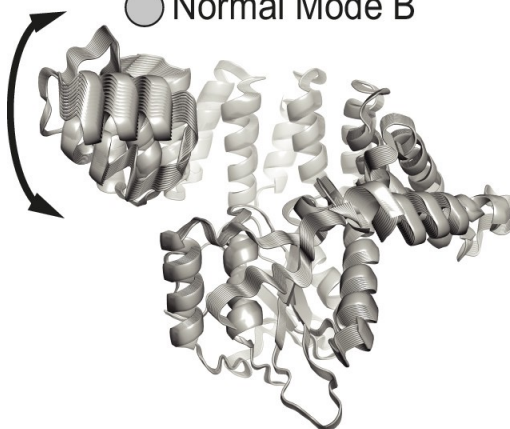
90° ↻



● Normal Mode A



○ Normal Mode B

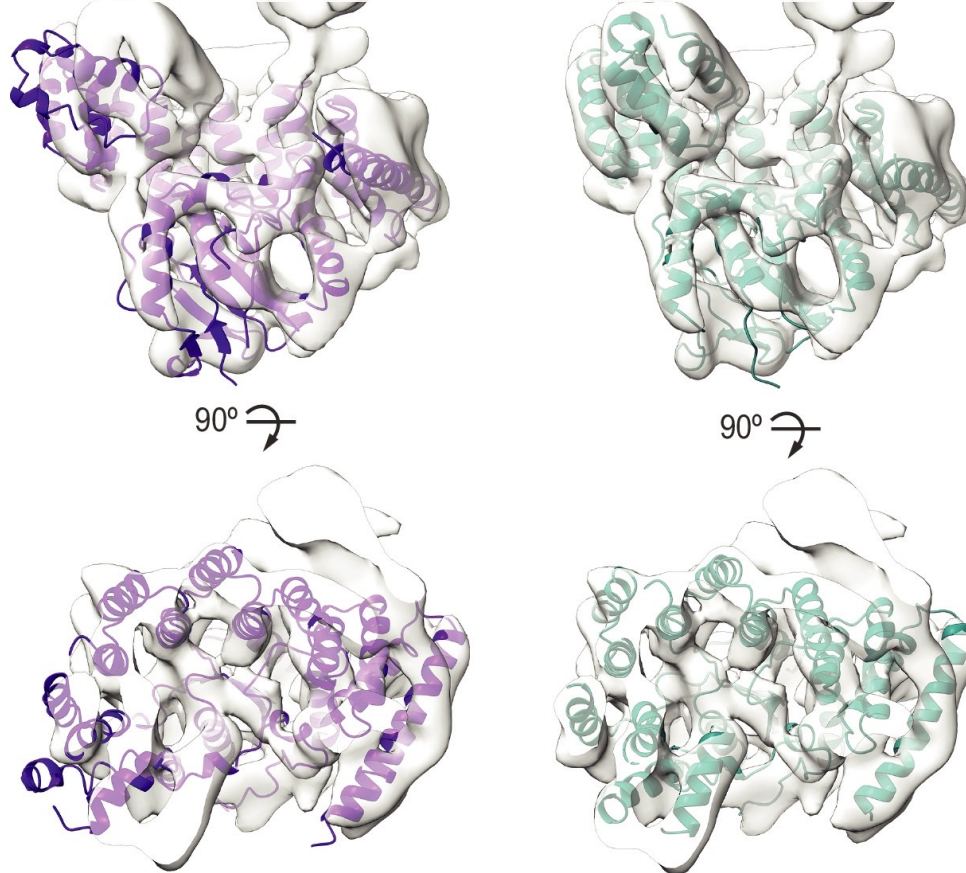


Normal mode analysis is a technique that can be used to describe the flexible states accessible to a protein or other molecule about an equilibrium position. It is based on the physics used to describe small oscillations (PMID: 31510014)

Used software EInémo to calculate normal modes (PMID: 15215461)

Integrating modelling and experimental data

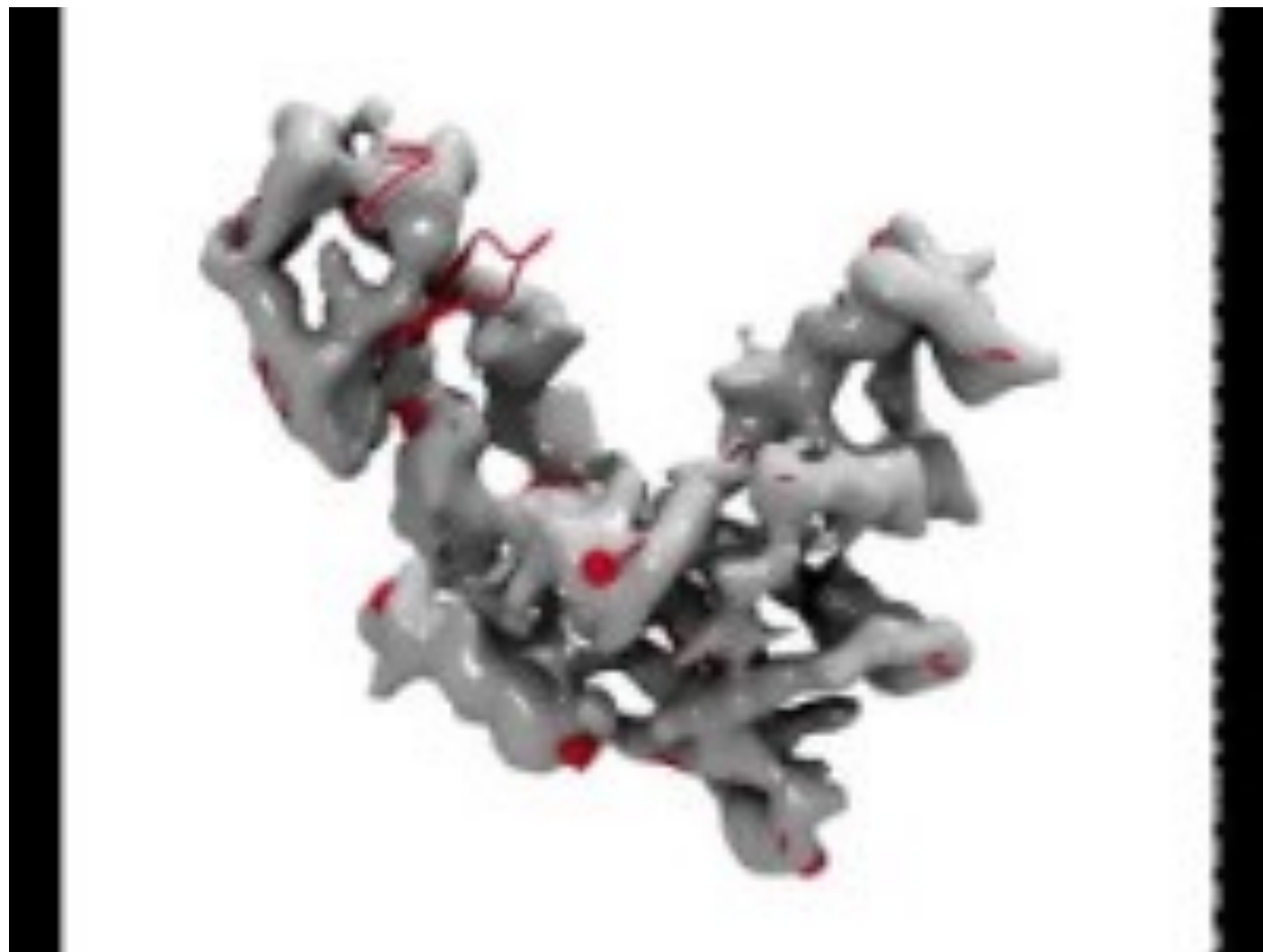
○ EM Maps ● EMC2•9 (rigid body) ● EMC2•9 Morph



- Flex-EM: uses Monte-Carlo search, conjugate-gradients minimization, and simulated annealing molecular dynamics (PMID: 18275820)
- Namdinator: automated molecular dynamics flexible fitting simulation and real space refinement (PMID: 31316797)
- ISOLDE: interactive real-time molecular dynamics flexible fitting (PMID: 29872003)

