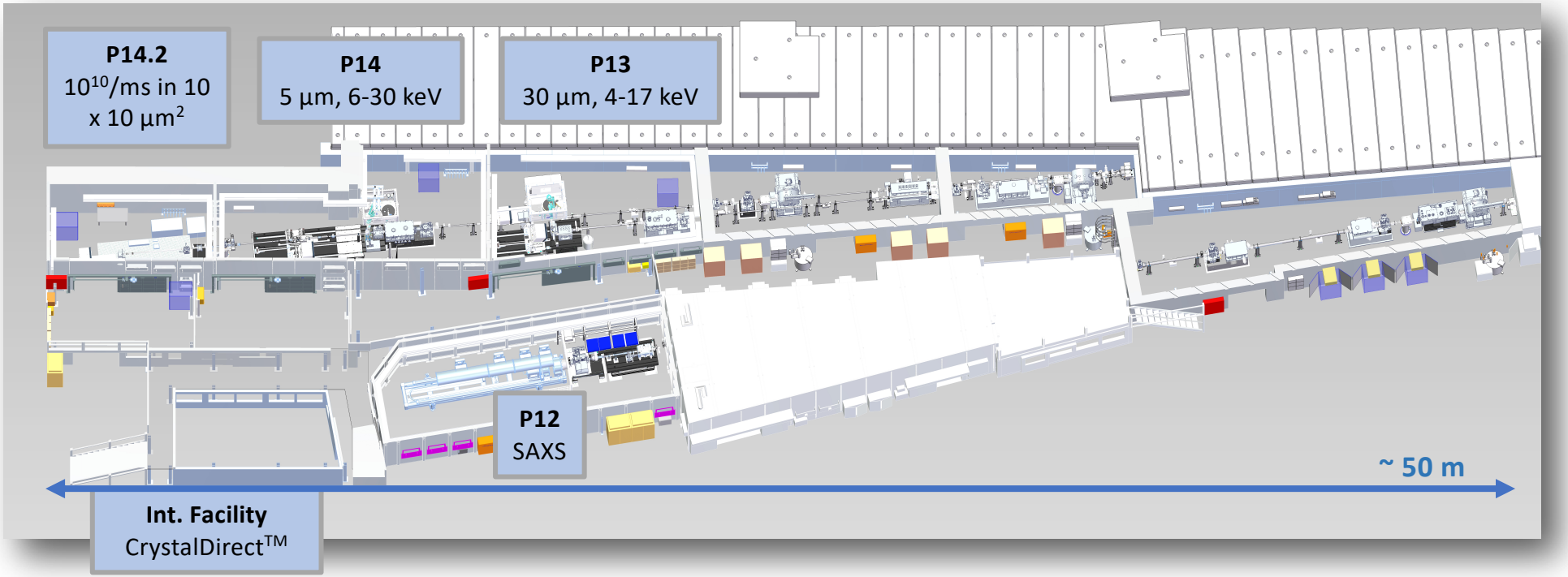
Sample quality control for Structural Studies

Maria Garcia Alai
EMBL Team Leader

EMBL

25th Sep. 2019

3 Beamlines at PETRA III



- EMBL Hamburg operates the entire beamlines beginning at the ‘frontend’
- The beamlines are embedded in the Integrated Facility for Structural Biology

SPC

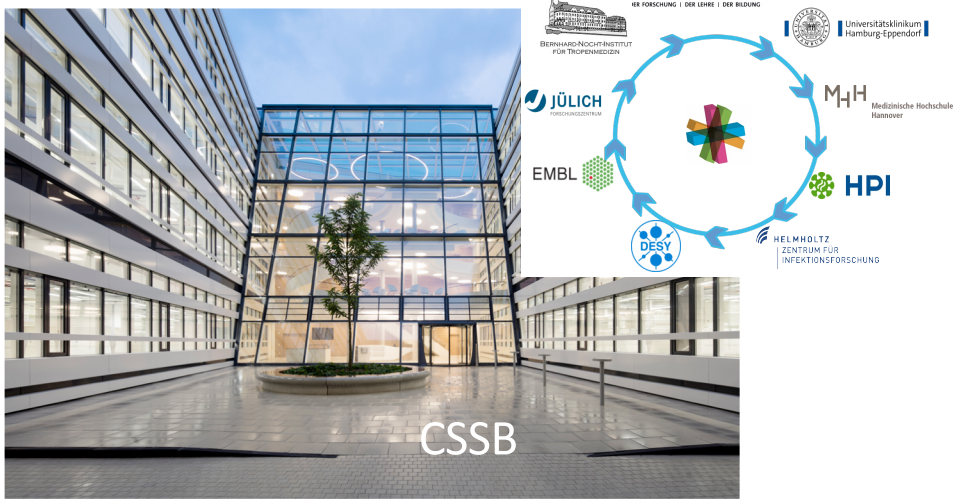
EMBL

Sample Preparation Characterization & Crystallization

High-throughput crystallization

HTX

Protein Characterization
(Molecular Biophysics)

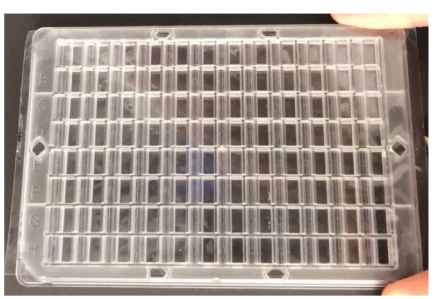


PC

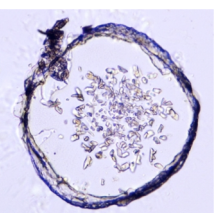
High-throughput Crystallization Facility

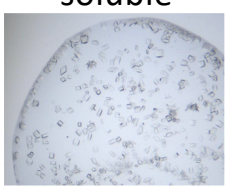
CRIMS

Screens	Robots	Plates	Imaging
• Initial screening 16 • Optimization 13 • LCP • Additive • Customized screens	• 2 TTP mosquito • Scorpion	• CrystalDirect • Swissci • Intelli • LCP • Hanging drop • Microbatch	• Formulatrix • 19 ºC and 4 ºC

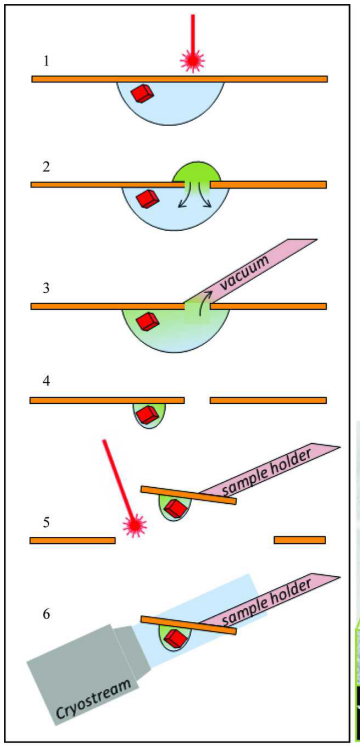


CRIMS

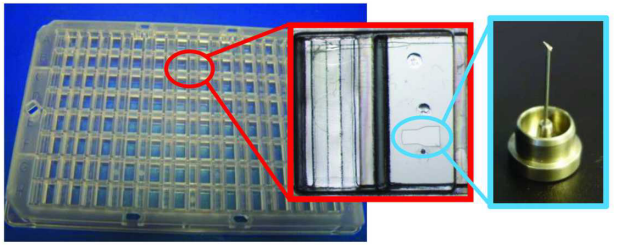




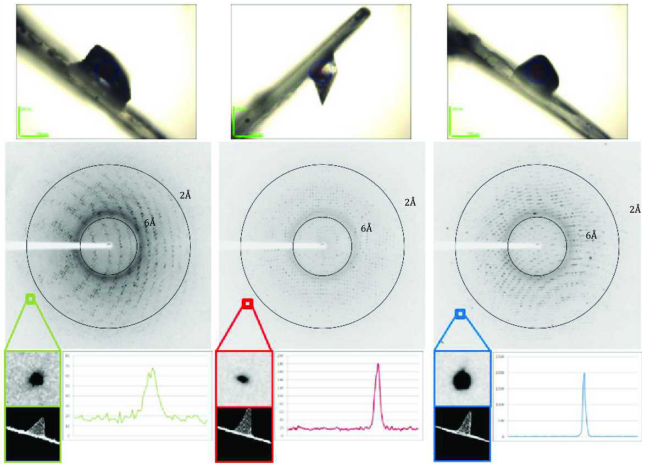
Crystal Direct Harvester



(a)

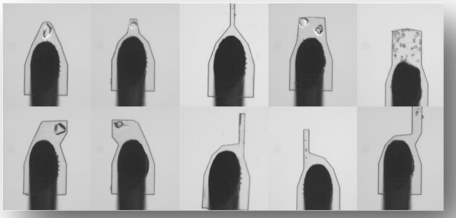


(b)



(c)

Zander et al. (2016). Automated harvesting and processing of protein crystals through laser photoablation.
Acta Cryst. (2016). D72, 454-466

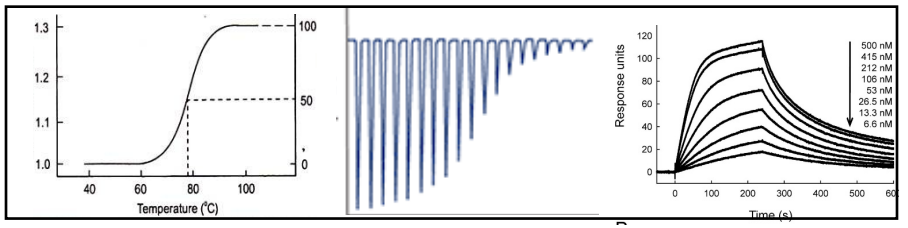


A robotic **CrystalDirect Harvester** (arinax.com) system for (remote) harvesting and cryo-cooling of crystals is operational

Molecular Biophysics Platform

Characterization | **Interaction** | **Kinetics**

iLAB



1) 

MALDI TOF: Bruker/CovalX

1. MS
2. TF
3. nDSF
4. DLS
5. CD
6. FT-IR

7. ITC
8. MST

9. SPR
10. Interferometry
11. Stopped-flow

8) 

Microscale Thermophoresis
Monolith NT.115 and NT.LabelFree
Nanotemper

6) 

Infrared spectroscopy
Bruker Vertex 70

10) 

Bioforte Octet Red96

9) 

Biacore T200 GE

2) 

Thermofluor

4) 

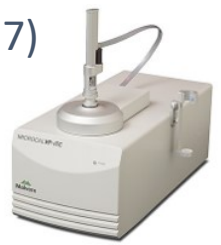
Differential light scattering
Wyatt

3) 

Differential Scale Fluorimetry
Prometheus 48
Nanotemper

5) 

Circular Dichroism
AppliedPhotophysics

7) 

Isothermal titration calorimetry
MicroCal

11) 

Stopped-flow
AppliedPhotophysics

Outline:

-Quality control
-Protein Folding
-Biophysical characterization

Quality control

-What does it mean “Protein quality control”
And… why do we care?

Quality control of purified proteins to improve research data reproducibility:

Little pain, lots to gain?

ARBRE-MOBIEU (Association of Resources for Biophysical research in Europe – MOlecular BIophysic in EUrope) and P4EU (Protein Production and Purification Partnership in EUrope)

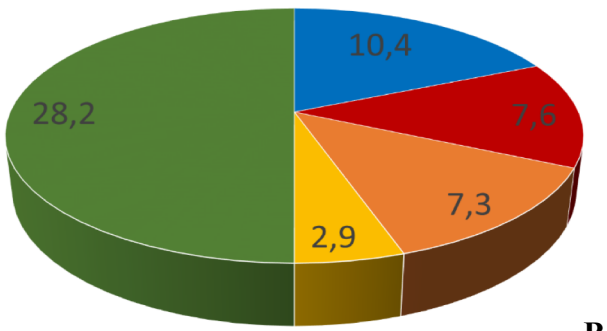


Figure 1. Categorization of preclinical research spending in the US into levels of reproducibility and common errors leading to data irreproducibility (2012 data, all figures in US $billions, adapted with permission from reference (3)). The 'Biological Reagents and Reference Materials' category includes approximately $0.5bn spent on poor quality commercial antibodies (4).

Begley, C.G. & Ioannidis, J.P. Reproducibility in science: improving the standard for basic and preclinical research. Circ. Res. 116, 116-26 (2015).