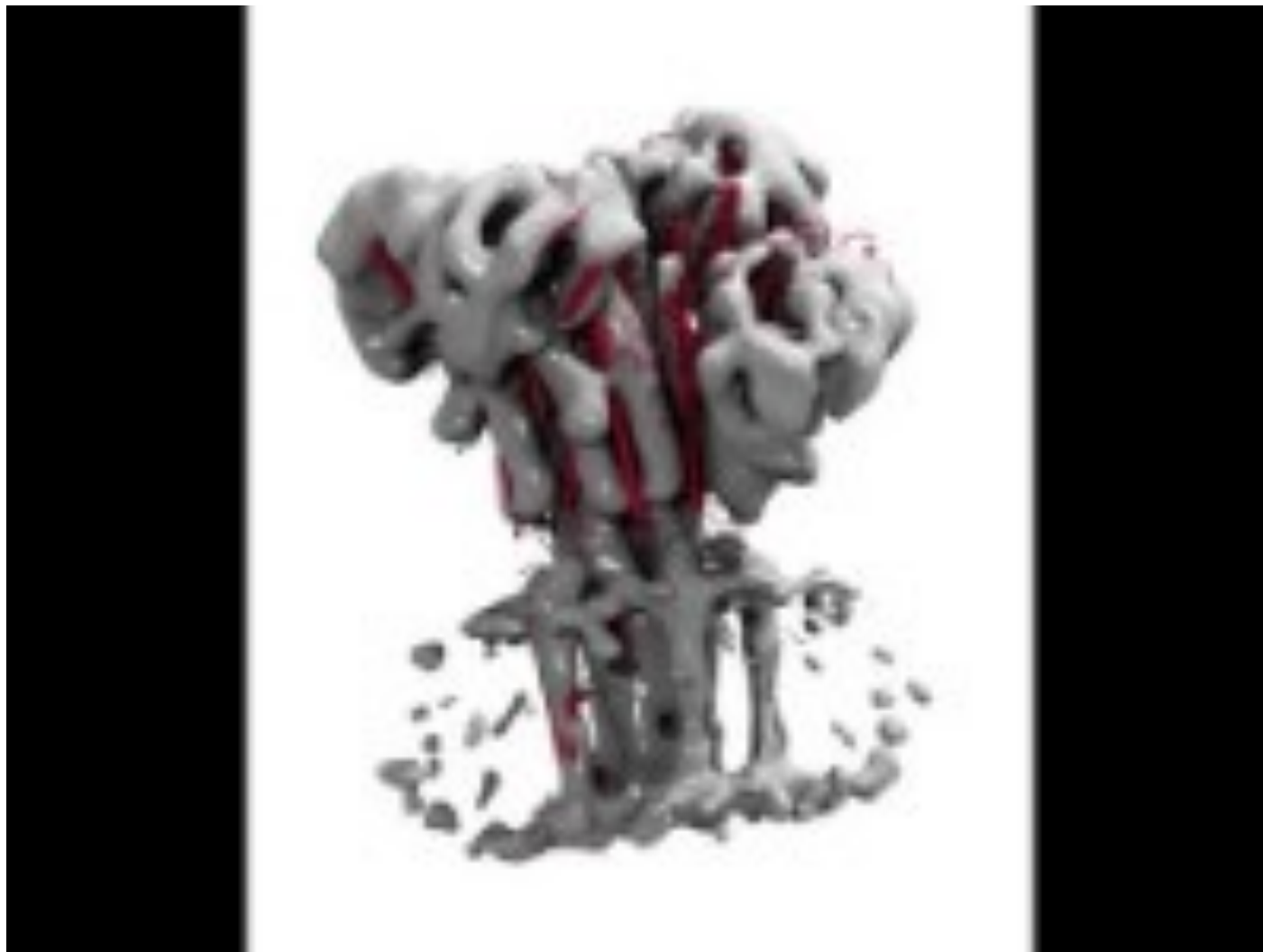


Namdinator automated MD flexible fitting



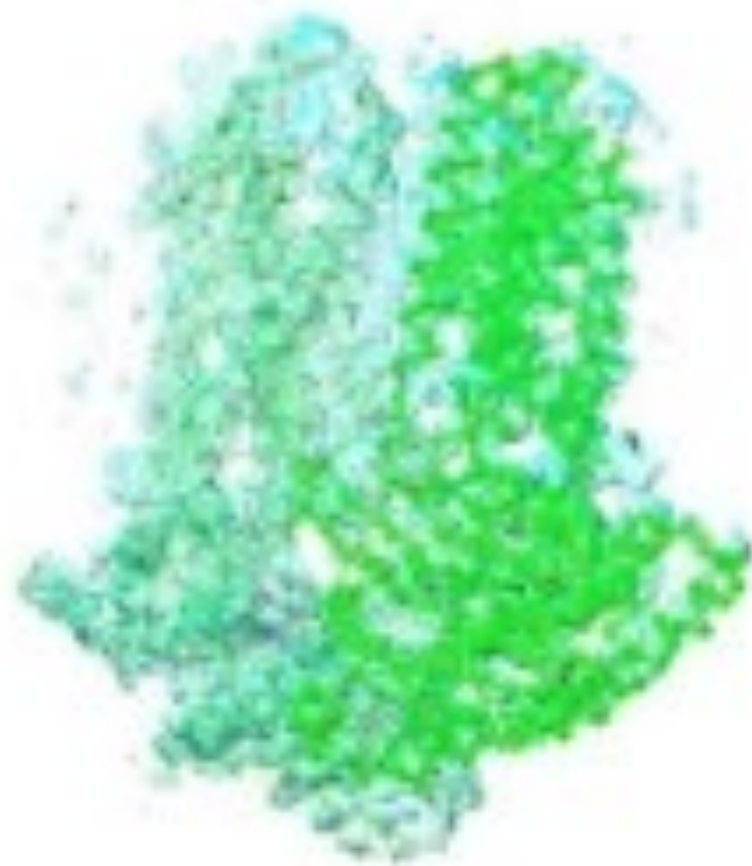


Namdinator automated MD flexible fitting

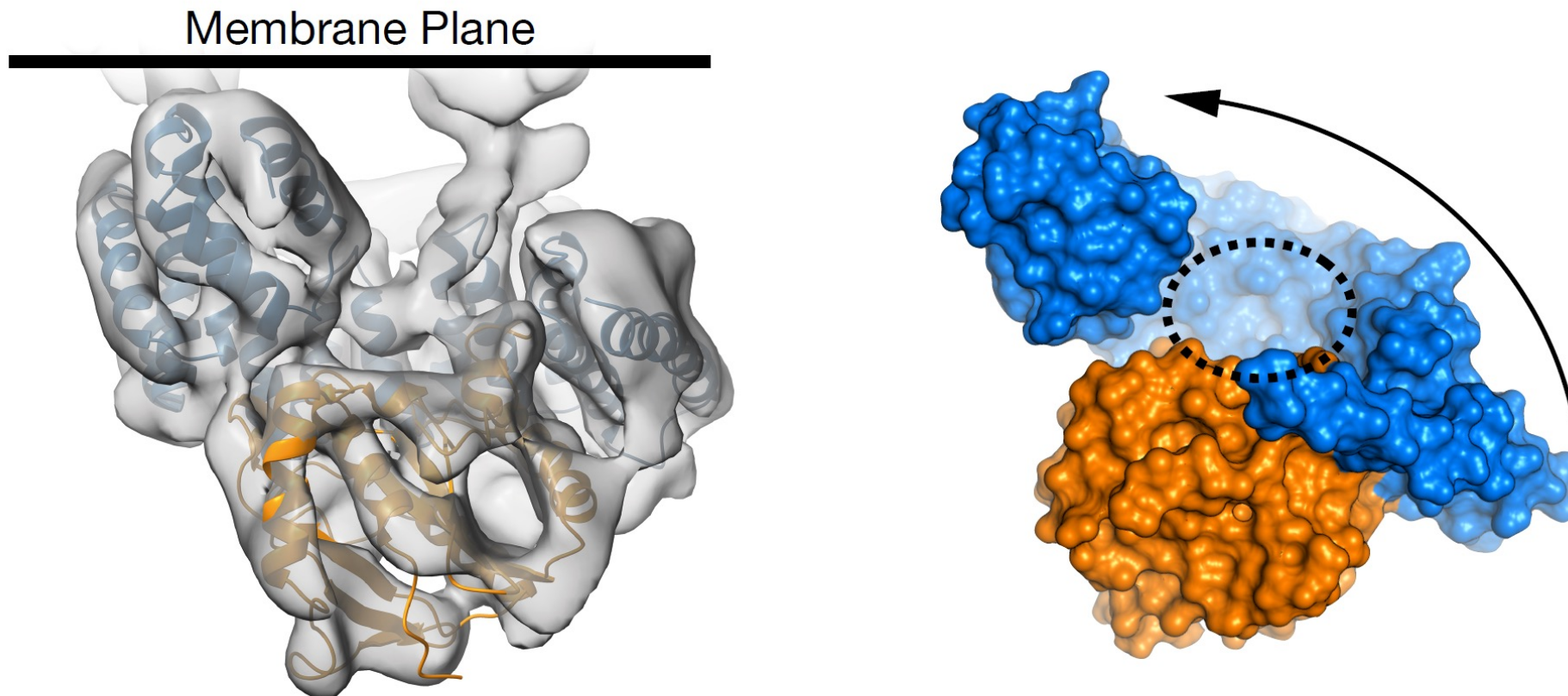




ISOLDE MD flexible fitting during model building



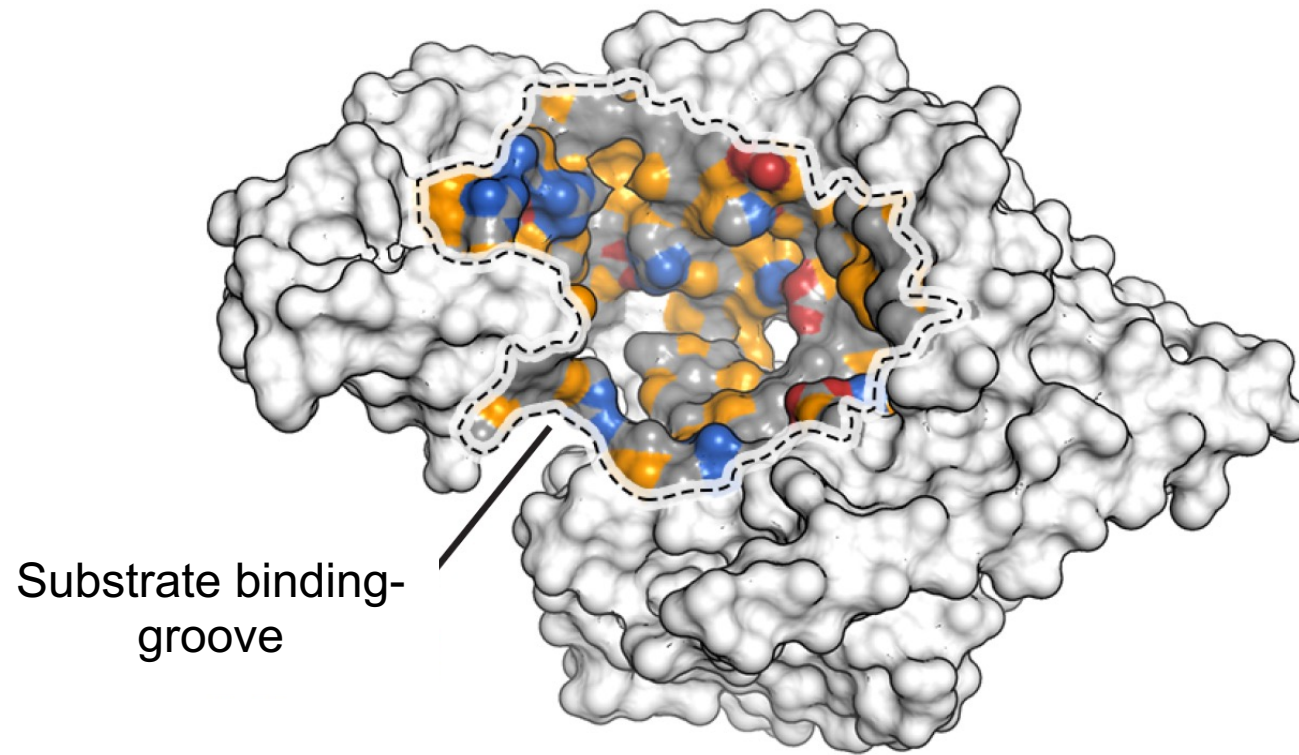
Substrate binding groove faces the membrane plan



○ EM ● EMC2 ● EMC9



EMC binding groove is only moderately hydrophobic

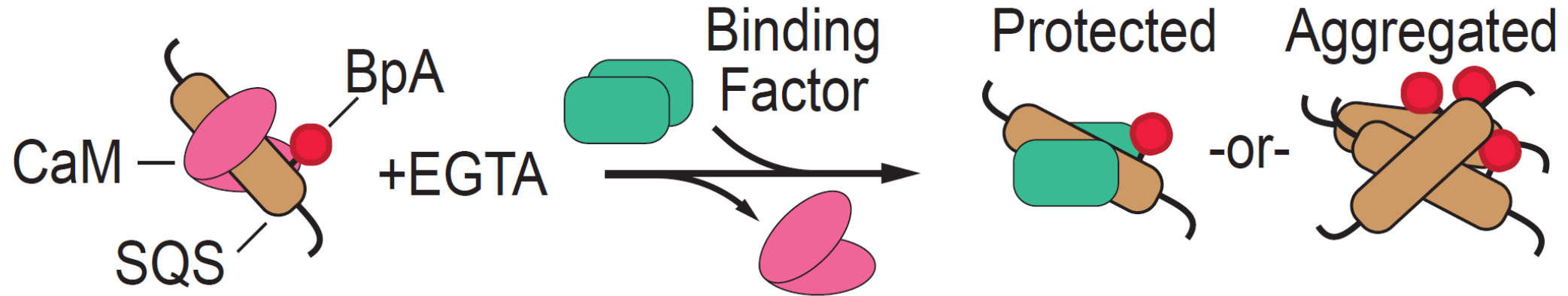


Substrate binding-groove

- C, CH, CH₂, CH₃ not attached to N or O
- Polar / Peptide backbone

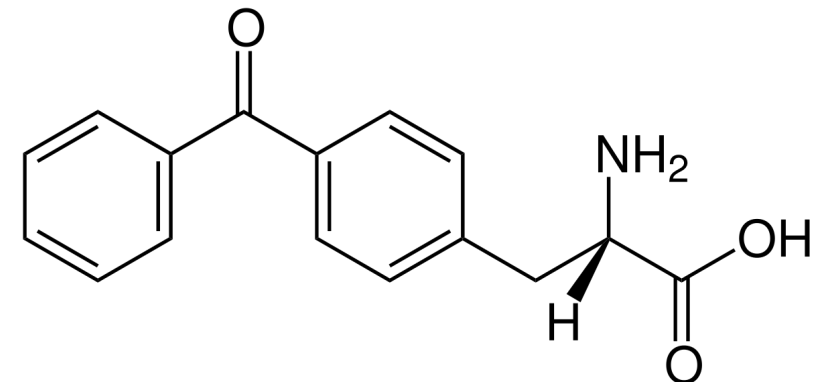
- Positively charged atoms
- Negatively charged atoms

Can cytosolic EMC engage TMDs?

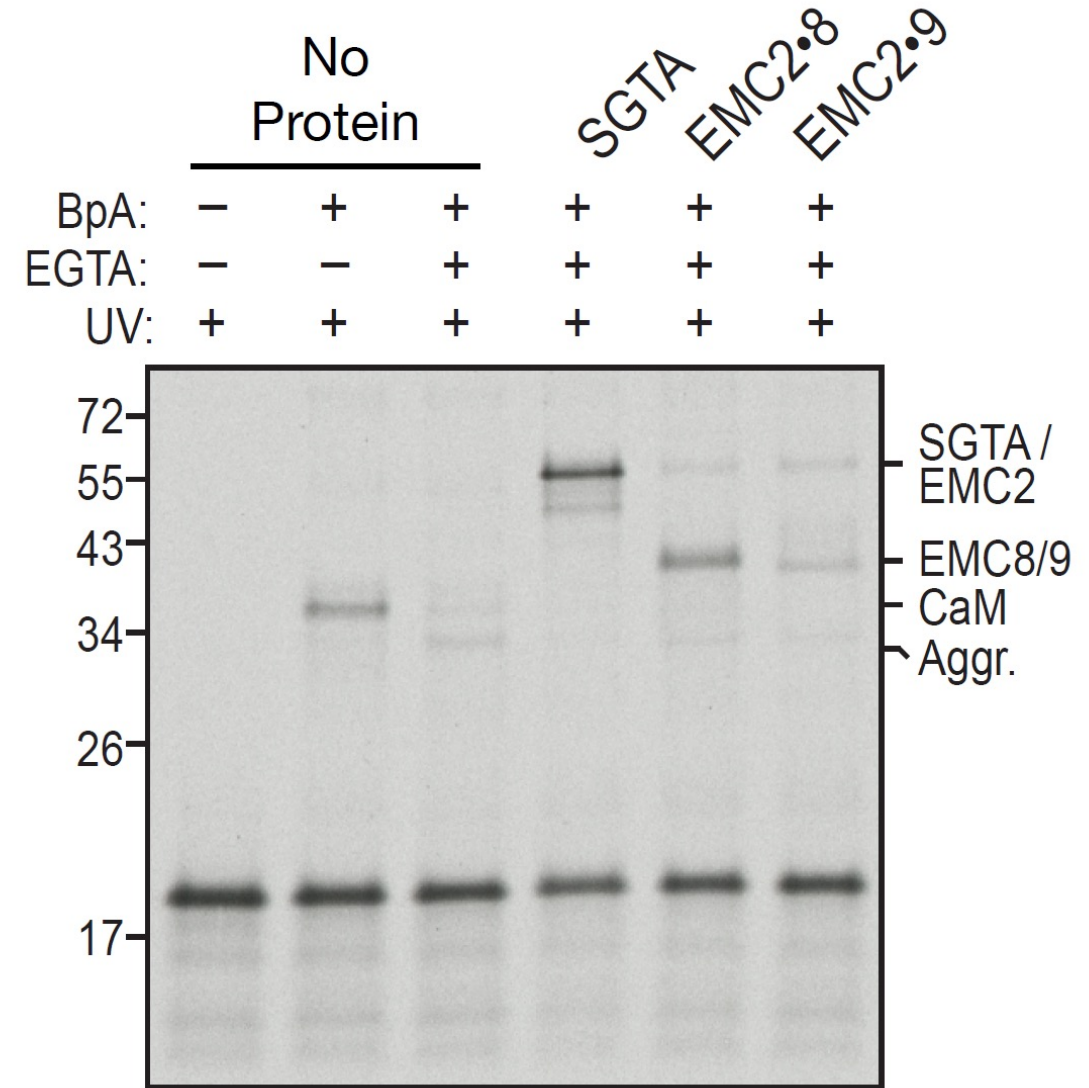
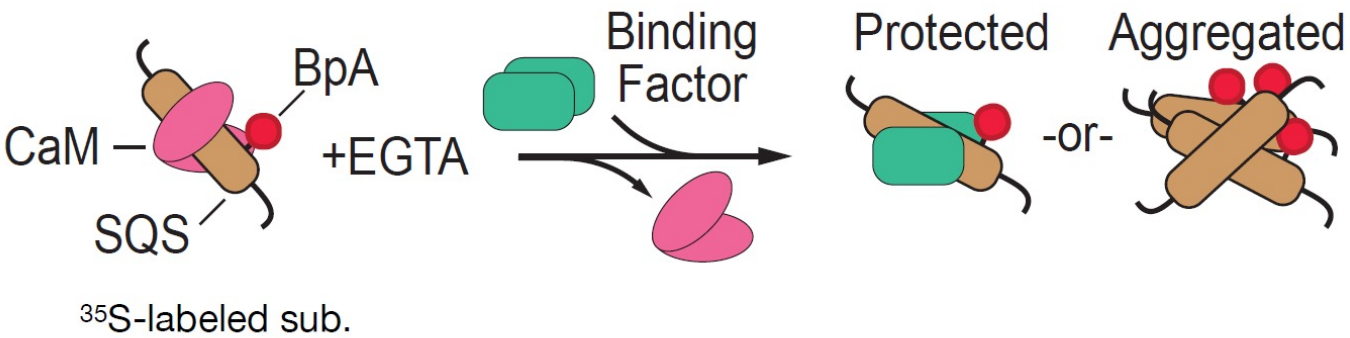


In vitro *E. coli* translation system with modifications:

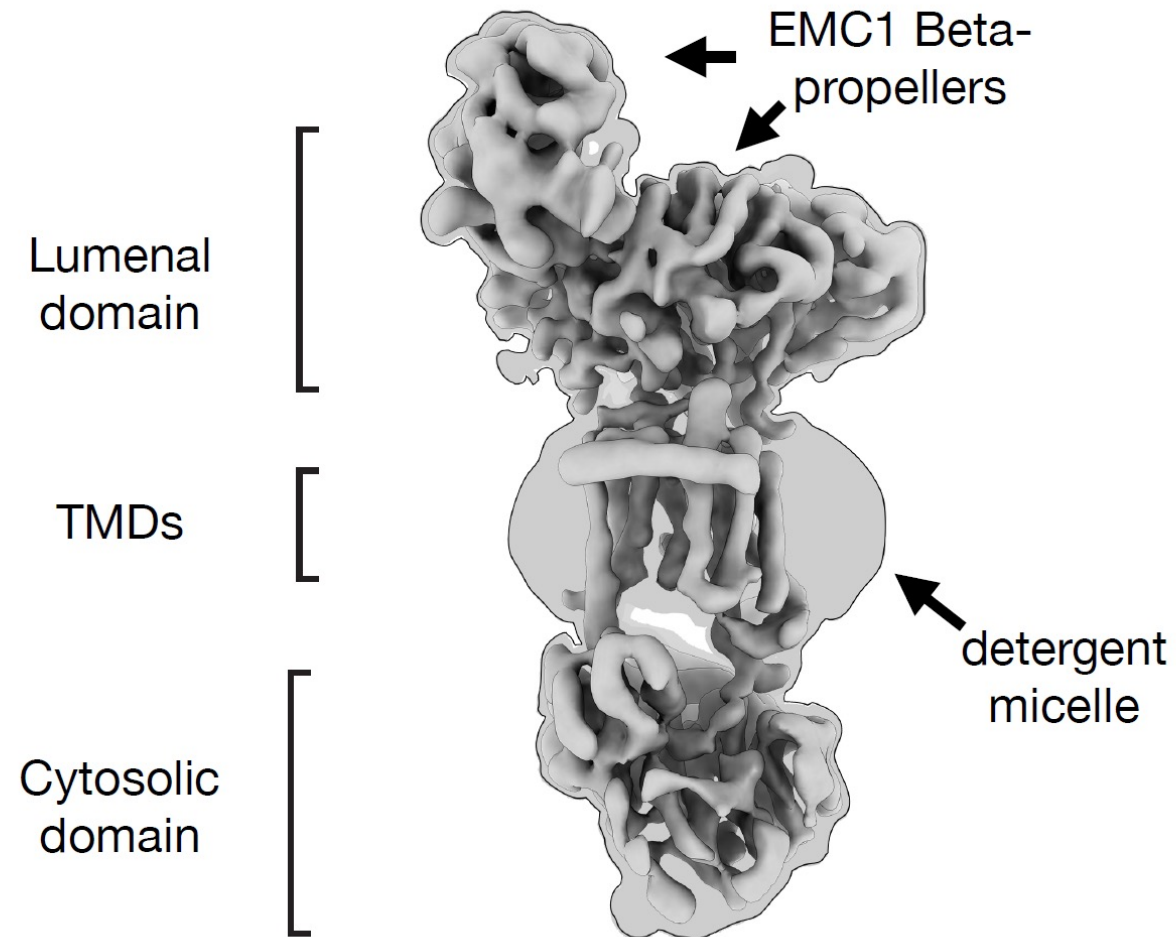
- omission of RF1
- tRNA purified from strain over-expressing amber suppressor tRNA
- Supplemented with amber suppression tRNA synthetase
- Benzoyl-phenylalanine (Bpa), a photo-crosslinking non-natural amino acid
- ³⁵S-Methionine for visualizing translations



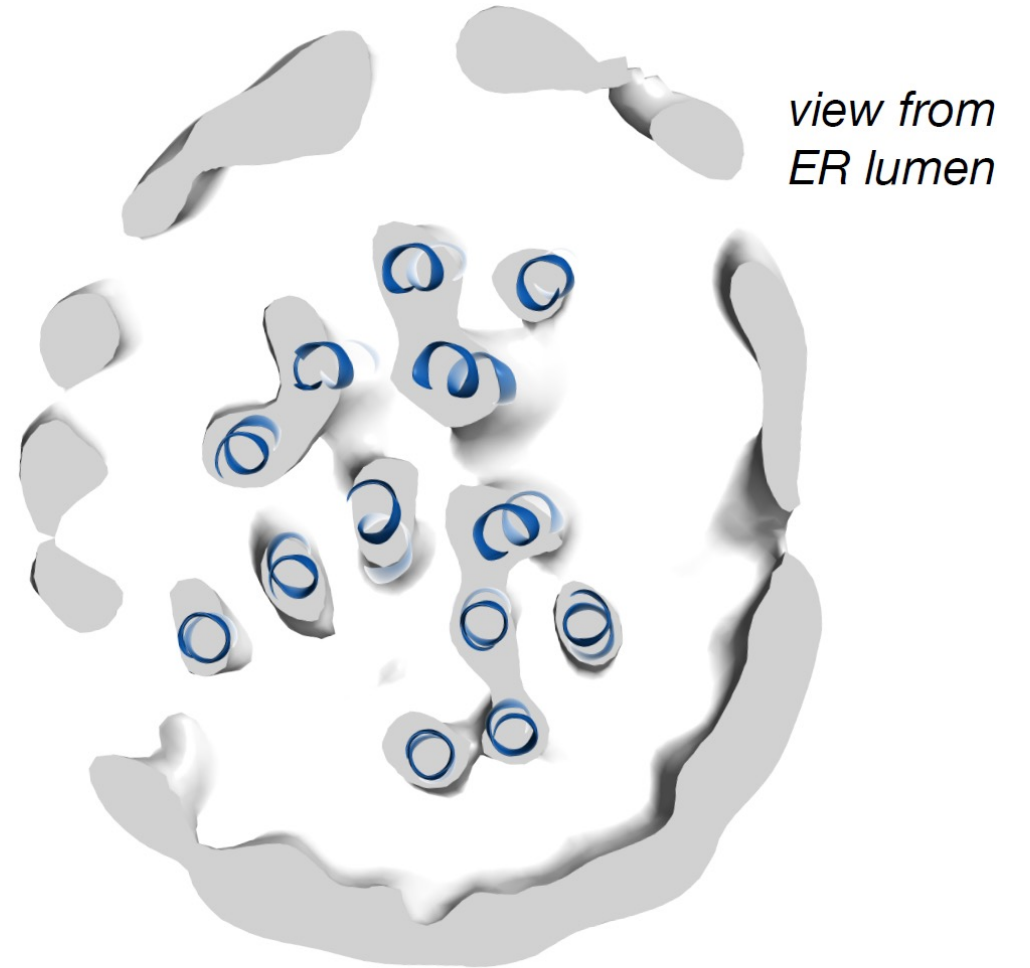
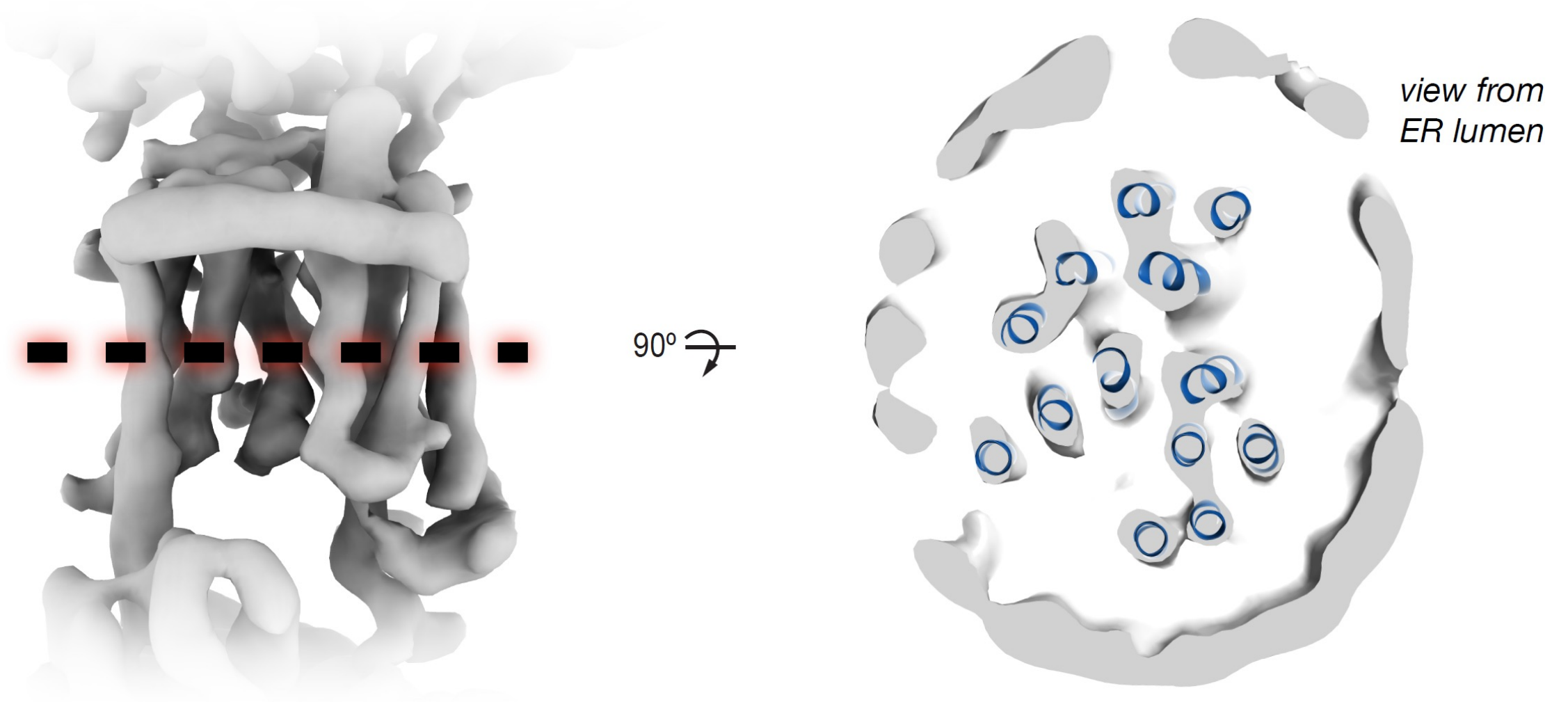
Can cytosolic EMC engage TMDs?



Low resolution cryoEM structure of EMC

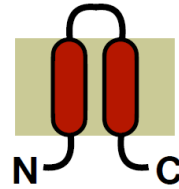


Identification of subunit TMDs

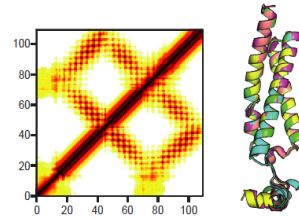


Integrated approaches to our TMD puzzle

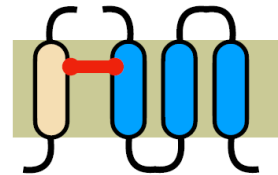
Topology and # of TMDs



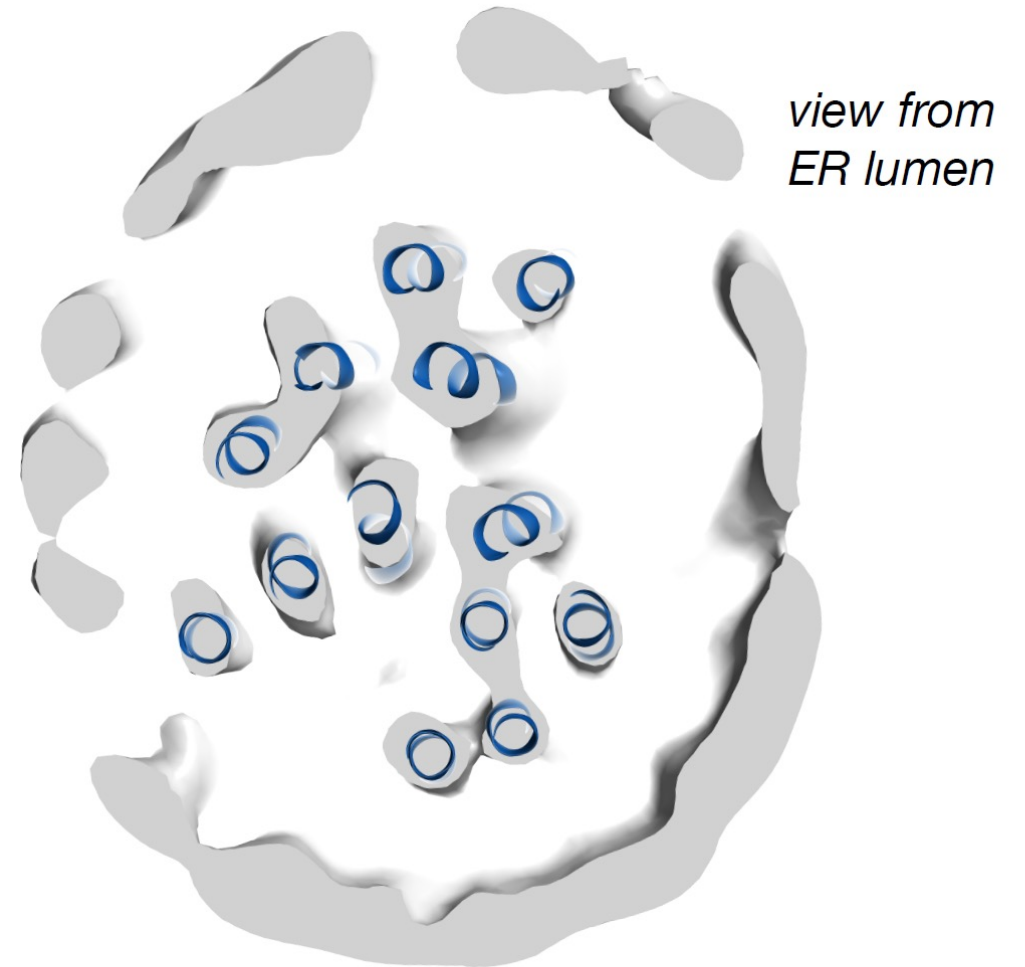
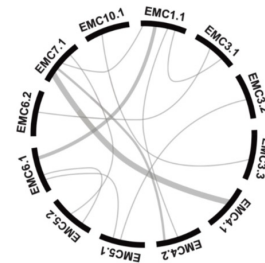
Protein prediction



In vivo photo-XL

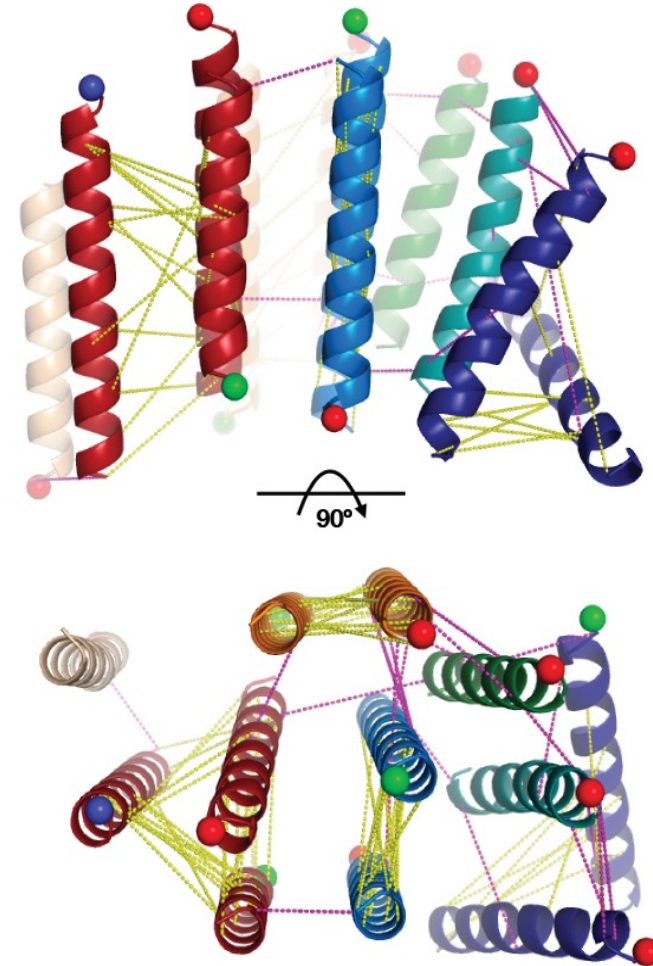
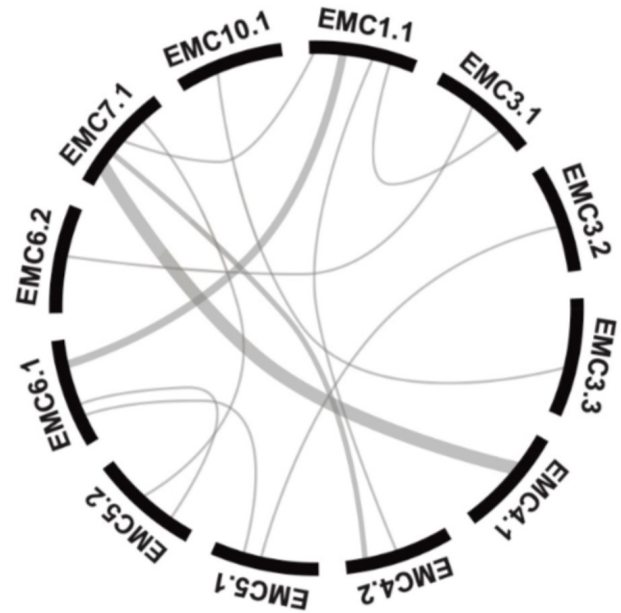


Co-evolution



Who is next to who with co-evolution

Determining intra and inter molecular contacts



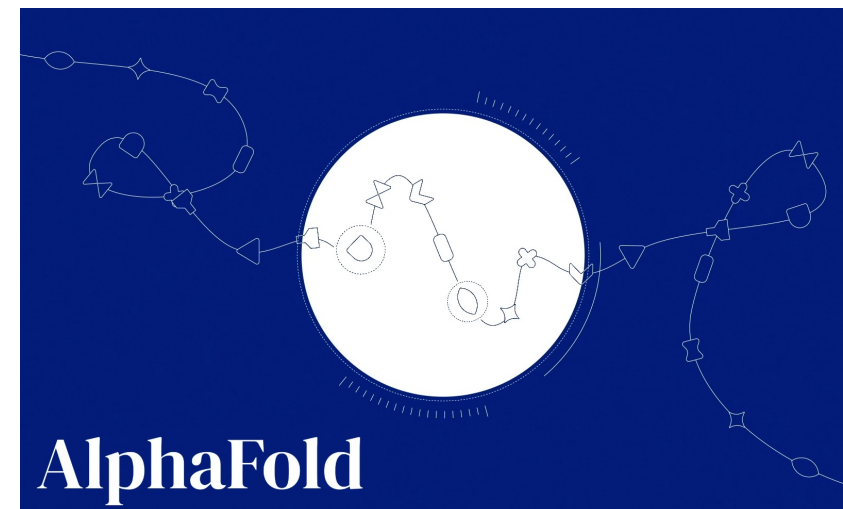
Analysis of correlated evolutionary sequence changes across proteins identifies residues that are close in space with sufficient accuracy to determine the three-dimensional structure of the protein complexes (PMID: 25255213)



AlphaFold2 structure prediction of proteins

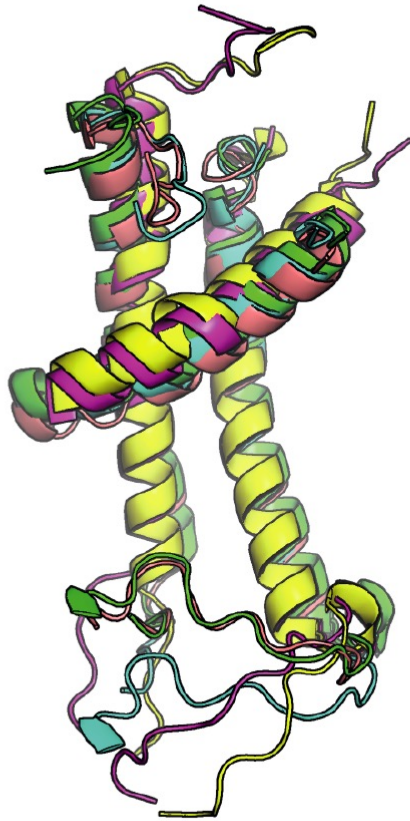
AF2 has solved the protein folding problem, but still many gaps that require experimentation

- // AlphaFold2 is an AI system that makes accurate predictions of a protein's 3D structure from its amino-acid sequence. AlphaFold Database now provides over 200 million protein predictions (<https://alphafold.ebi.ac.uk/>)
- // AF2 revolutionized structural biology but there are still limitations:
 - // Most accurate models are predictions of monomers
 - // AF multimer is very powerful in predicting protein complexes but not always accurate
 - // Cannot infer mutational analysis
 - // Does not predict ligands, cofactors, post-translational modifications
 - // AlphaFill can help infer or 'transplant' missing elements (<https://alphafill.eu/>)
- // Closing the gaps with next generation applications:
 - // AlphaLink (predict conformational changes)
 - // <https://www.nature.com/articles/s41587-023-01704-z>
 - // CombFold (predict large protein assemblies)
 - // <https://www.nature.com/articles/s41592-024-02174-0>
 - // AlphaPulldown (in situ protein interaction screen)
 - // <https://pubmed.ncbi.nlm.nih.gov/36413069/>
 - // Generative *de novo* design
 - // <https://www.nature.com/articles/s41586-023-06415-8>

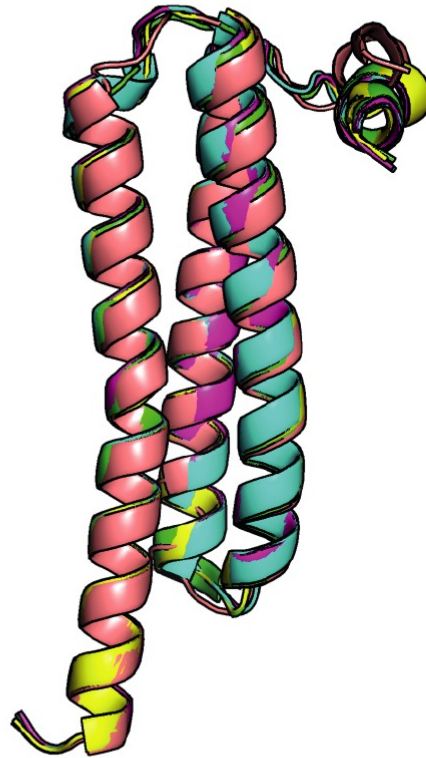


Prediction of subunits unambiguously matches density maps

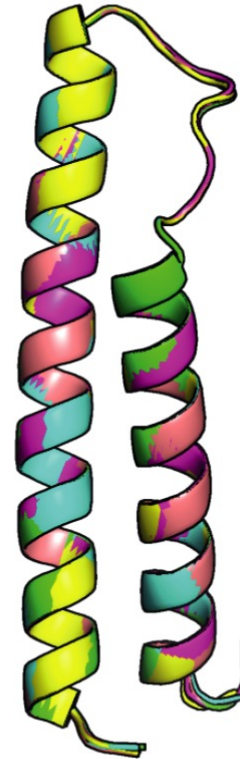
EMC3



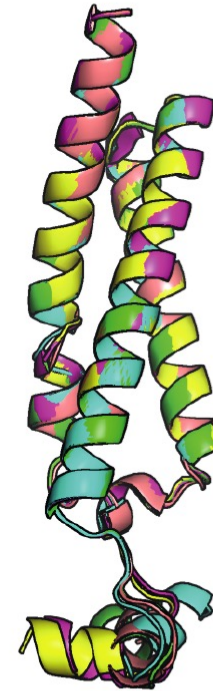
EMC4



EMC5



EMC6

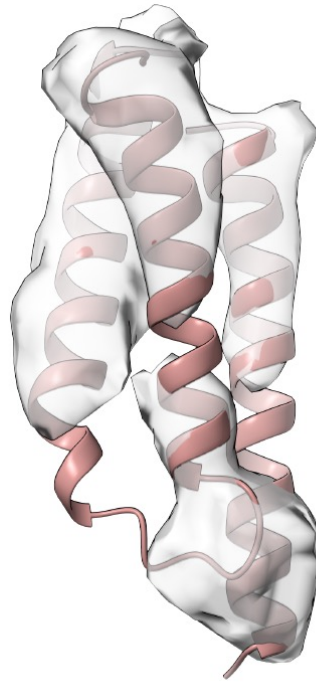


Prediction of subunits unambiguously matches density maps

EMC3



EMC4



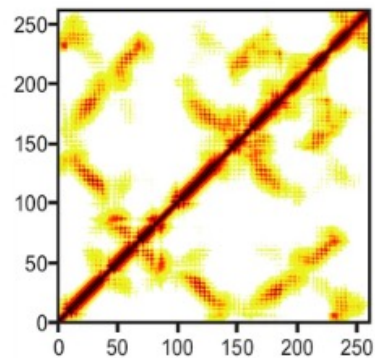
EMC5



EMC6



EMC3 (MSA = 3,733 seq.)



Model Align

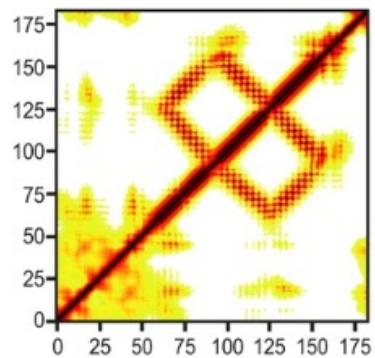


Model / Refine Align



RMSD = 4.3Å

EMC4 (MSA = 1,761 seq.)



Model Align

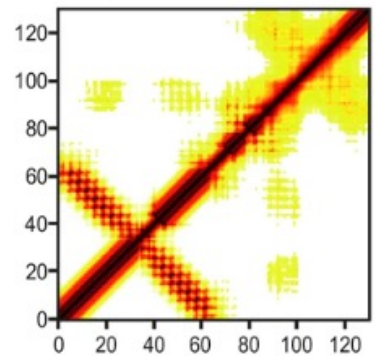


Model / Refine Align



RMSD = 1.8Å

EMC5 (MSA = 1,573 seq.)



Model Align

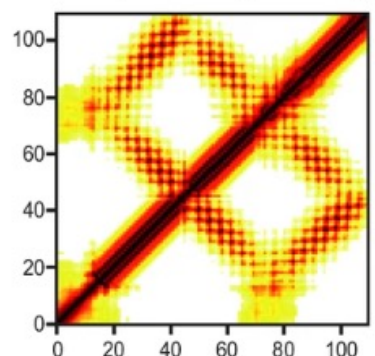


Model / Refine Align



RMSD = 1.0Å

EMC6 (MSA = 2,372 seq.)



Model Align

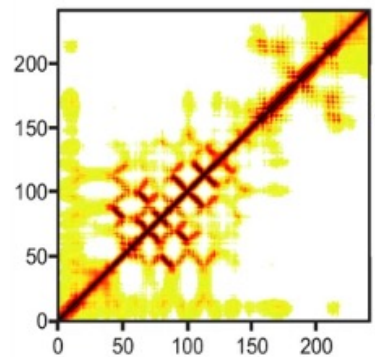


Model / Refine Align



RMSD = 2.4Å

EMC7 (MSA = 1,788 seq.)



Model Align

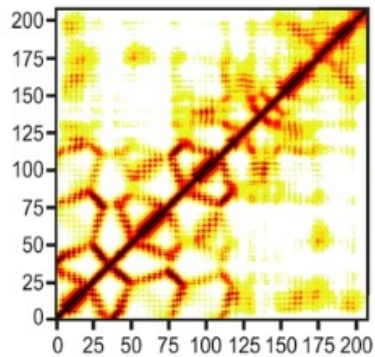


Model / Refine Align



RMSD = 2.4Å

EMC9 (MSA = 2,929 seq.)