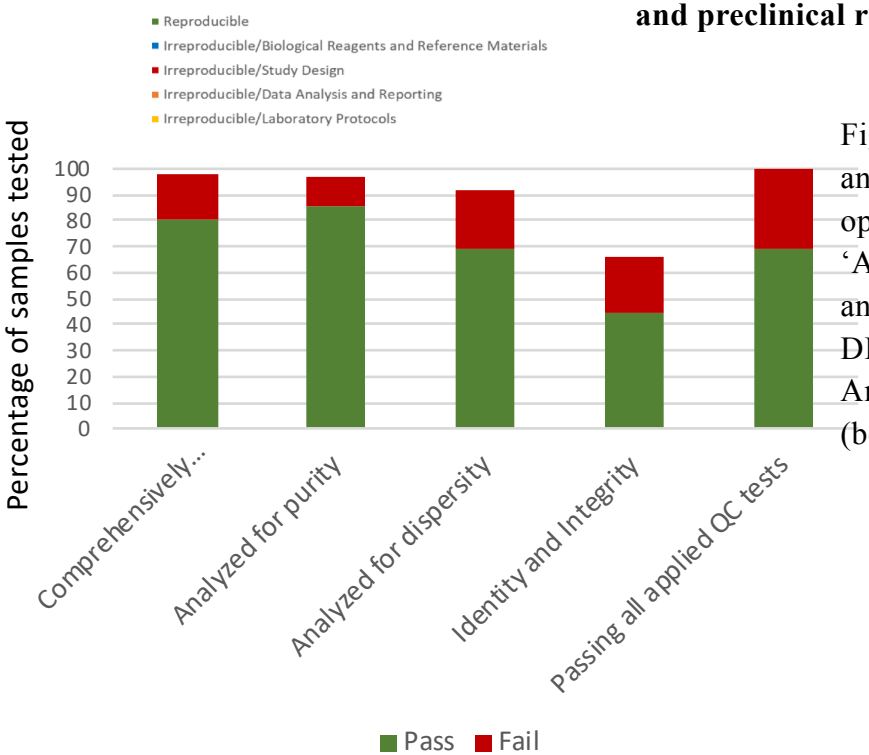


Figure 2. Summary of sample testing and results. 'Comprehensively documented' is an evaluation of the documentation supplied with the protein samples and reflects our opinion on whether this is sufficient to easily reproduce the sample. Samples 'Analysed for Purity' have been evaluated using SDS-PAGE, CE, RPLC or similar analytical techniques. Samples 'Analysed for Dispersity' were evaluated using SEC, DLS, SEC-MALS or Field-Flow Fractionation, Field-Flow Fractionation-MALS or Analytical Ultra-Centrifugation. Identity and Integrity was evaluated using MS (bottom up or top down as appropriate).

- 186 samples from 47 laboratories
- 30% samples failed at least in one QC test

Quality control of purified protein
Best practice recommendations

Guideline

ii) Minimal quality control parameters that should be tested on protein sample

- Purity & integrity
- Homogeneity (aggregation state)
- Identity

iii) Extended quality control parameters

- General quality test by UV spectroscopy
- Homogeneity Conformational stability/folding state
- Optimization of storage conditions
- Batch-to-batch consistency

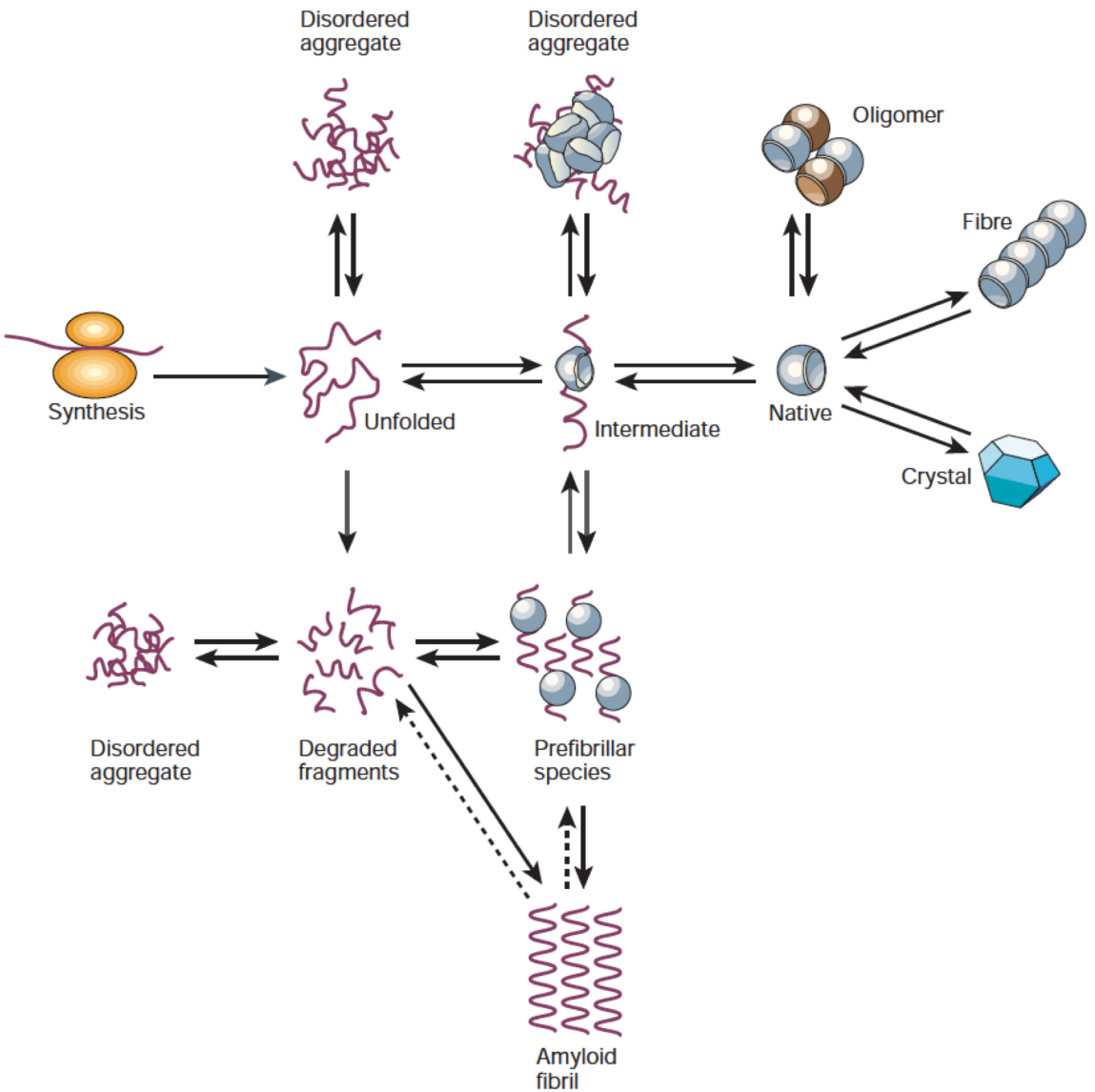
Protein Folding

20^n different possible polypeptide chains of n amino acids long

- It is estimated that the there are between 10^{78} to 10^{82} atoms in the known, observable universe.

Only a very small fraction of this vast set of conceivable polypeptide chains would adopt a single, stable three-dimensional conformation—by some estimates, less than one in a billion.

Protein folding



Christopher M. Dobson. Nature 426, 884-890. 2003

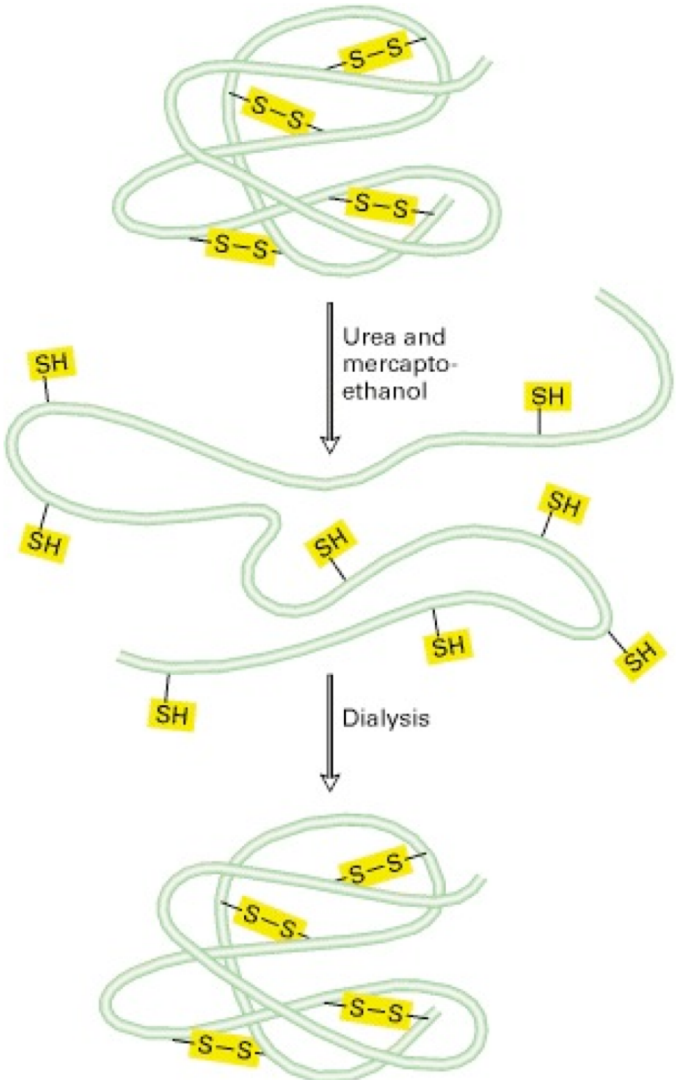
Quiz:

- Who has ever done a refolding experiment?

Quiz:

- Who has ever done a refolding experiment?
- Who has ever done a mini-prep?

In vitro denaturation and renaturation of proteins



Ribonuclease

Christian Anfinsen's experiment

- Treatment with an 8 M urea solution containing mercaptoethanol completely denatures most proteins.
- The urea breaks intramolecular hydrogen and hydrophobic bonds, and the mercaptoethanol reduces each disulfide bridge (–S–S–) to two sulfhydryl (–SH) groups.
- When these chemicals are removed by dialysis, the –SH groups on the unfolded chain oxidize spontaneously to re-form disulfide bridges, and the polypeptide chain simultaneously refolds into its native conformation and activity is reestablished.

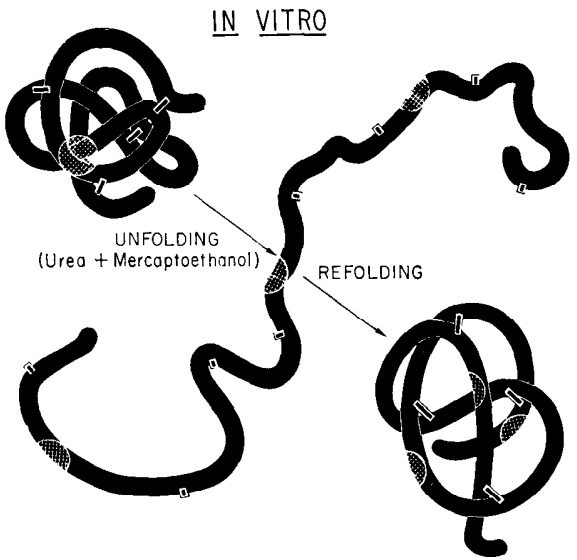
Nobel prize in Chemistry 1972

The thermodynamic hypothesis

”3D structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, etc.) is the one in which the **Gibbs free energy** of the *whole system* is **lowest**”

- the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a *given environment.*
- In terms of natural selection: a protein molecule only makes stable, structural sense when it exists under conditions similar to those for which it was selected “the so-called physiological state”.