Model Align



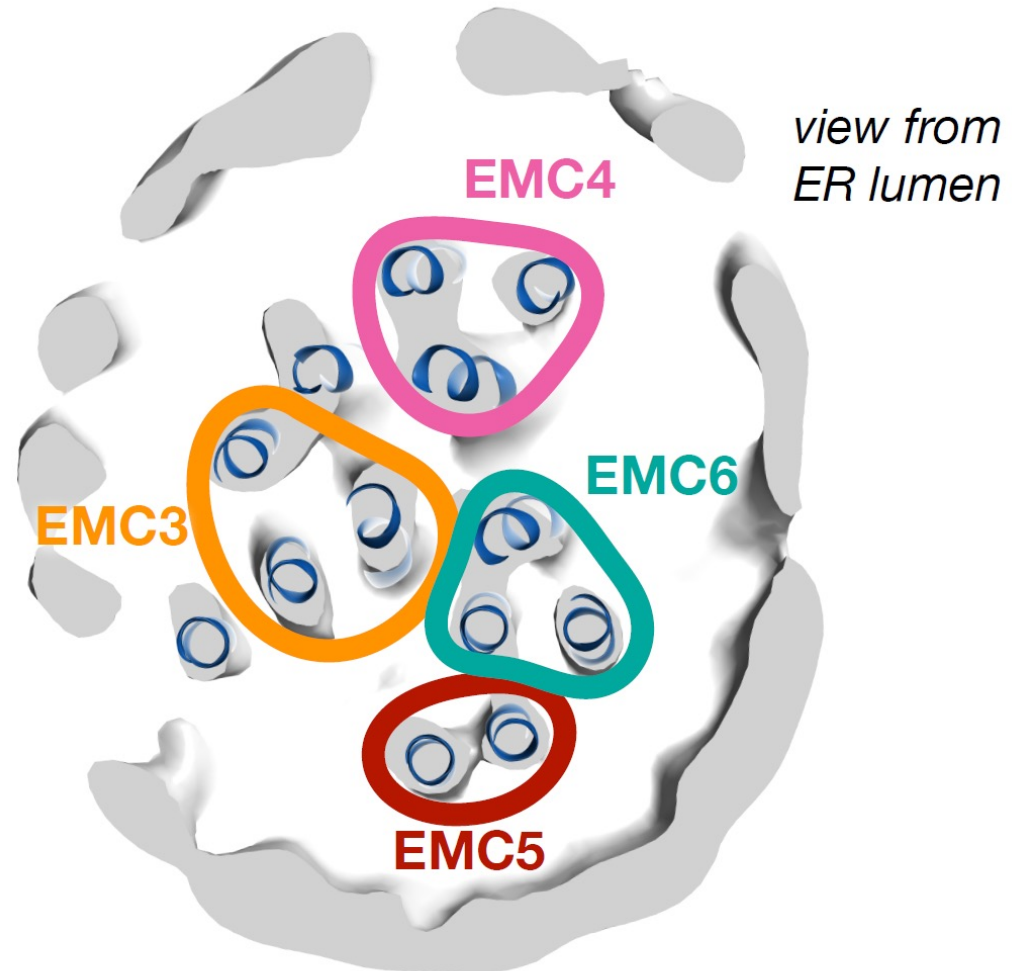
Model / Xtal Align



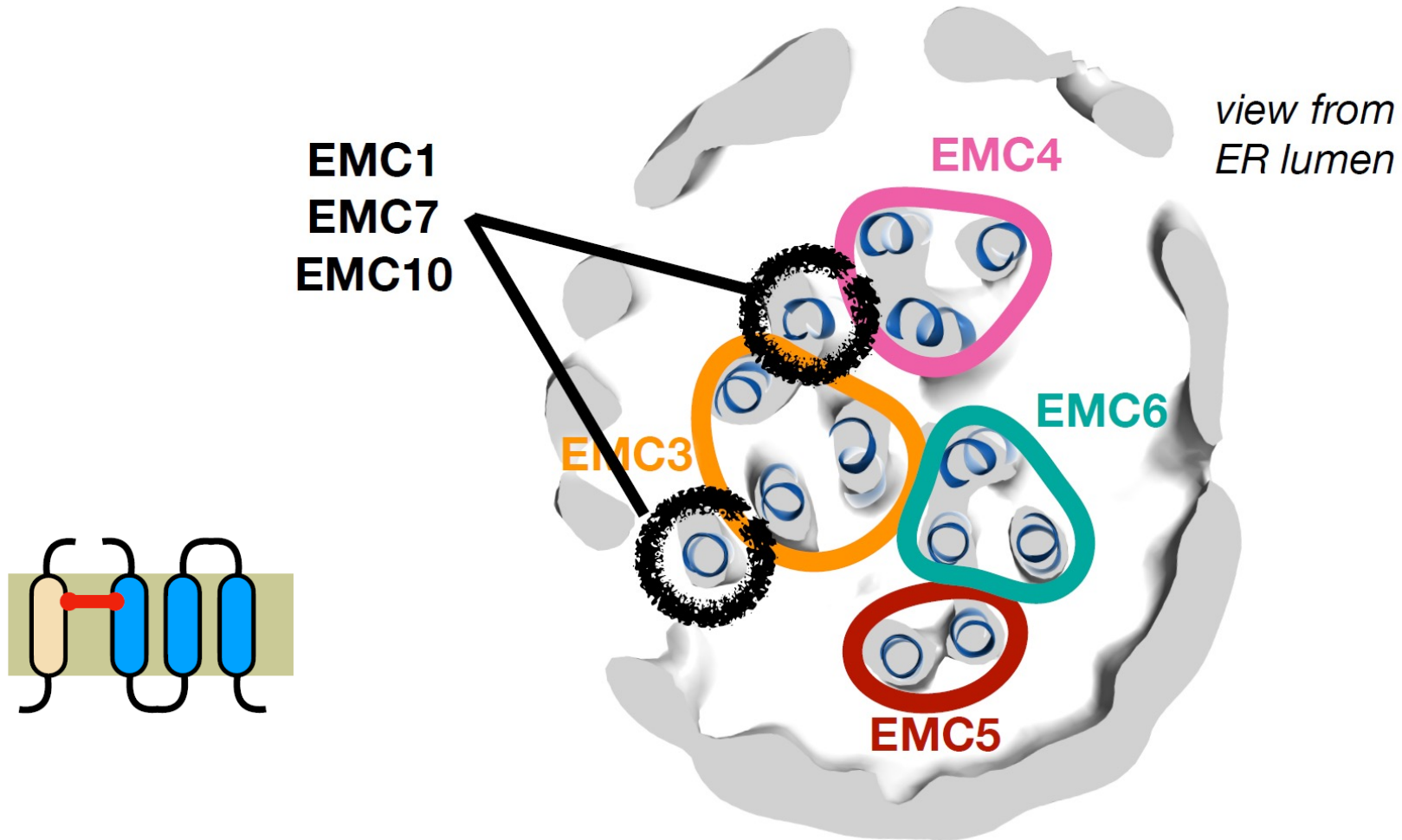
RMSD to 6Y4L = 3.5Å



Placement of multi-TMD subunits

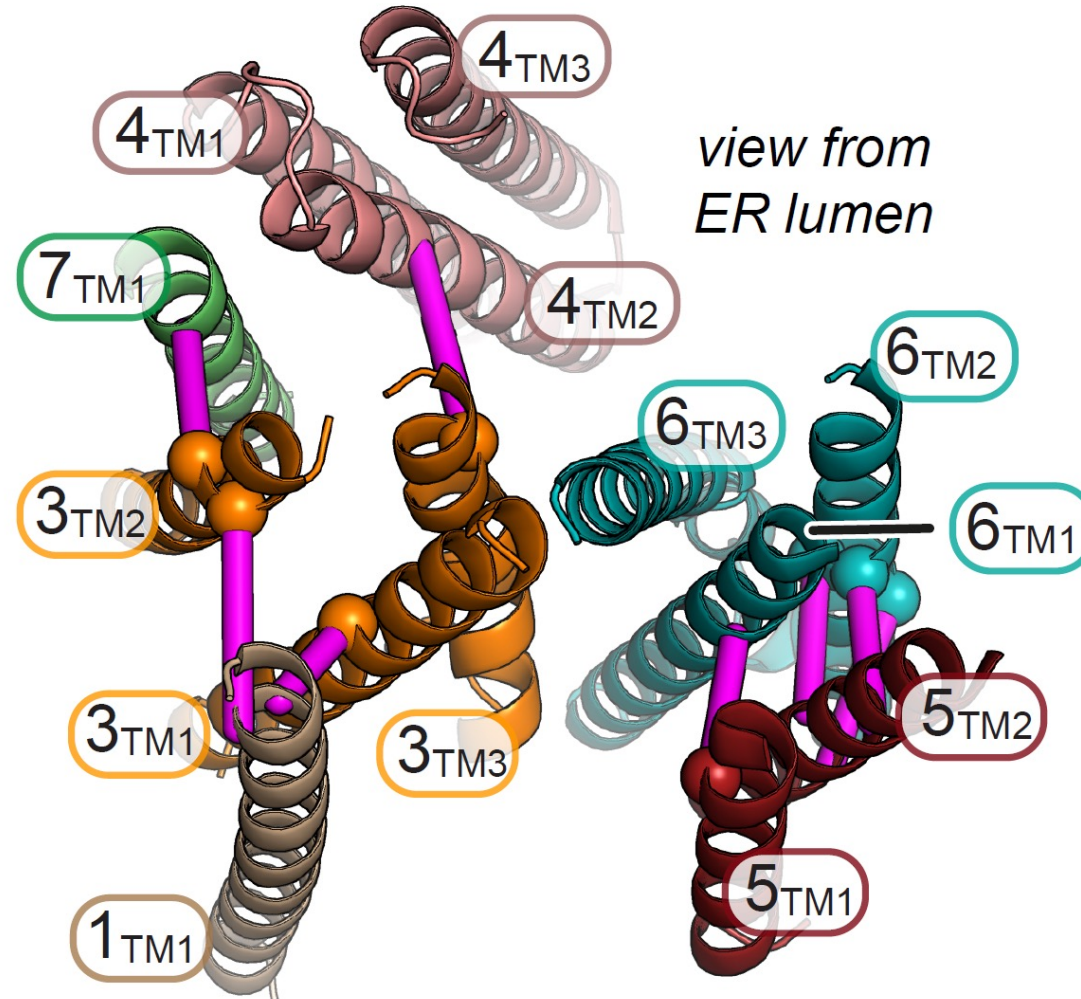


Placing single pass TMDs with in vivo photo-crosslinking

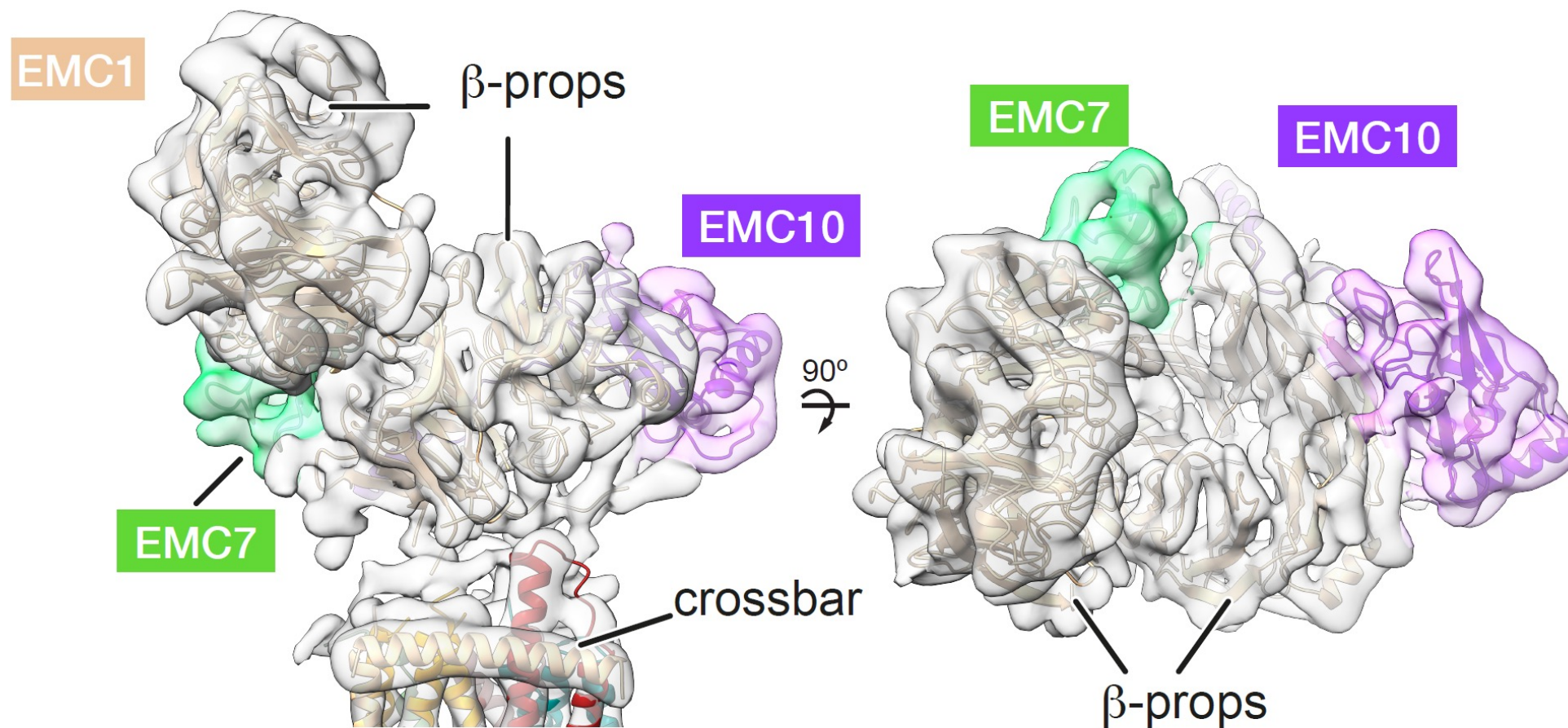


Organization of membrane subunits

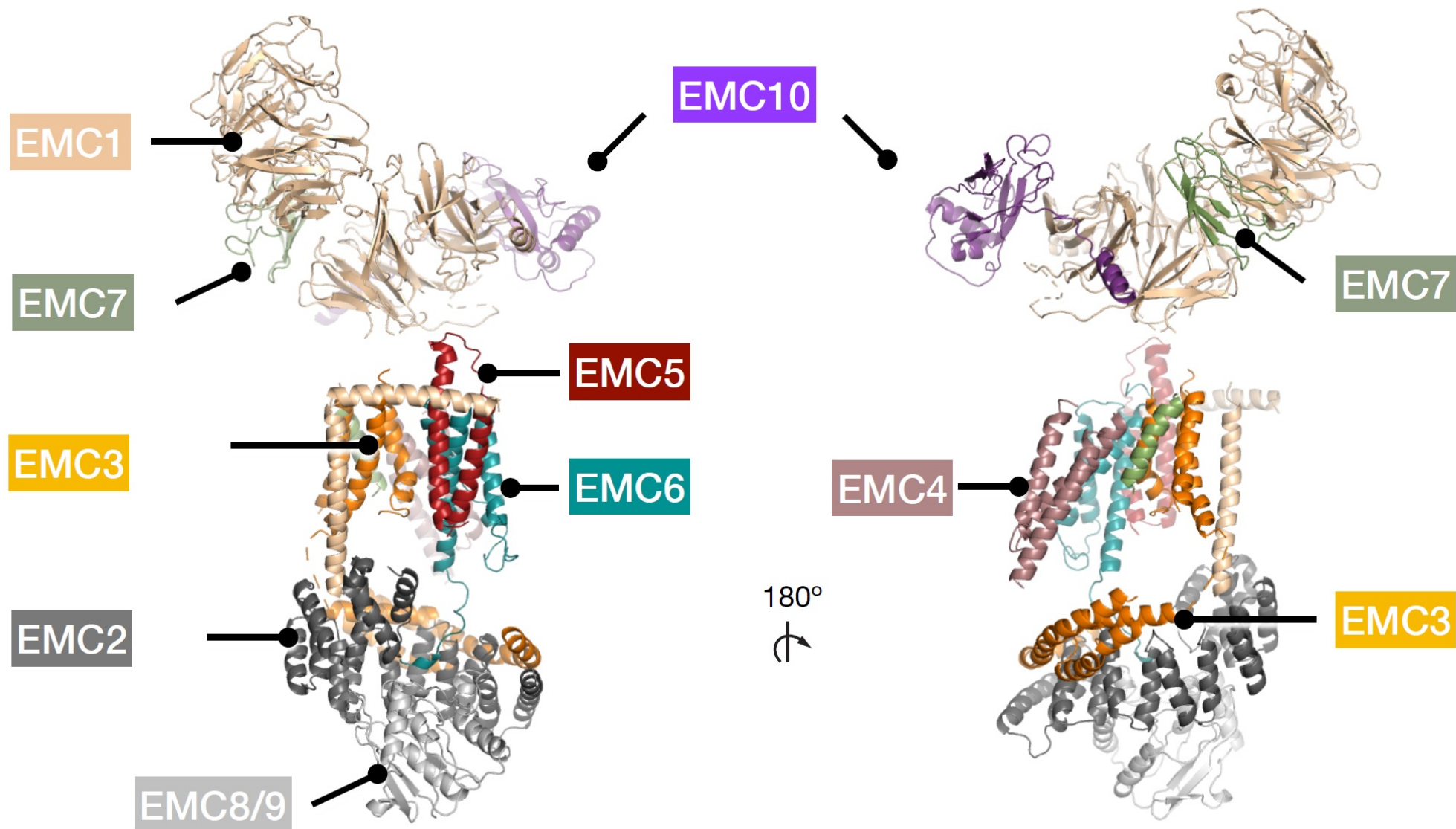
CryoEM, AI prediction, and in cell crosslinking