The Two-states model



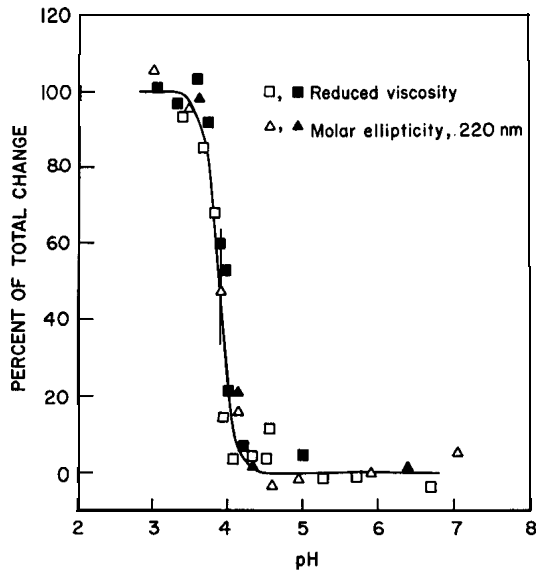
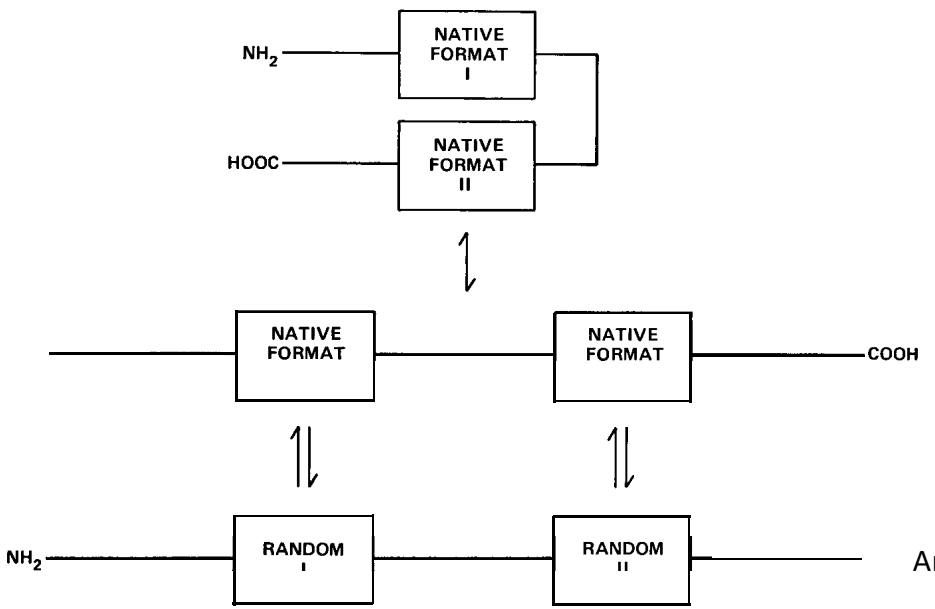
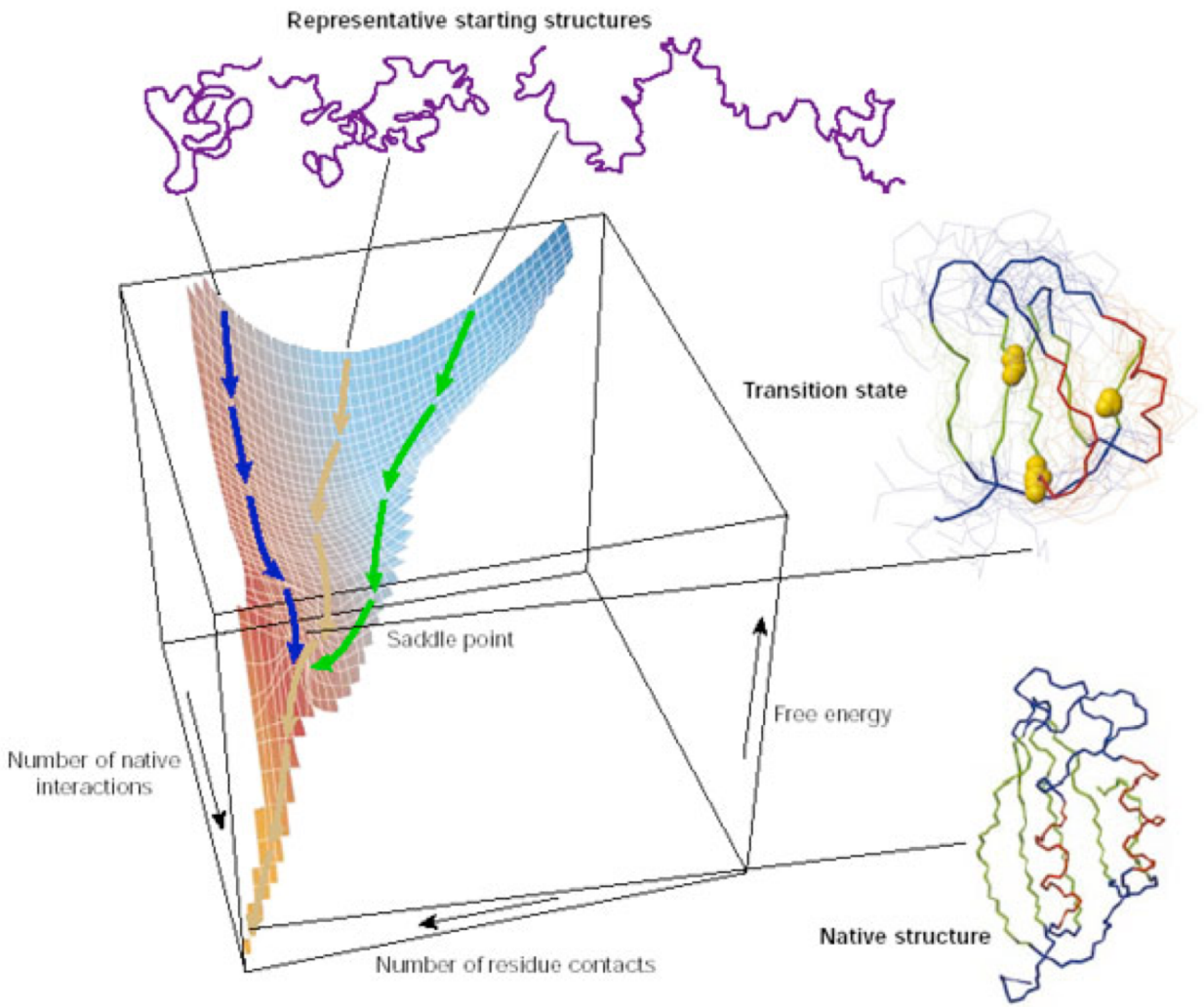


Fig. 9.
Changes in reduced viscocity and molar ellipticity at 220 nm during the acid-induced transition from native to denatured nuclease.
□ and ■ , Reduced viscosity; △ and ▲, molar ellipticity at 220 nm. □ and △, Measurements made during the addition of acid; ■ and ▲, measurements made during the addition of base.
A. N. Schechter, H. F. Epstein and C. B. Anfinsen, unpublished results.



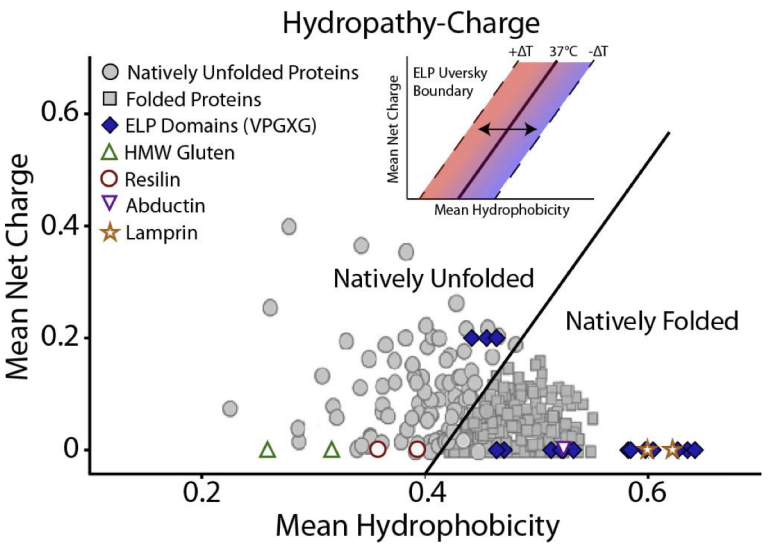
Anfinsen, Nobel Lecture, December 11, 1972

The critical region: the transition state

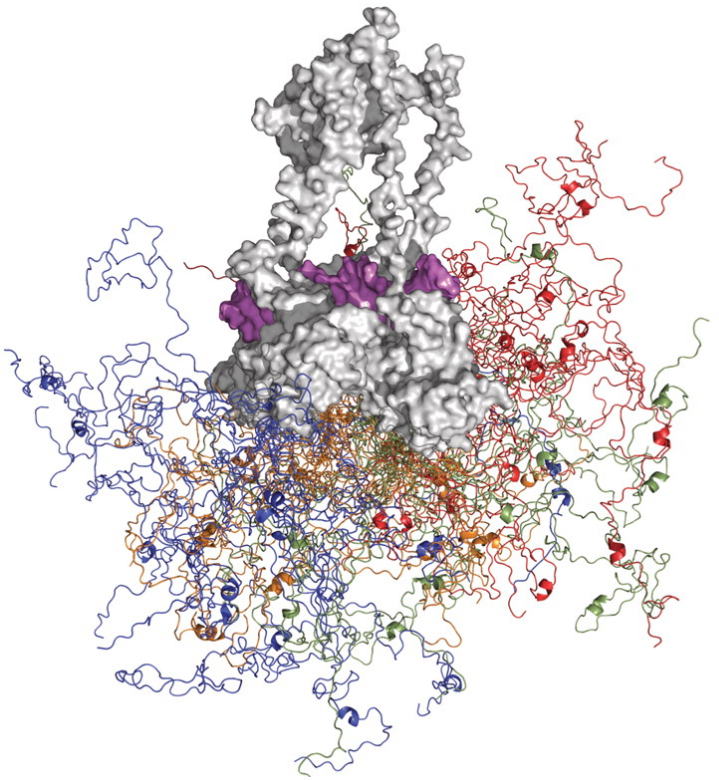


Christopher M. Dobson. Nature 426, 884-890. 2003

Intrinsically disordered proteins



Roberts S1, Dzuricky M2, Chilkoti A1. FEBS Lett. 2015 Sep 14;589(19 Pt A):2477-86.



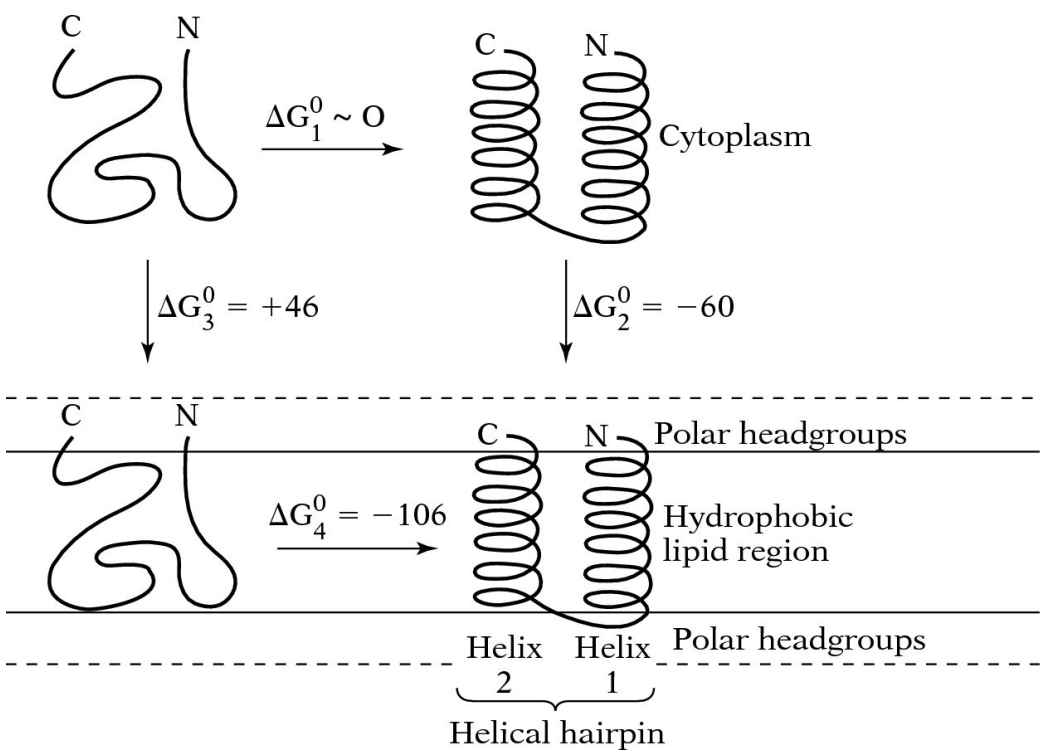
Mark Wells et al. PNAS 2008;105:5762-5767

How to study them?

- Bioinformatics
- SAXS
- CD
- NMR

Protein folding on membrane proteins

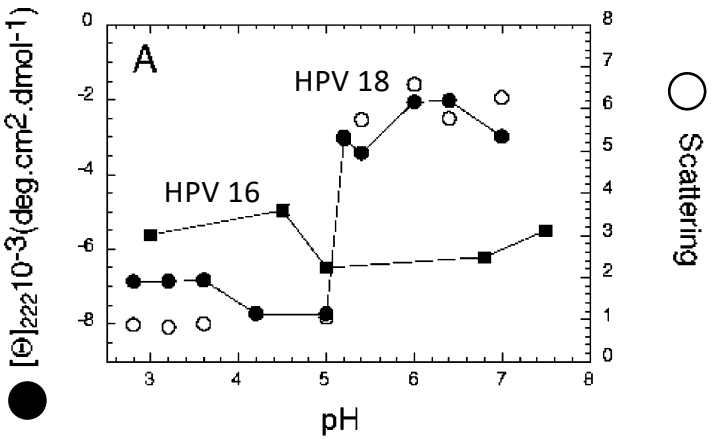
Folding α-helical membrane proteins



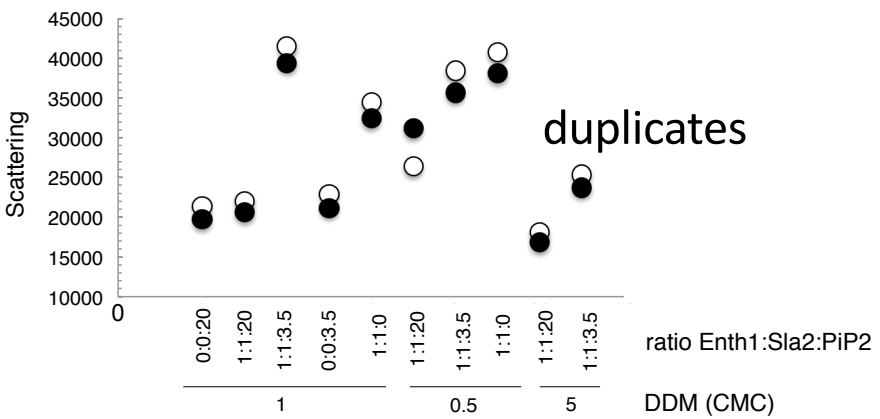
- Helical hairpin hypothesis
- Insertion of a hairpin structure composed of two helices into the nonpolar interior of the bilayer
- Insertion is driven by free energy arising from burying hydrophobic helical surfaces
- Alternative pathway of inserting unfolded peptide/random coil is energetically unfavored

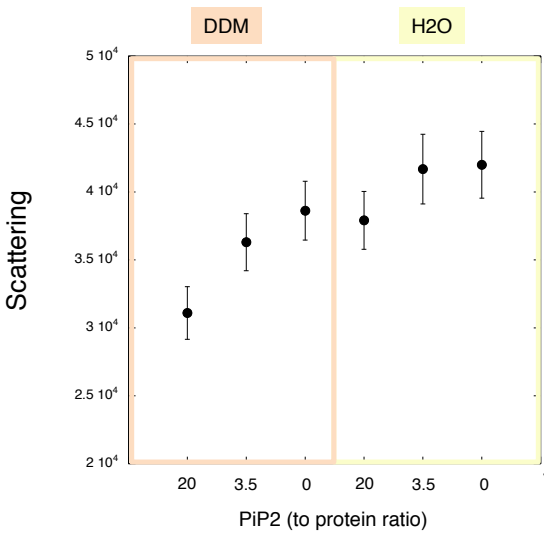
Sampling Protein Folding

- Solubility
- Aggregation
- Secondary structure contain
- Stability



light scattering at 360 nm (Fluorimeter)

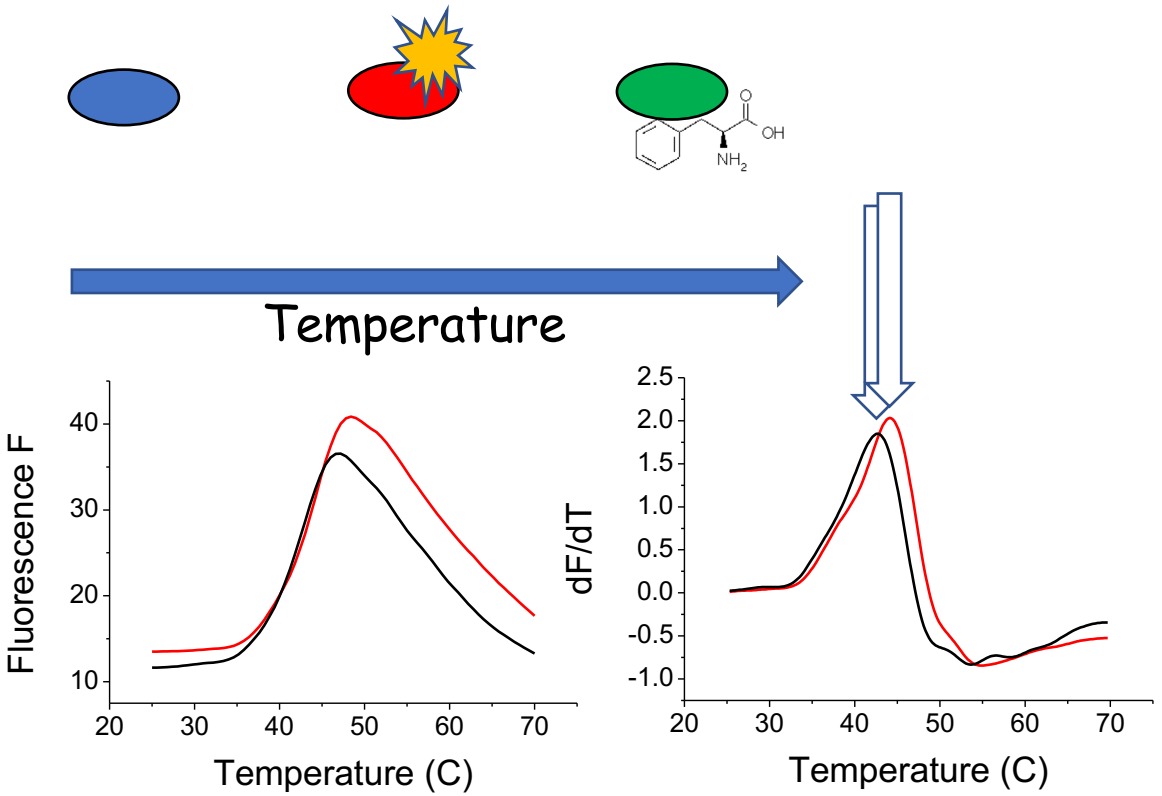




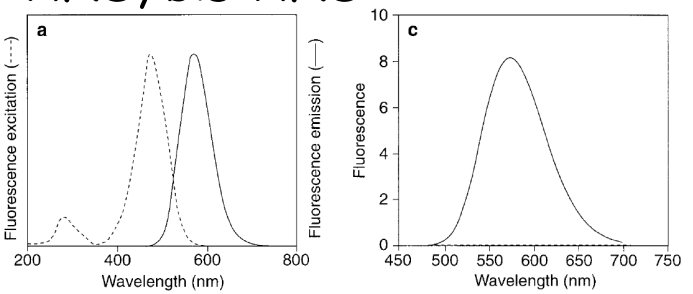
Sampling Protein Folding

- Solubility
- Aggregation
- Secondary structure contain
- Stability
    - Screen for protein Stability

Differential Scanning Fluorimetry



ANS, bis-ANS

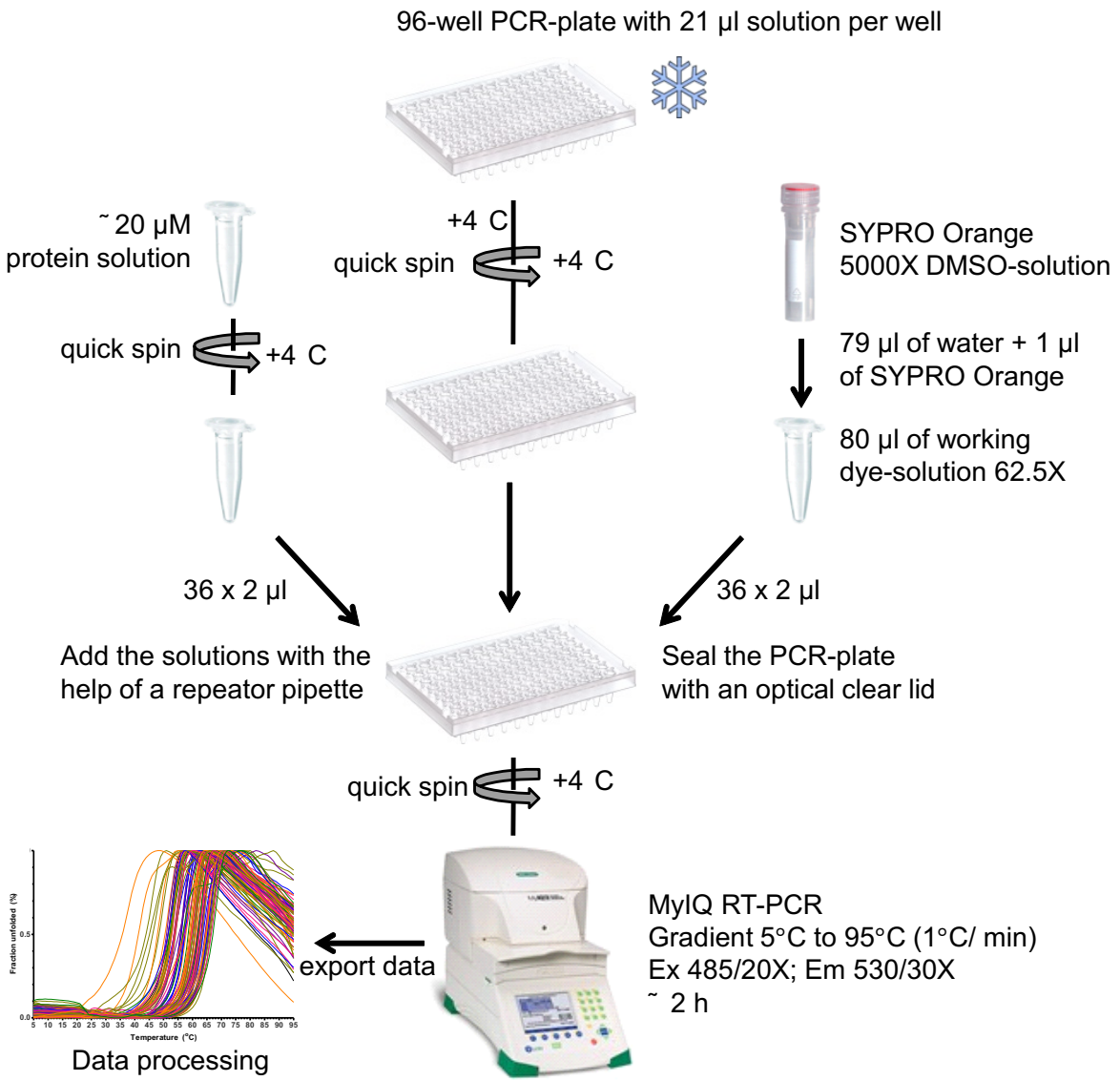


SYPRO Orange, Mol. Probes,
Steinberg et al Anal. Biochem. 1996

Pantoliano et al J Biomol Screen. 2001

48- 96 samples

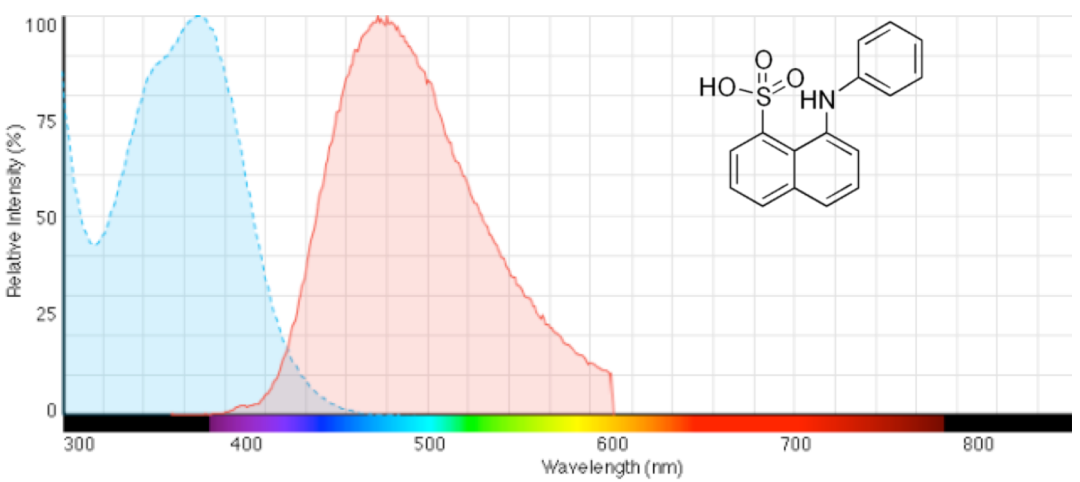
Thermofluor



Thermofluor

Sypro Orange/ ANS fluorescent properties will change as it binds to hydrophobic regions on the protein surface