Protein prediction for models of the luminal subunits



Composite model of EMC based integrated approaches

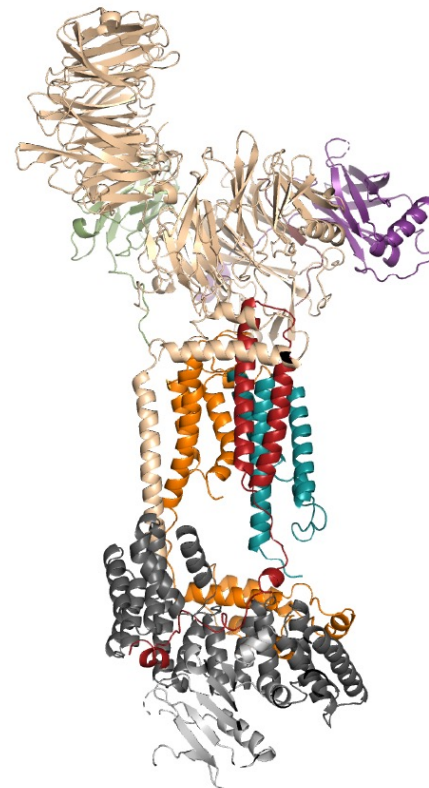


In agreement with subsequent full-length structure

Human EMC 2.2/6.5Å
PDB: 6Y4L/6Z3W

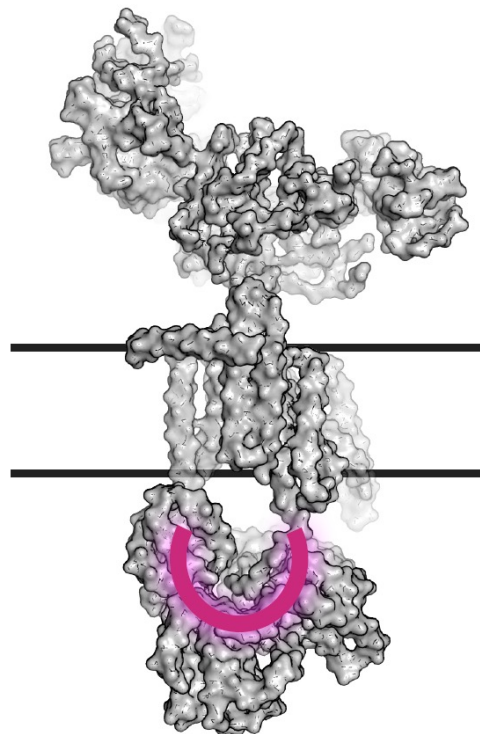


Human EMC 3.5Å
PDB: 6WW7

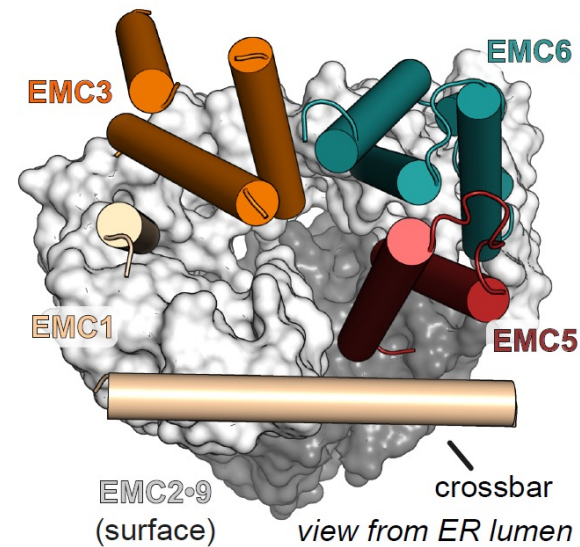


Mechanism of EMC mediated membrane protein insertion

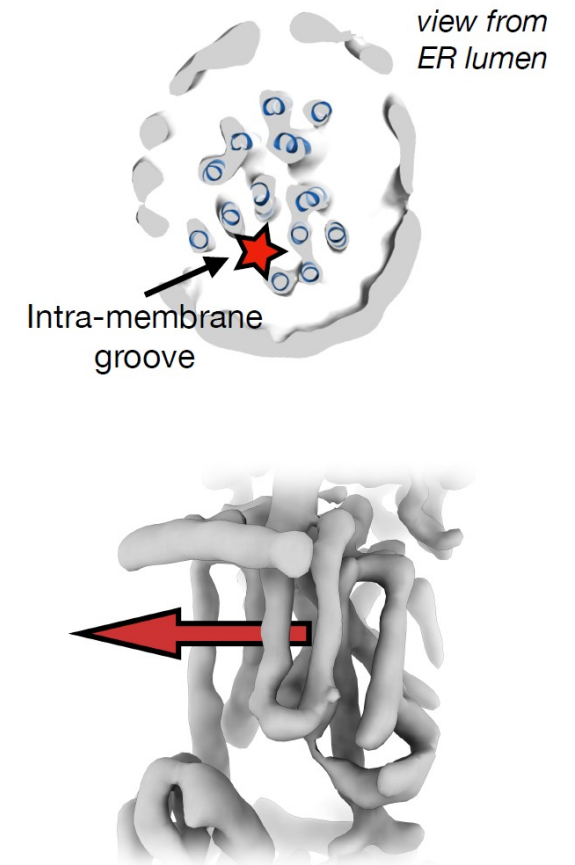
1. Cytosolic EMC subunits bind to client TMDs



2. Pathway links EMC's cytosolic and membrane regions




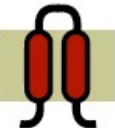
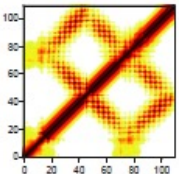
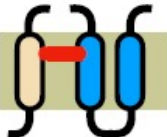
3. TMD architecture promotes insertion



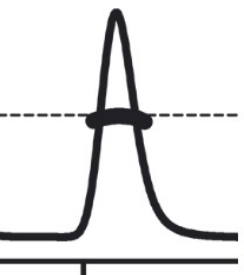
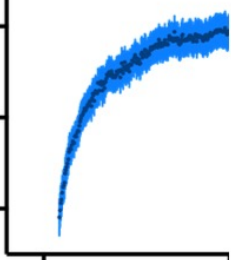
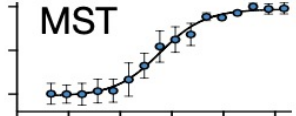
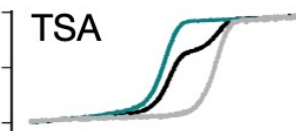
John P O'Donnell, Ben P Phillips, Yuichi Yagita, Szymon Juskiewicz, Armin Wagner, Duccio Malinverni, Robert J Keenan, Elizabeth A Miller, Ramanujan S Hegde (2020) The architecture of EMC reveals a path for membrane protein insertion eLife 9:e57887

Integrated approaches to understanding mechanism

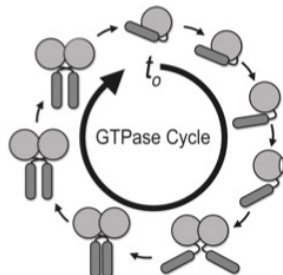
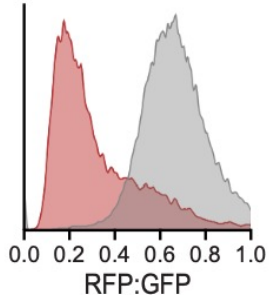
Integrated Structural Approaches

- Crystallography & cryoEM 
- Topology & TMD # 
- Modeling 
- Photo-XL 

Biophysical Analysis

- SEC-MALS 
- Kinetics 
- MST 
- TSA 

Assays & Cellular Function

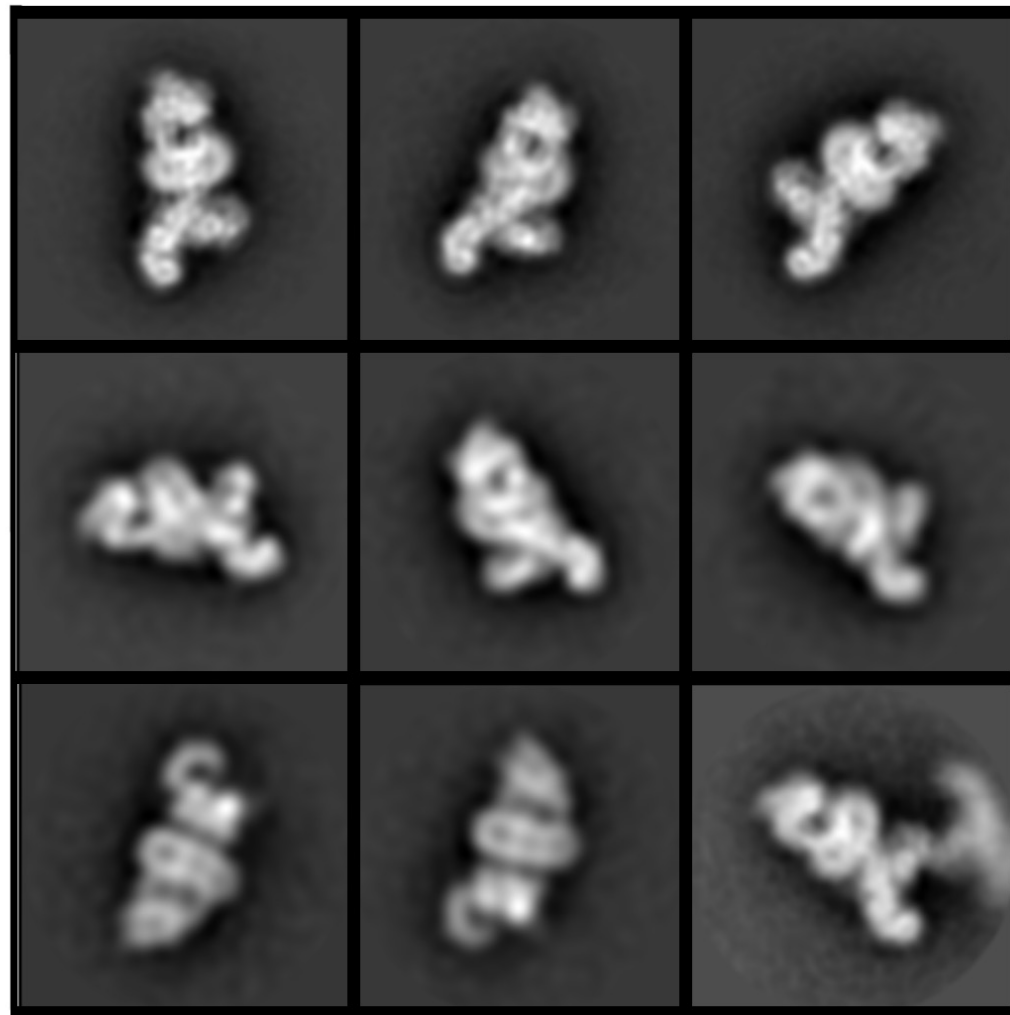
- GTPase Cycle 
- RFP:GFP 



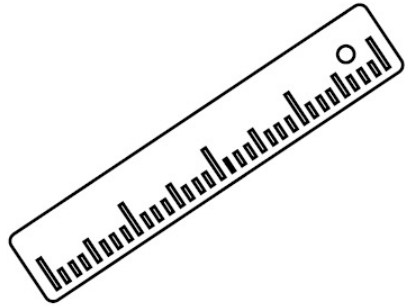
Thank you



2D classification of particles



Size of complex



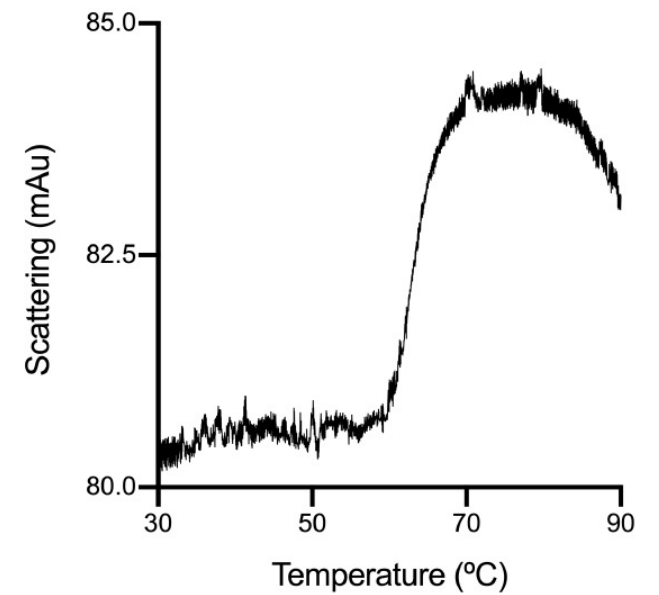
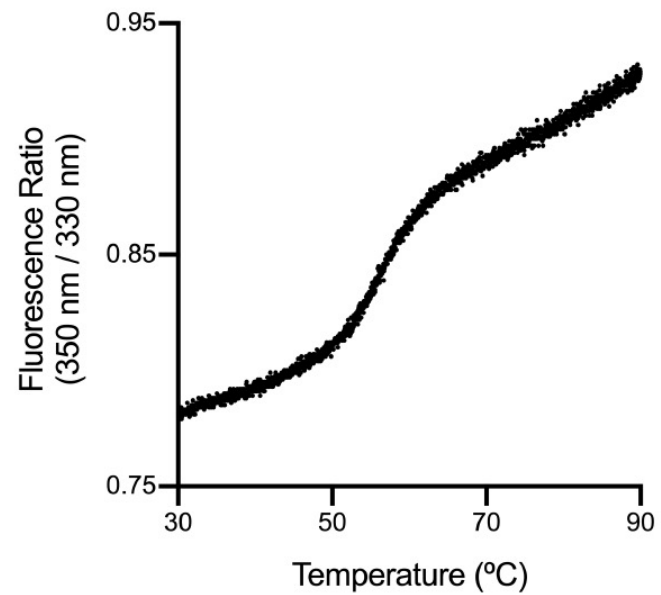
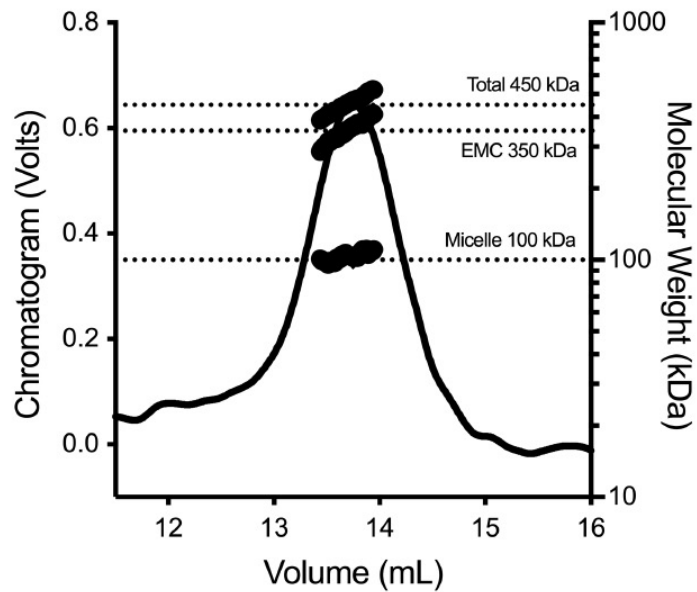
Thermal Unfolding