8-anilino, 1-naphthalene sulfonate



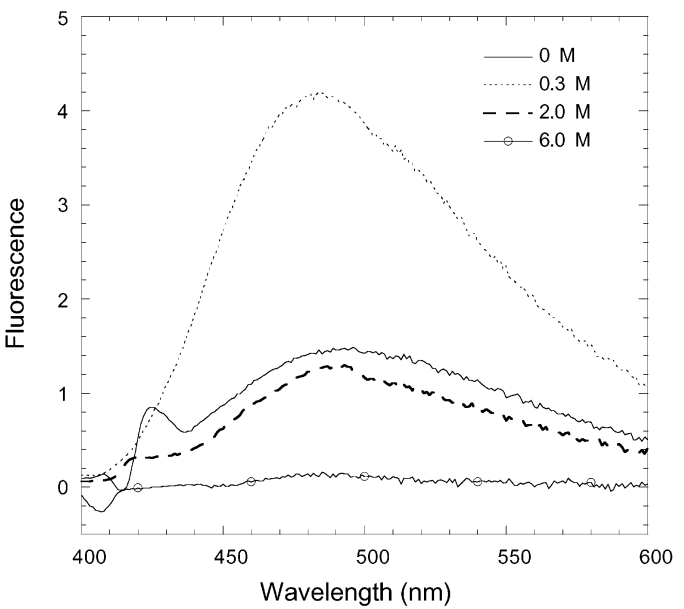
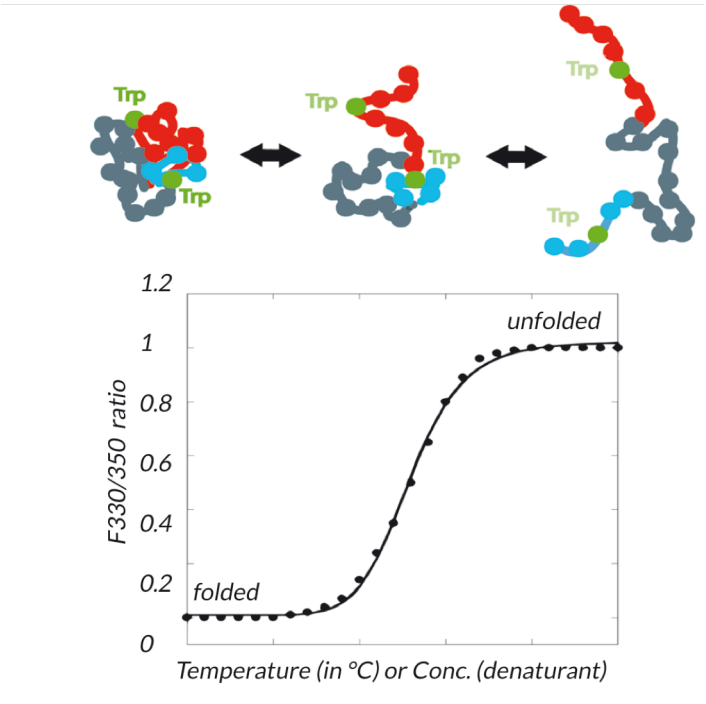
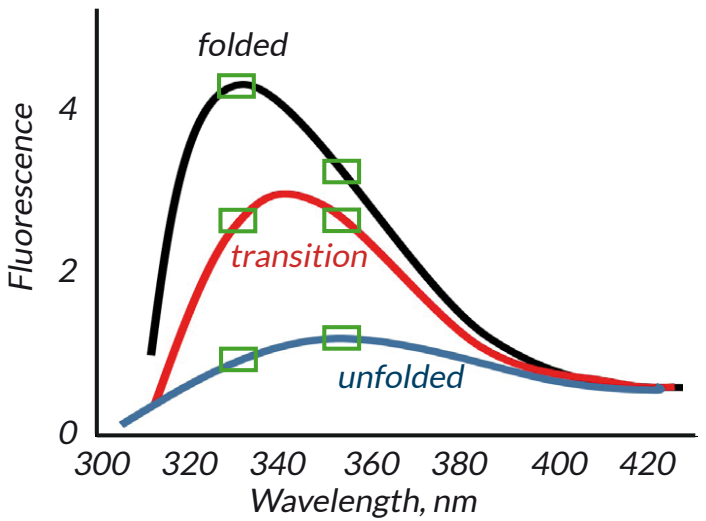


FIGURE 7: ANS binding of E7 after the GdmCl-induced conformational transition at the diferent denaturant concentrations indicated.

Not compatible with detergents!!!

nanoDSF

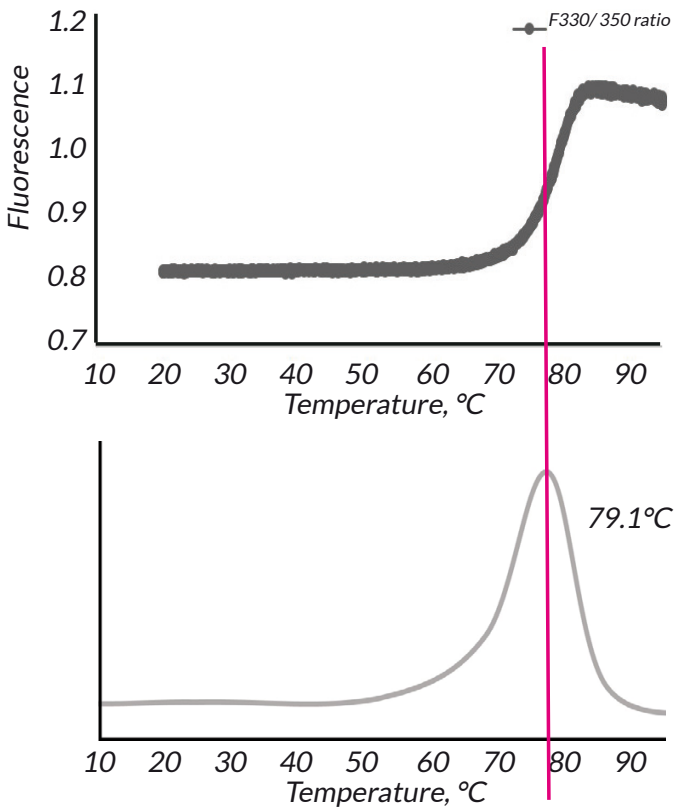


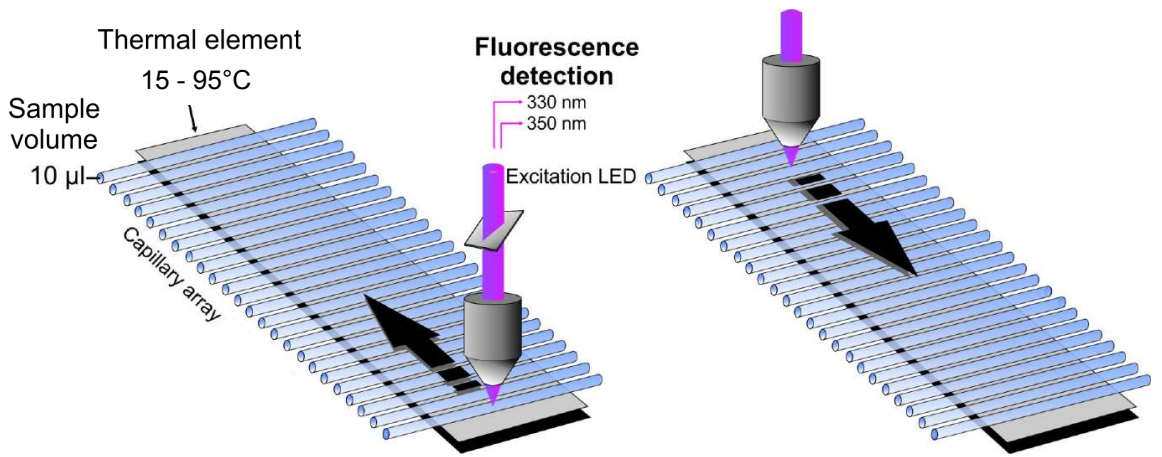


NanoTemper Technologies Prometheus NT.48

Principle behind the nanoDSF. Increasing temperature causes protein unfolding that can be assessed by monitoring changes of tryptophan fluorenscence at 330nm and 350nm wavelength.

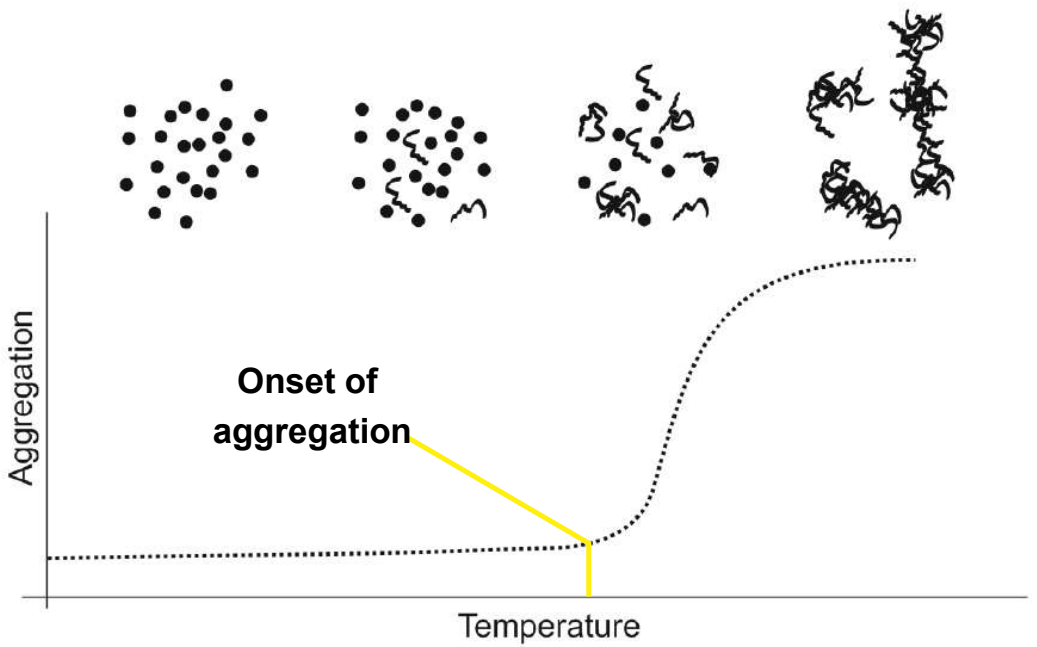
The thermal unfolding transition midpoint (Tm)

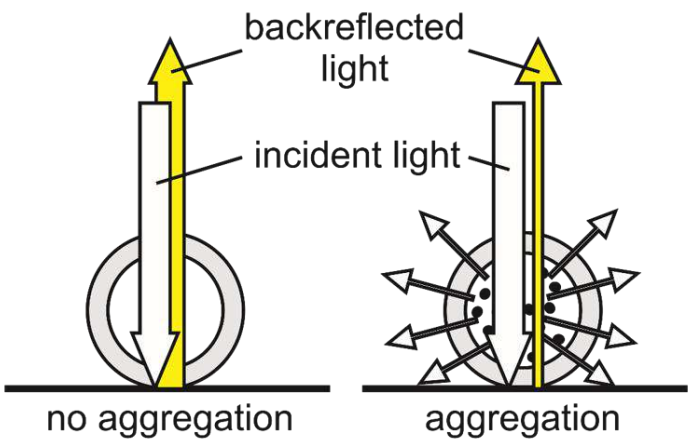




- ratio F330/350 against the temperature
- Tm is determined by first derivate analysis

Static light scattering





High-throughput screening for IMPs stability

40% samples processed in the SPC are membrane proteins

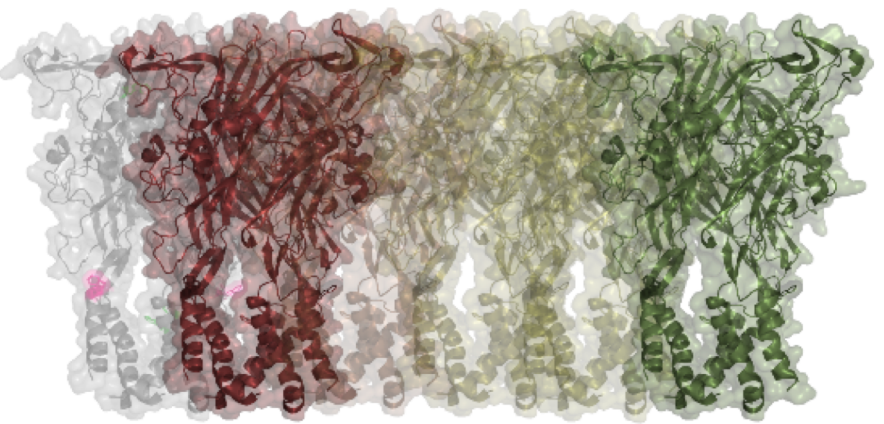
Introduction

- IMP stability in detergent or membrane-like environments is the bottleneck for structural studies
- Detergent solubilization from membranes is usually the first step in the workflow
- Looking for a simple high-throughput screening method to identify optimal conditions for membrane protein stabilization

High-throughput screening for IMPs stability

- following **nDSF and scattering** upon thermal denaturation
- (de-)stabilization **effects of detergents**
- find **suitable conditions** for downstream handling during purification
- **thermodynamic parameters (Tm, Tagg, Tonset)**
- We selected 9 IMPs to benchmark our protocol

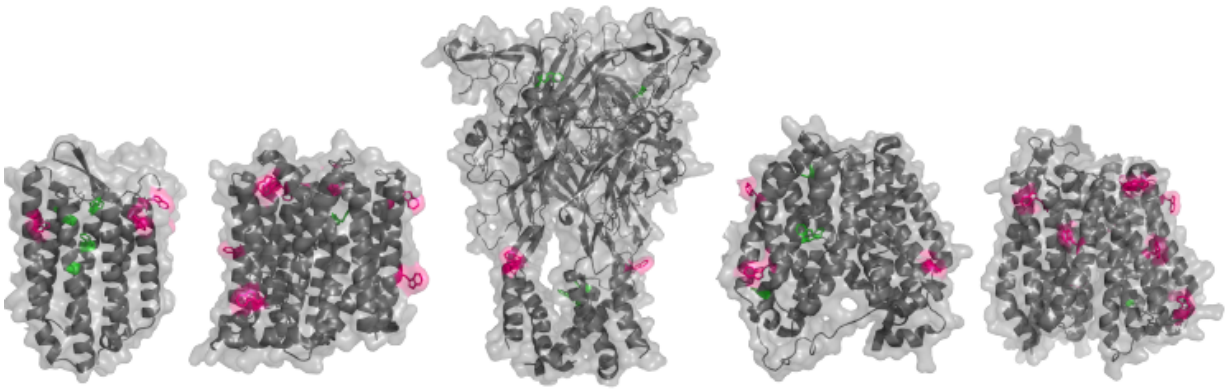
Objective



Unstable Stable

Kotov *et al.* Scientific Reports 2019

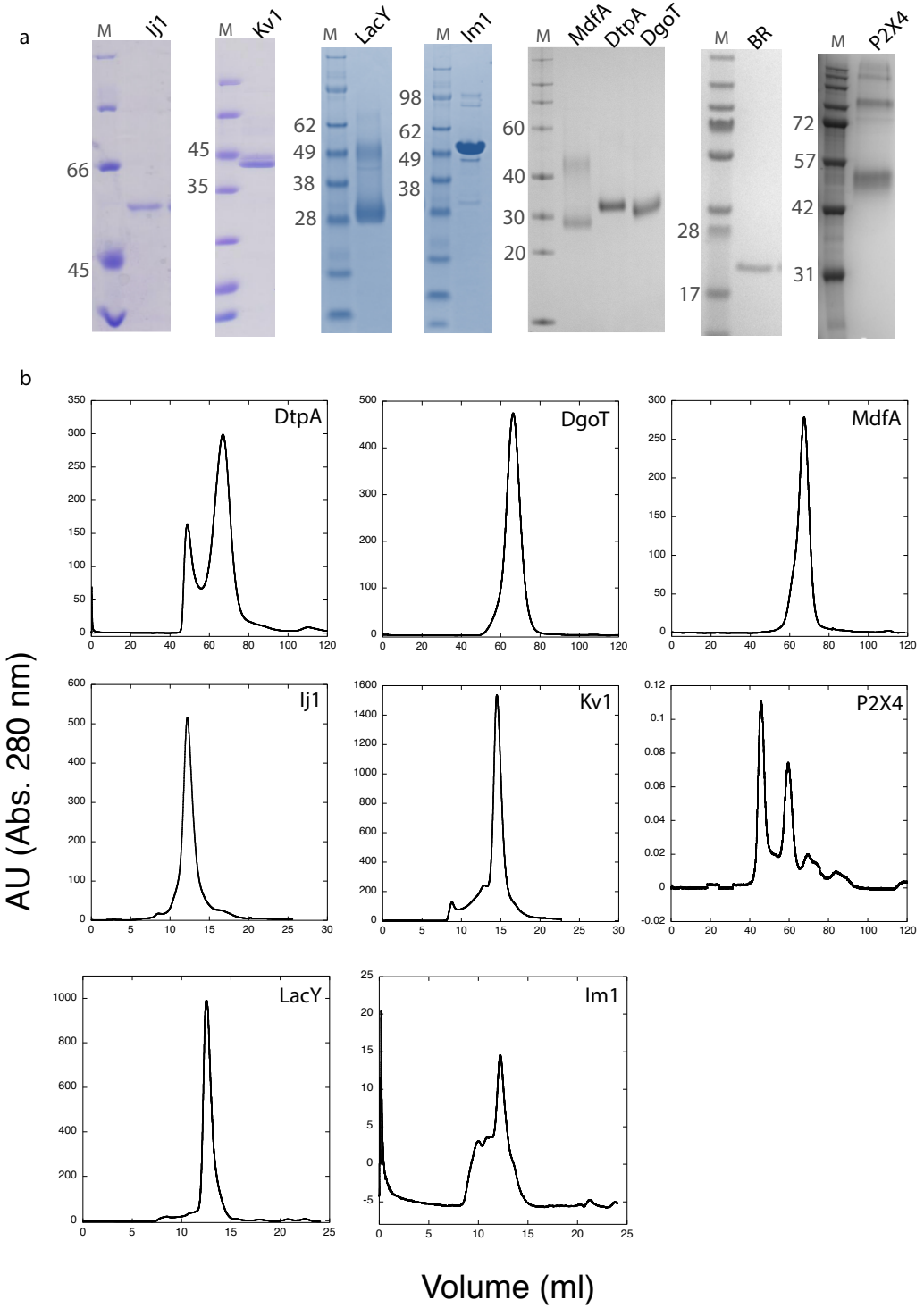
We selected 9 Integral membrane proteins (targets)



BR	MdfA	P2X4	LacY	DtpA	?	?	?	?
					MFS transporter	7TM pathogen	ABC transporter	kinase
					DgoT	**Kv1**	**Ij1**	**Im1**

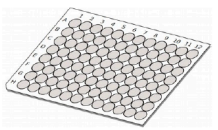
Protein	Organism	Family	Function	Number of Trp residues	PDB ID
DgoT	*E. coli*	MFS transporters	putative galactonate transporter	14	6E9N, 6E9O
MdfA	*E. coli*	MFS transporter	multi drug resistance	9	4ZP0, 4ZOW, 4ZP2, 6GV1, 6EUQ
DtpA	*E. coli*	MFS transporter	peptide transporter	10	6GS1, 6GS4, 6GS7
Kv1	*Pseudomonas aeruginosa*	unknown	unknown	17	—
Ij1	*E. coli*	ABC-Transporter	ion transport	22	—
P2X4	*Homo sapiens*	P2X ionotropic receptors	regulator in mediating neuropathic pain	6	4DW0, 4DW1 (zebrafish)
BR	*Halobacterium salinarum*	7TM receptor	proton pump	8	4MD1, 4MD2, 4XXJ
LacY	*E. coli*	MFS transporter	transport of beta-galactosides	5	1PV6
Im1	*E. coli*	HisKA	Kinase	2	—

- Membranes solubilized in 1–2% DDM
- DDM as starting detergent in SEC



Our pipeline

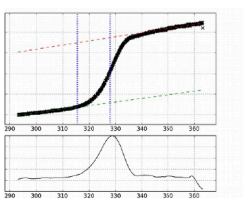
Purification with “best guess” detergent (DDM)



Dilution of initial detergent and addition of excess screen detergent

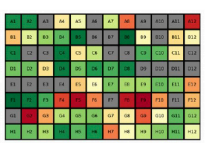
Equilibrate 1 hour for detergent exchange

Simultaneous measurement of protein unfolding and aggregation



Non-linear curve-fitting to estimate T_m and T_{agg} and onset temperatures

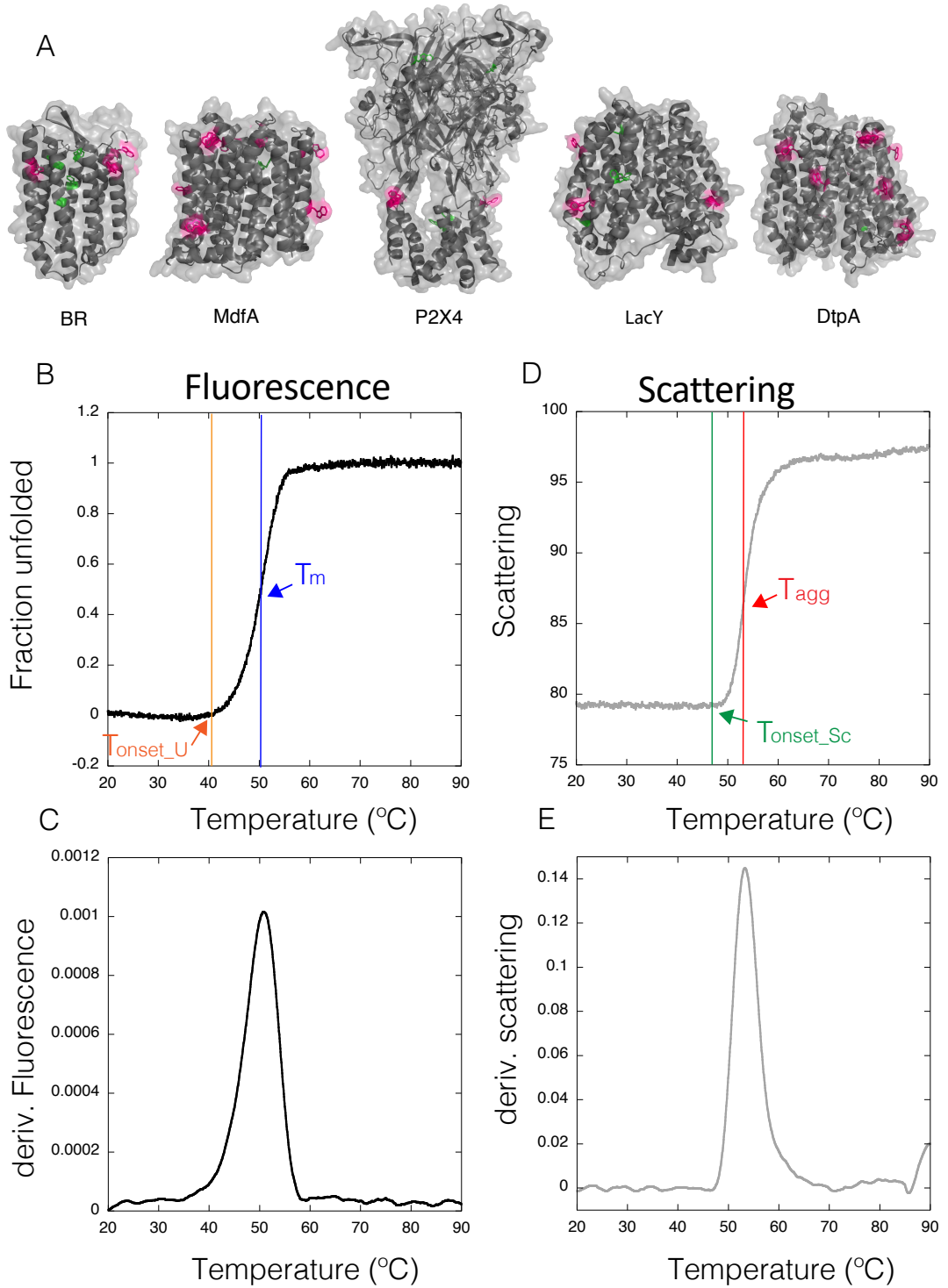
Removal of failed samples based on fit parameter errors



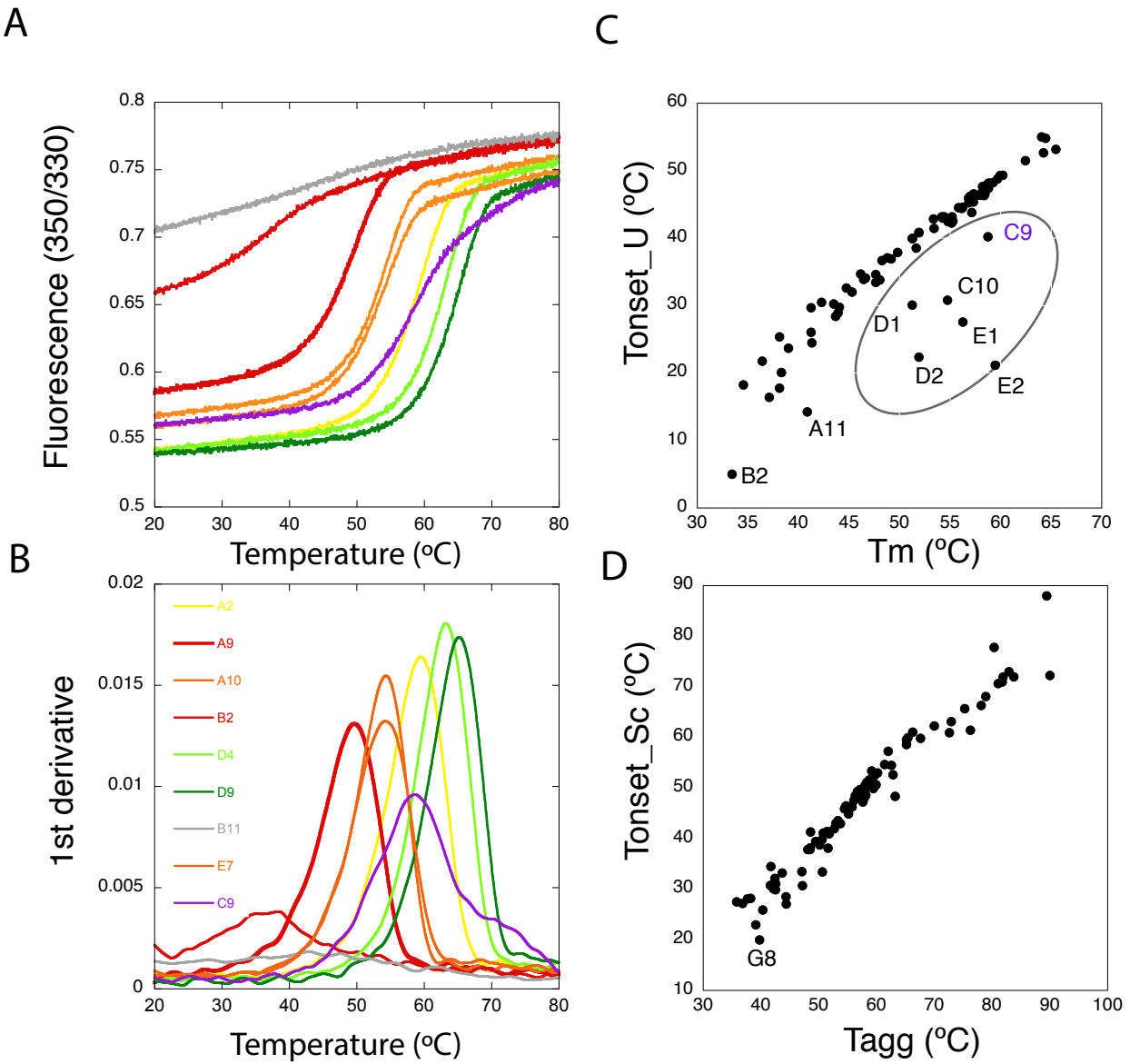
Top hits in thermal stability and solubility