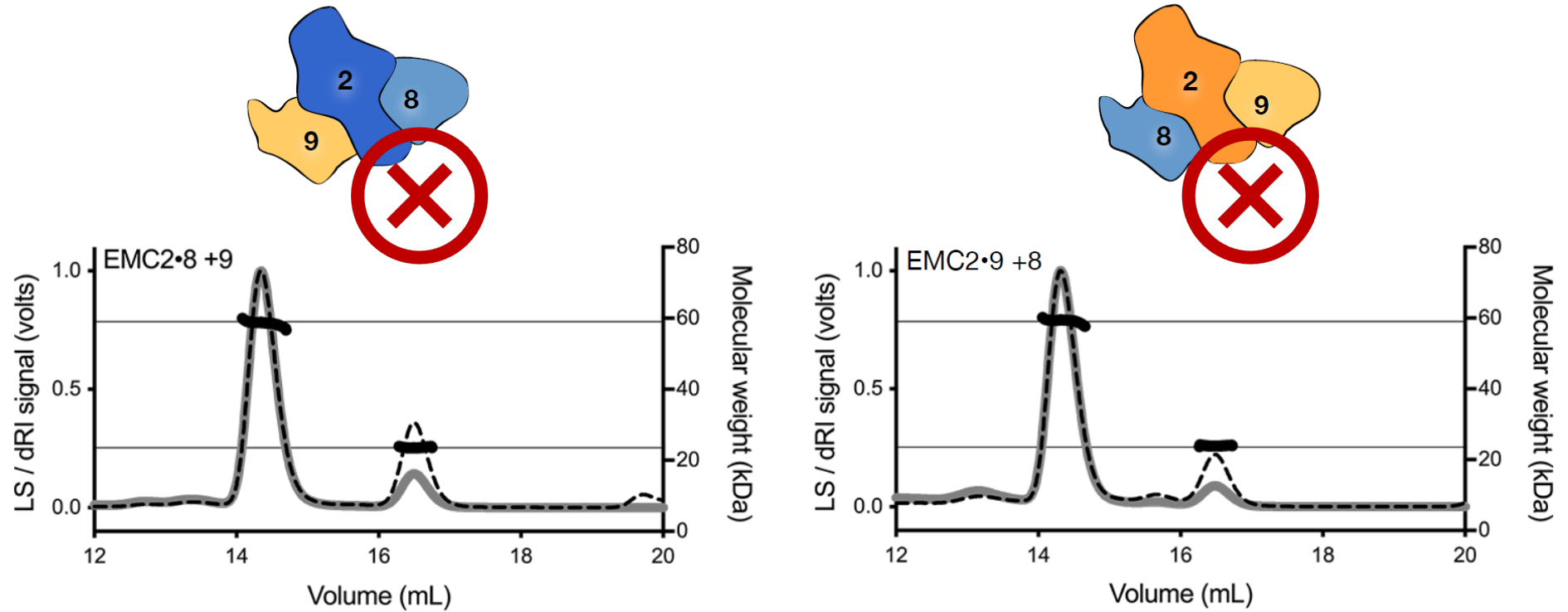
Aggregation



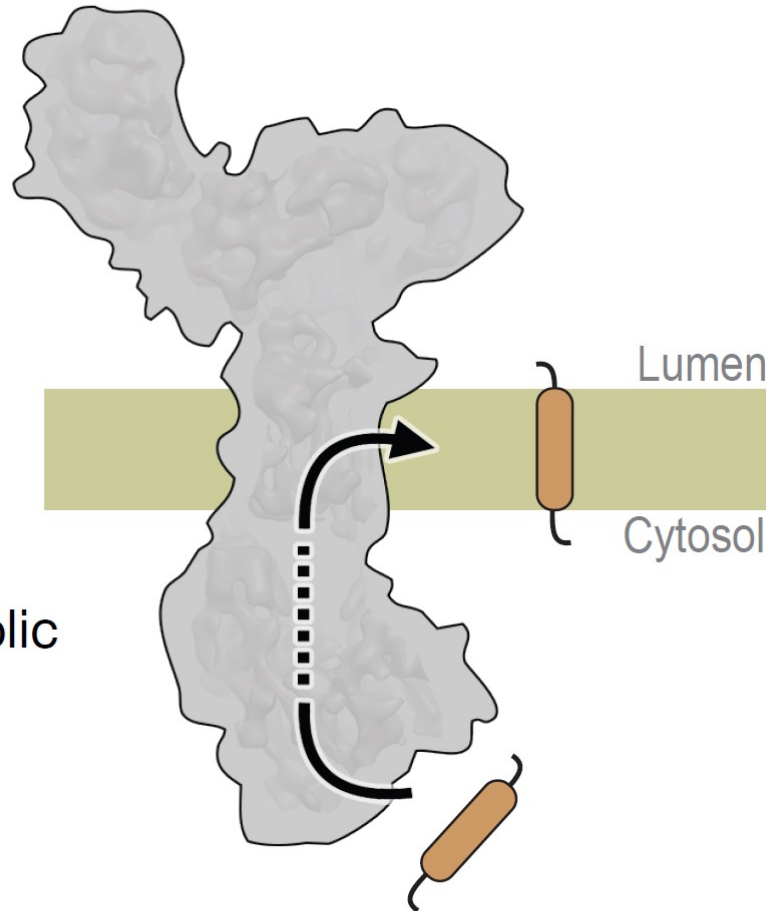
Heptagon [CC BY-SA 3.0 (<https://creativecommons.org/licenses/by-sa/3.0>)]



What is the stoichiometry of the complexes?



How does EMC facilitate insertion?

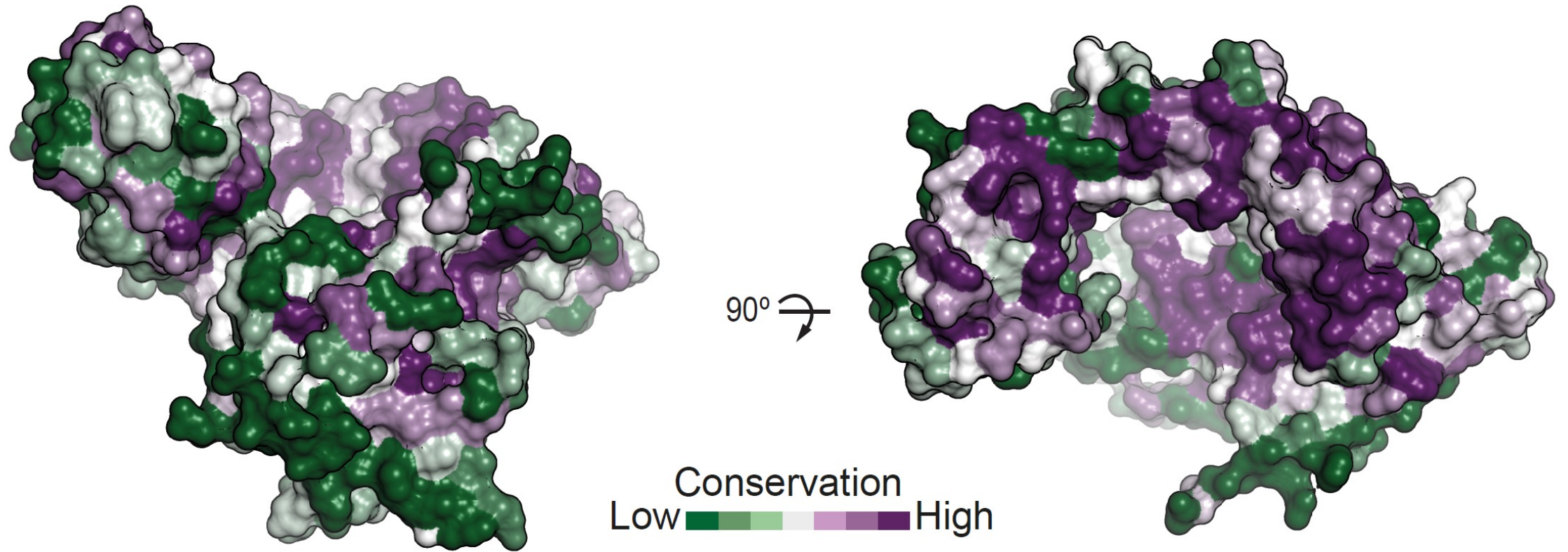


2. Pathway links EMC's cytosolic and membrane regions

3. TMD architecture promotes insertion

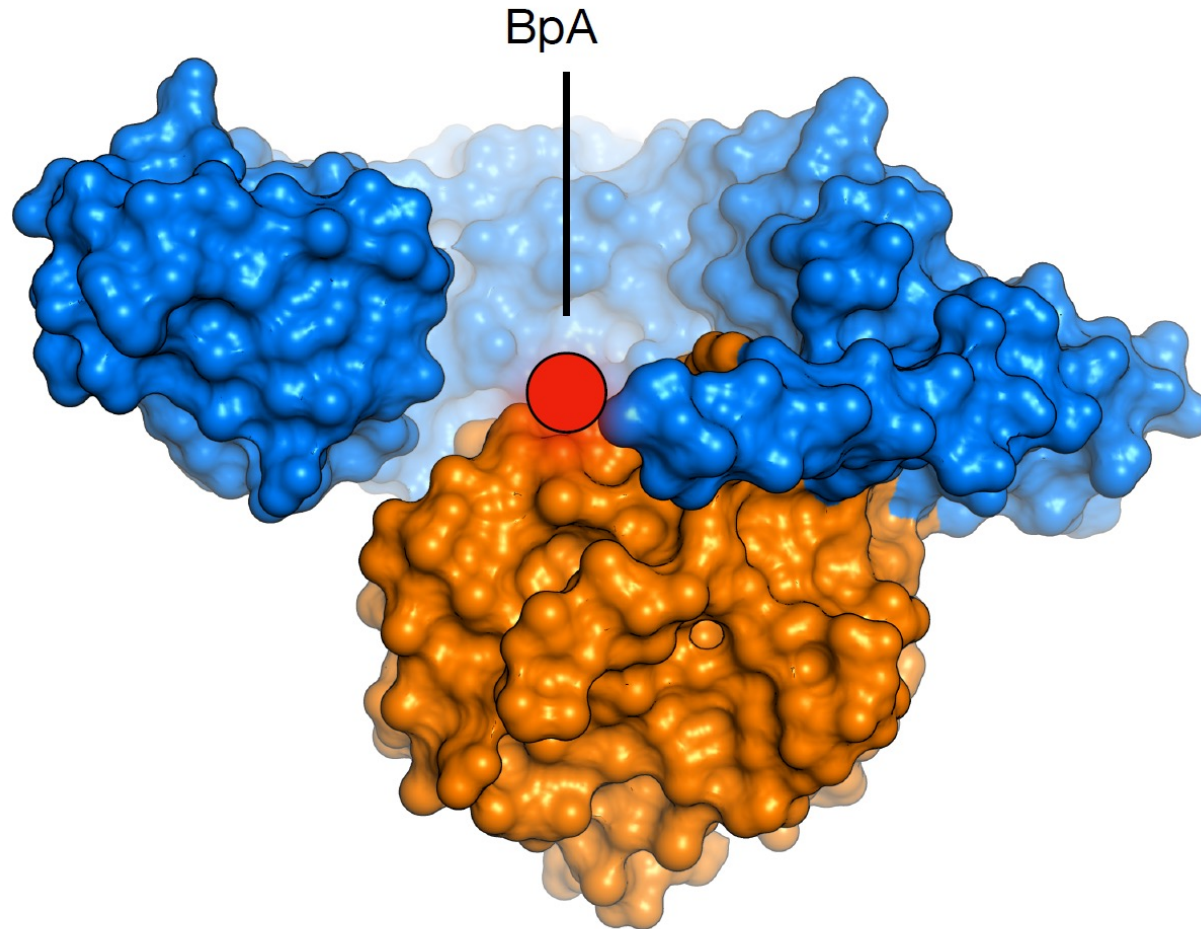
1. Cytosolic EMC subunits bind to client TMDs