Kotov *et al.* Scientific Reports 2019

nDSF measurements



Tm vs Tonset


A
Fluorescence (350/330)
Temperature (°C)
B
1st derivative
Temperature (°C)
A2
A9
A10
B2
D4
D9
B11
E7
C9
C
Tonset_U (°C)
Tm (°C)
C9
C10
D1
E1
D2
E2
A11
B2
D
Tonset_Sc (°C)
Tagg (°C)
G8

Correlation between IMP stability and micelle size

Sample	Readout	n	Spearman's ρ
DtpA	Ratio	21	0.70
	Scattering	27	0.24
DgoT	Ratio	15	0.43
	Scattering	33	0.25
LacY	F330	8	0.00
	Scattering	27	-0.04
Kv1	Ratio	38	0.46
	Scattering	37	0.66
Ij1	Ratio	8	0.45
	Scattering	29	0.65
Im1	Ratio	9	0.78
MdfA	Ratio	33	0.74
	Scattering	36	0.20
P2X4	Ratio	42	0.80
BR	F330	28	0.84
	Scattering	28	0.73

Coefficients approaching zero show no correlation between variables while those approaching 1 indicate a positive correlation (Y values increase as the X values increase).

Correlation between IMP stability and micelle size

Sample	Readout	n	Spearman's ρ
DtpA	Ratio	21	0.70
	Scattering	27	0.24
DgoT	Ratio	15	0.43
	Scattering	33	0.25
LacY	F330	8	0.00
Im1	Ratio	9	0.78
MdfA	Ratio	33	0.74
	Scattering	36	0.20
P2X4	Ratio	42	0.80
BR	F330	28	0.84
	Scattering	28	0.73

Regarding crystallization, shorter chain detergents are preferred as they allow for better crystal packing and better diffracting crystals. The goal is to find the shortest possible detergent that does not cause the protein to unfold!

Coefficients approaching zero show no correlation between variables while those approaching 1 indicate a positive correlation (Y values increase as the X values increase).

Discussion

Implications for sample optimization

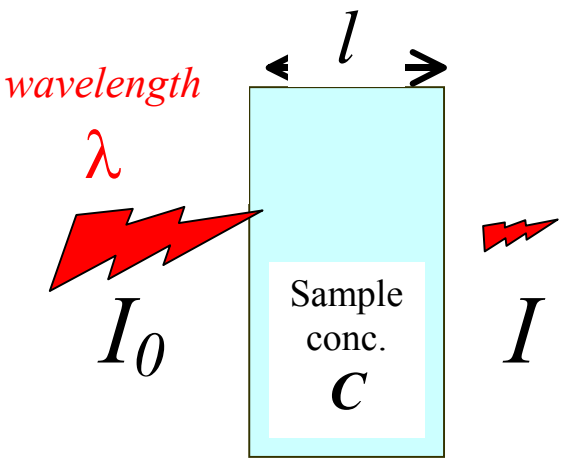
- Analyse stability and solubility of IMPs by diluting them from their initial solubilization condition into different detergents
- Identify groups of detergents with characteristic stabilization and destabilization effects for selected targets
- Fos-choline and PEG family detergents may lead to membrane protein destabilization and unfolding
- Finding conditions that are suitable for downstream handling of membrane proteins during purification

Absorption Spectroscopy

- Shine light through a sample and measure the proportion absorbed as a function of wavelength.
- Absorbance $A = \log(I_0/I)$
- Beer-Lambert law:

  $A(\lambda) = \varepsilon(\lambda)lc$

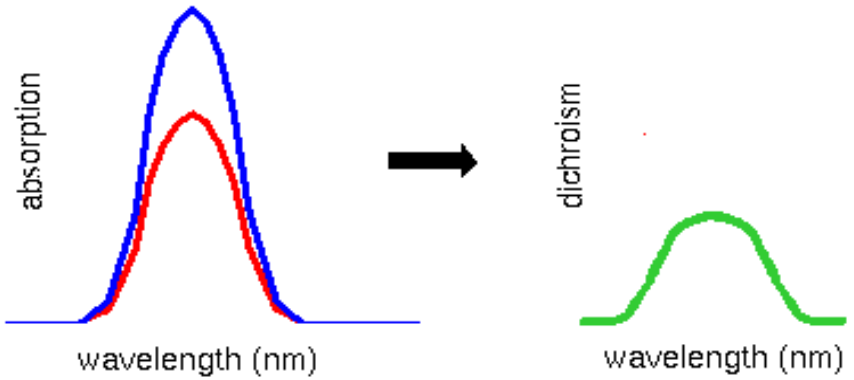
  ε: extinction coefficient

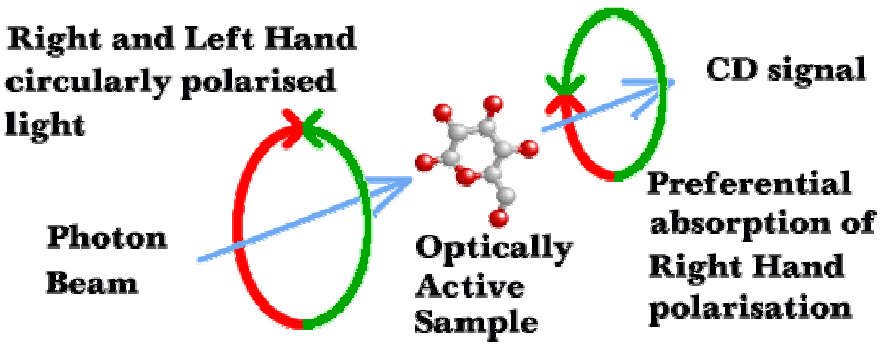


- The longer the path or the more concentrated the sample, the higher the absorbance

Circular Dichroism

- CD measures the difference between the absorption of left and right handed circularly-polarized light. polarized light:

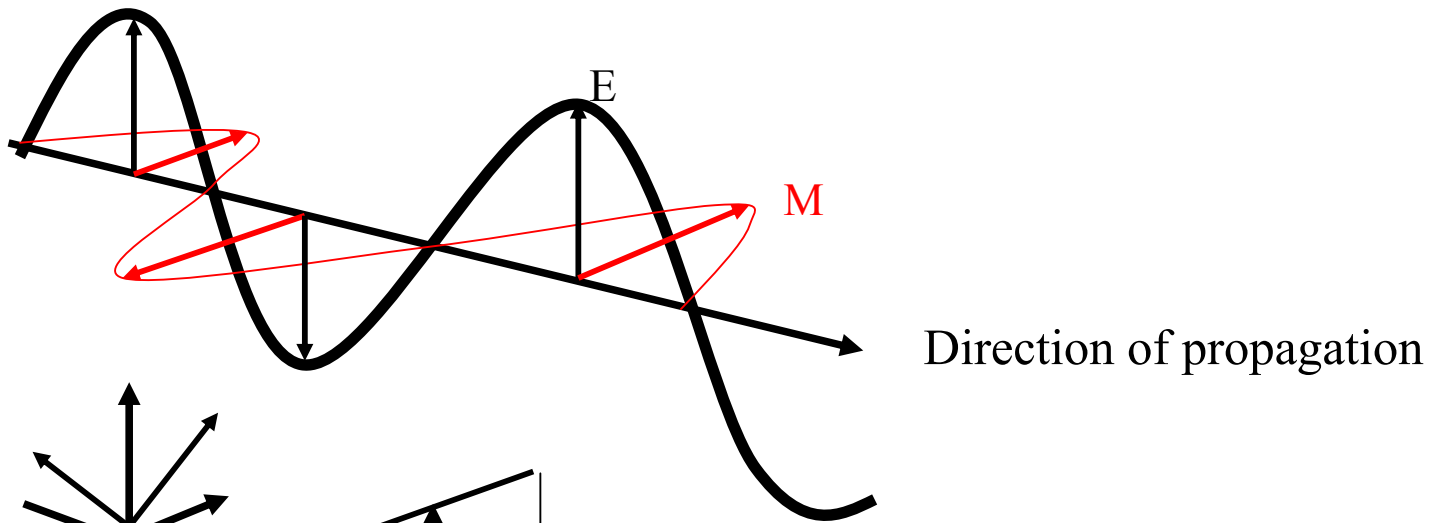


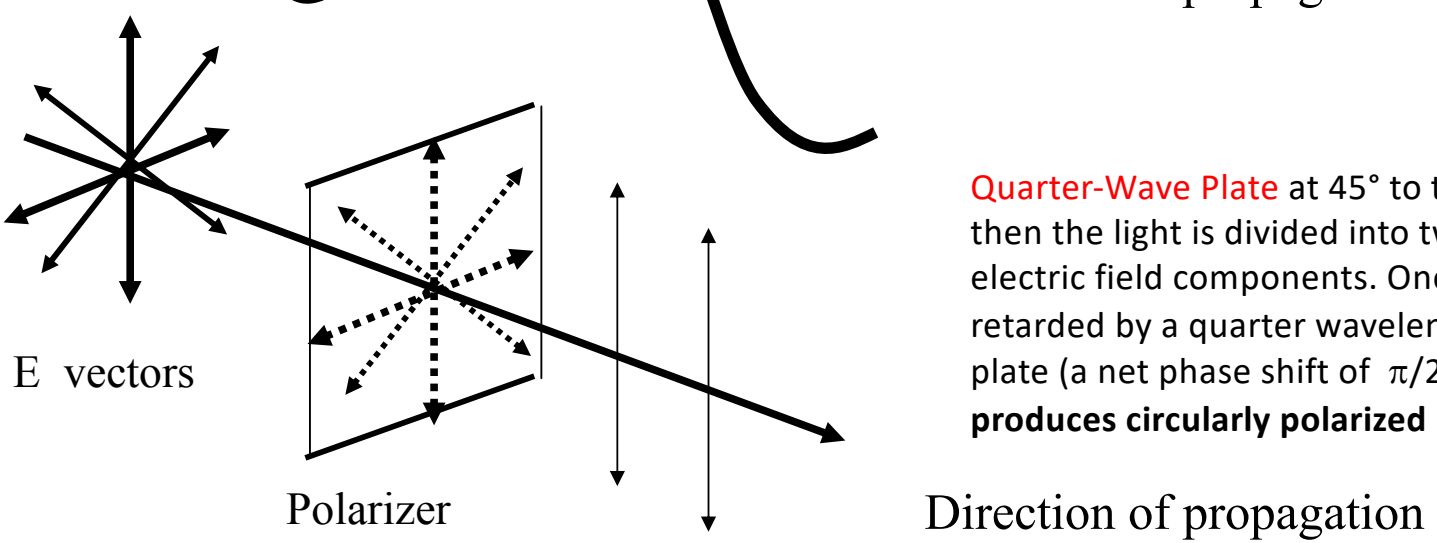


- This is measured as a function of wavelength, & the difference is always very small (<<1/10000 of total). After passing through the sample, the L & R beams have different amplitudes & the combination of the two unequal beams gives elliptically polarized light.Hence, CD measures the ellipicity of the transmitted light (the light that remains that is not absorbed):