Cytosolic vestibule and rim of EMC2 are highly conserved

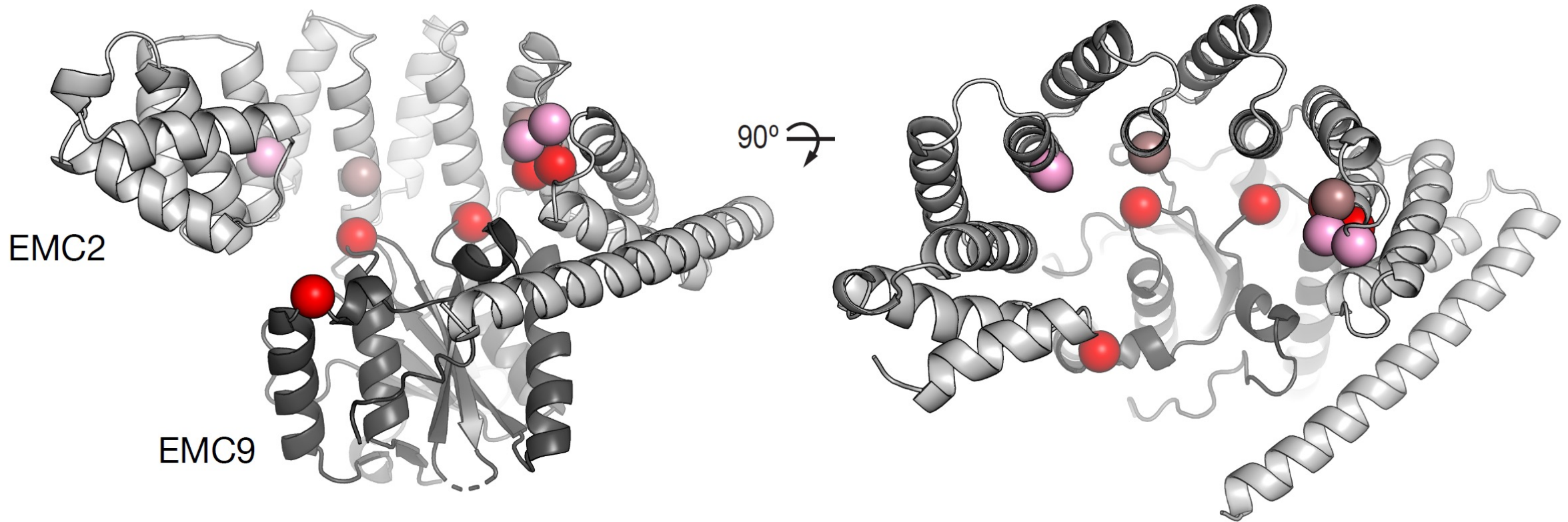




Map binding groove by crosslinking back to substrate



EMC2•8/9 engage substrate in cytosolic vestibule



Crosslink Efficiency

● High ● Med ● Low



In vivo SQS insertion reporter assay





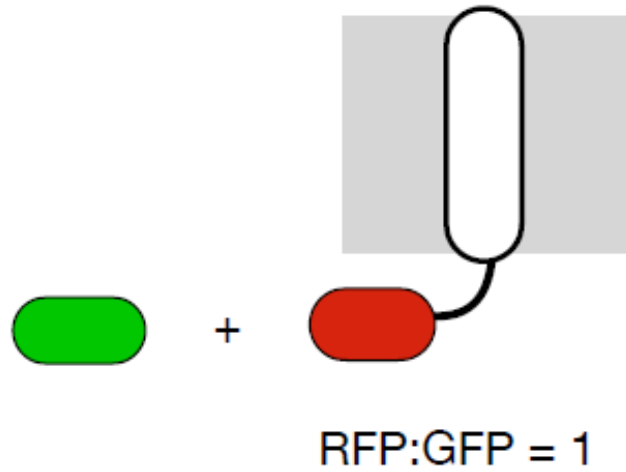
In vivo SQS insertion reporter assay



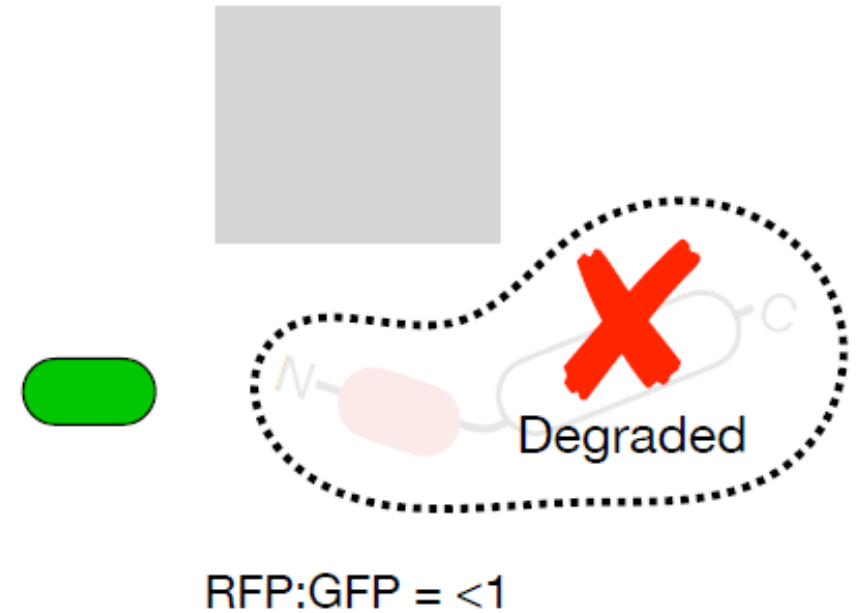
In vivo SQS insertion reporter assay



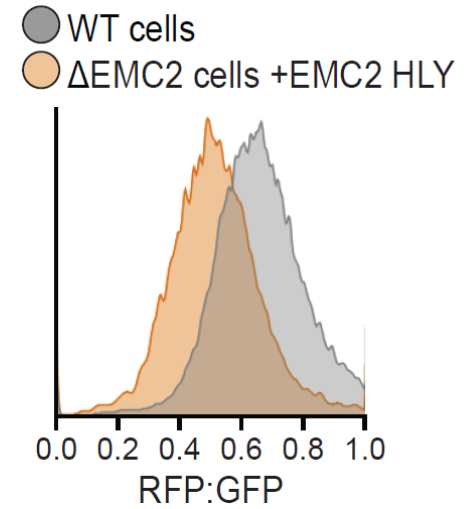
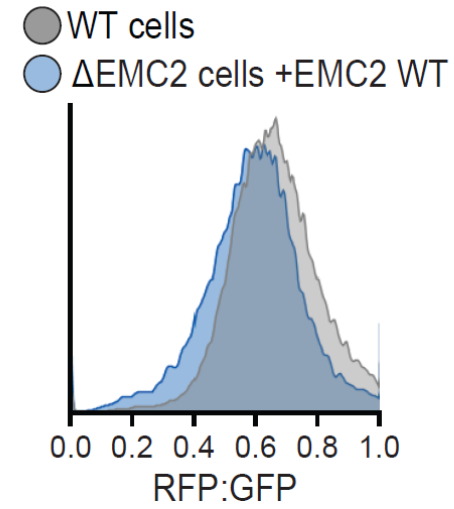
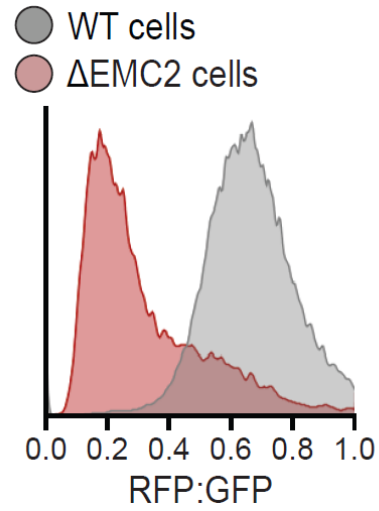
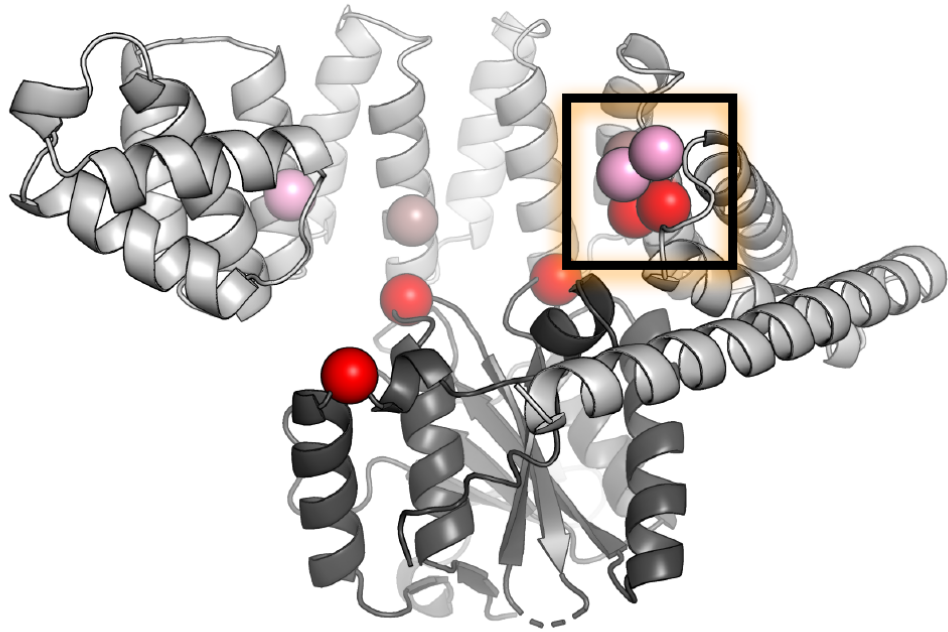
Proper insertion



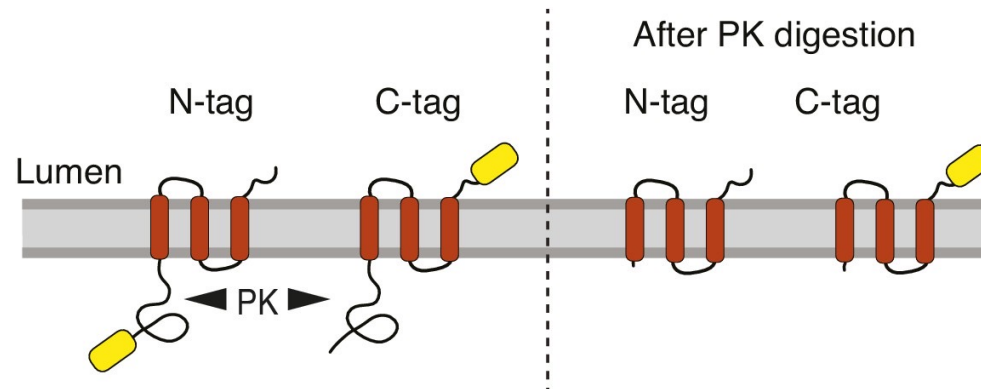
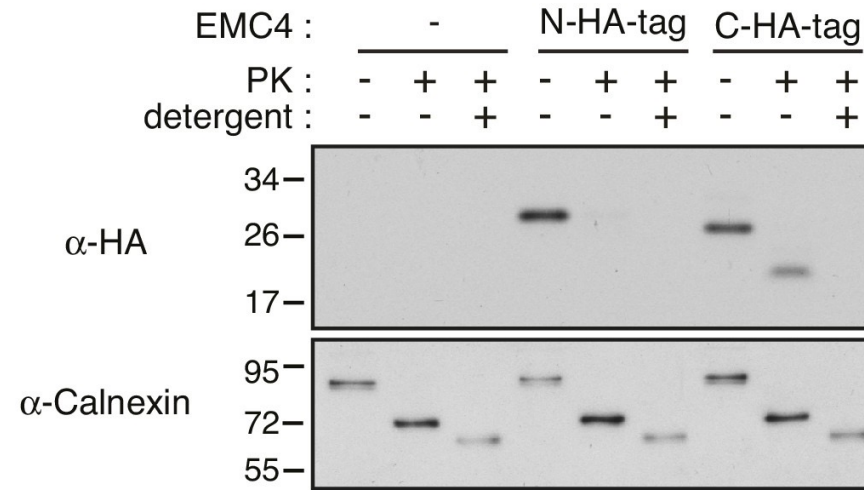
No insertion



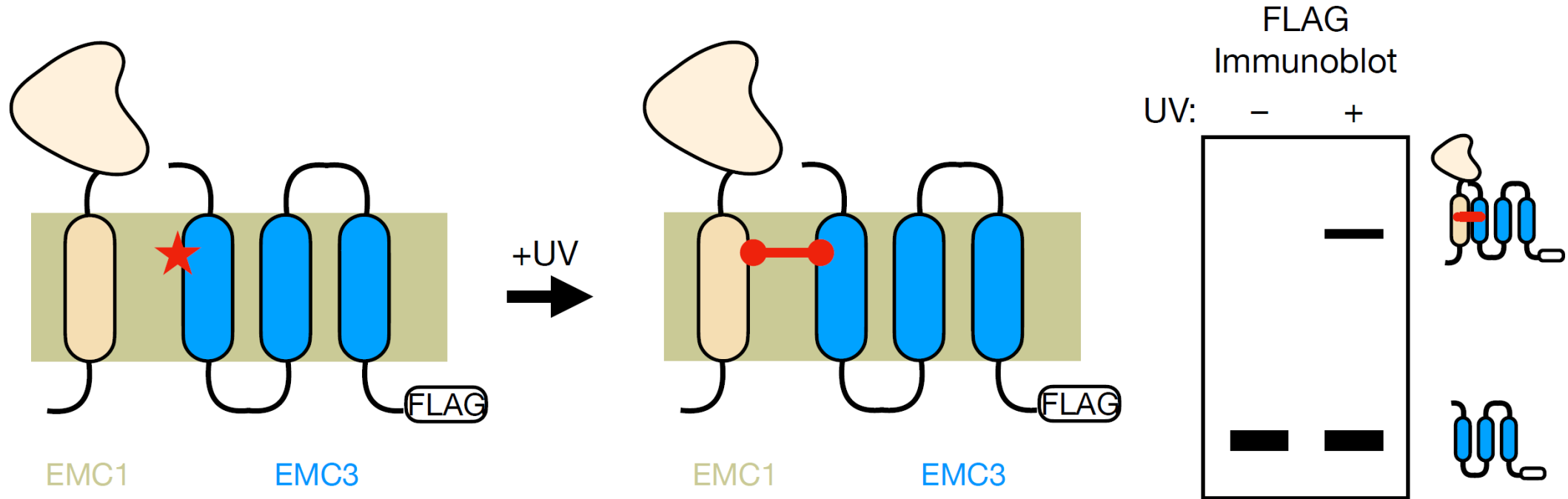
Mutation of vestibule impedes insertion



Membrane protein topology and # of TMDs

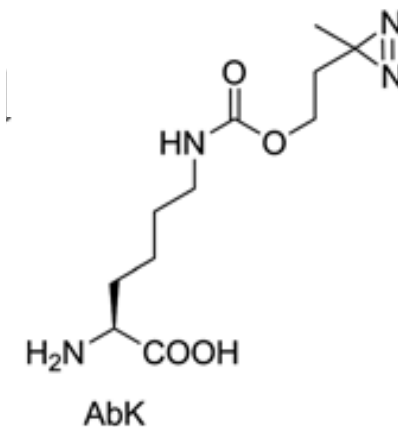


Example cellular crosslink



Site-specific photo-crosslinking in mammalian cells

- *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (PylRS) and tRNA^{Pyl}_{CUA} pair
- photo-crosslinking amino acid AbK
- UV irradiated in cells in native state



Placement of single pass TMD subunits