Circular Dichroism

Plane Polarized Light



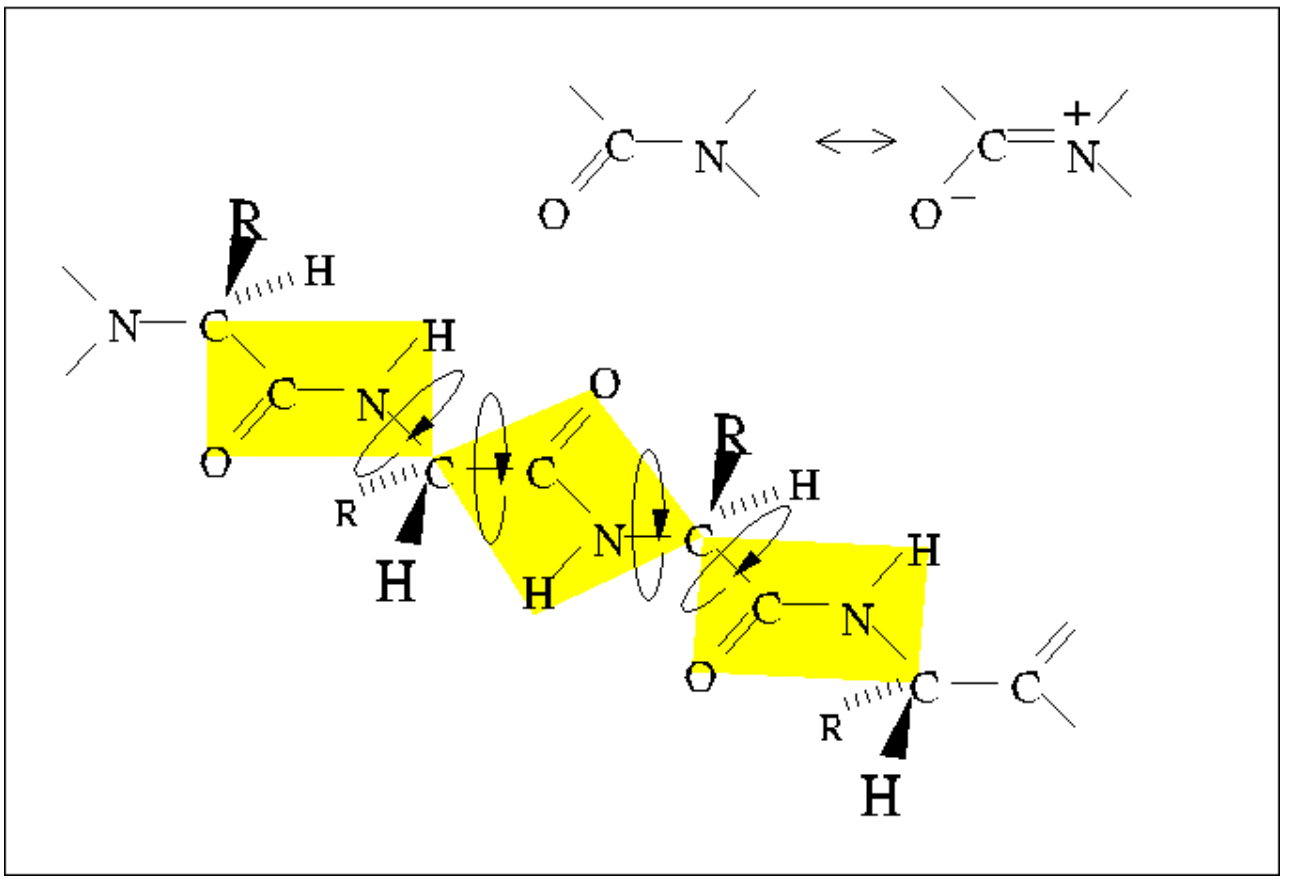


Quarter-Wave Plate at 45° to the optic axis, then the light is divided into two equal electric field components. One of these is retarded by a quarter wavelength by the plate (a net phase shift of $\pi/2$). This **produces circularly polarized light.**

Circular Dichroism

•The peptide bond is inherently asymmetric & is always optically active.

• Any optical activity from side-chain chromophores is induced & results from interactions with asymmetrical neighbouring groups.



Units

- Molar Ellipticity

$$[\theta] = \frac{\theta}{c.l} \quad \text{deg } cm^2 dmol^{-1}$$

$$[\theta] = \frac{\theta.M}{10.C.l} \quad \text{deg } cm^2 dmol^{-1}$$

$$[\theta] = 3300\,(\varepsilon_L - \varepsilon_R) = 3300\,\frac{(A_L - A_R)}{Cl}$$

- Differential absorbance

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R = \frac{[\theta]}{3300}\ M^{-1} cm^{-1}$$

- [θ] and Δε can also be expressed per residue and this is useful for comparing systems of differing size

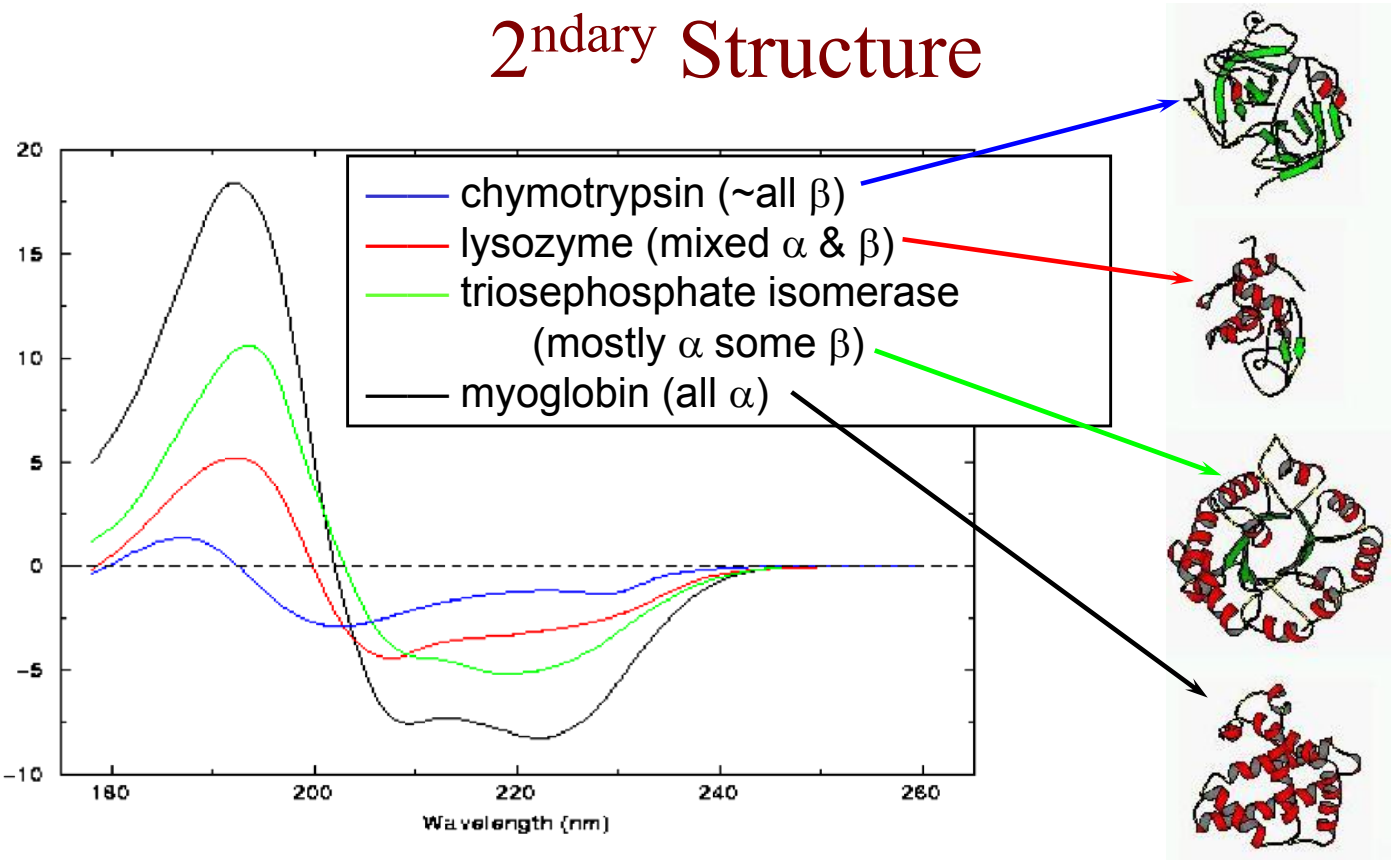
- θ; signal millidegree, c; [sample] dmol L^{-1}, C; [sample] mg mL^{-1}, l; pathlength cm; M; molecular weight

[θ] MRW = θ/(10 x cr x l)

Cr (mean residue molar concentration)= n x c, where n is the number of peptide bonds in the protein

Far UV CD Spectra

Total Signal for a Protein Depends on its 2ndary Structure

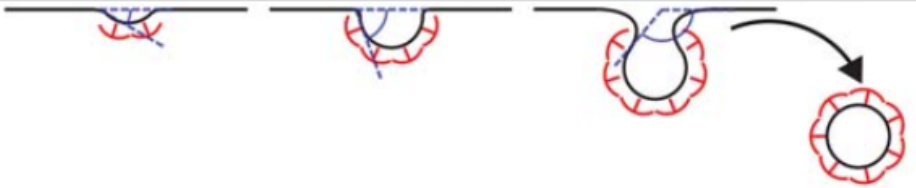


- Notice the progressive change in θ_{222} as the amount of helix increases from chymotrypsin to myoglobin

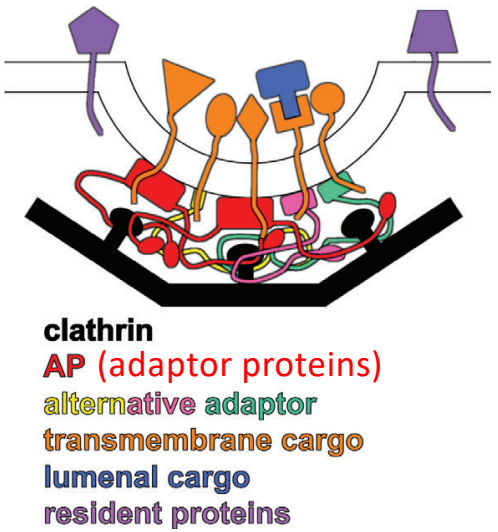
A real example

Assembling a puzzle…

Clathrin mediated endocytosis



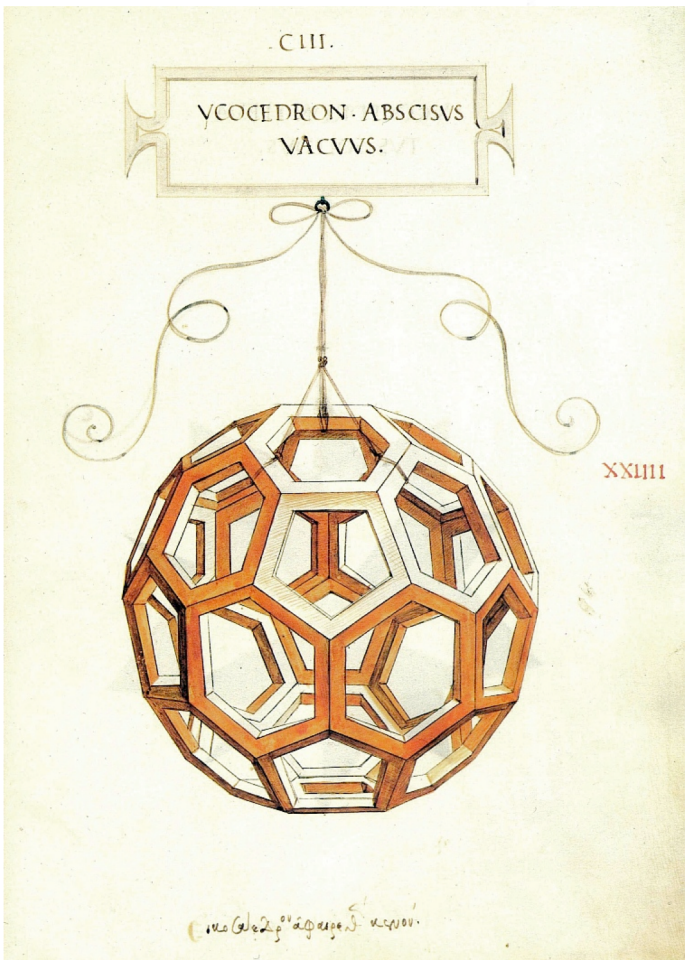
A



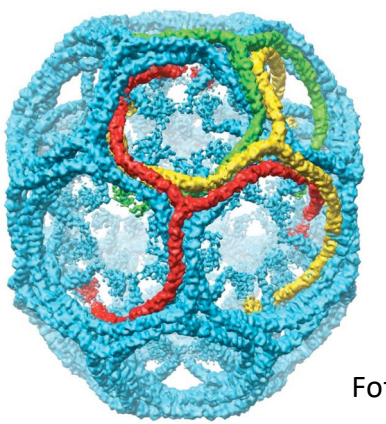
Robinson MS, Review, Traffic 2015

B

-Model of a Clathrin coated vesicle based on EM and proteomics analysis

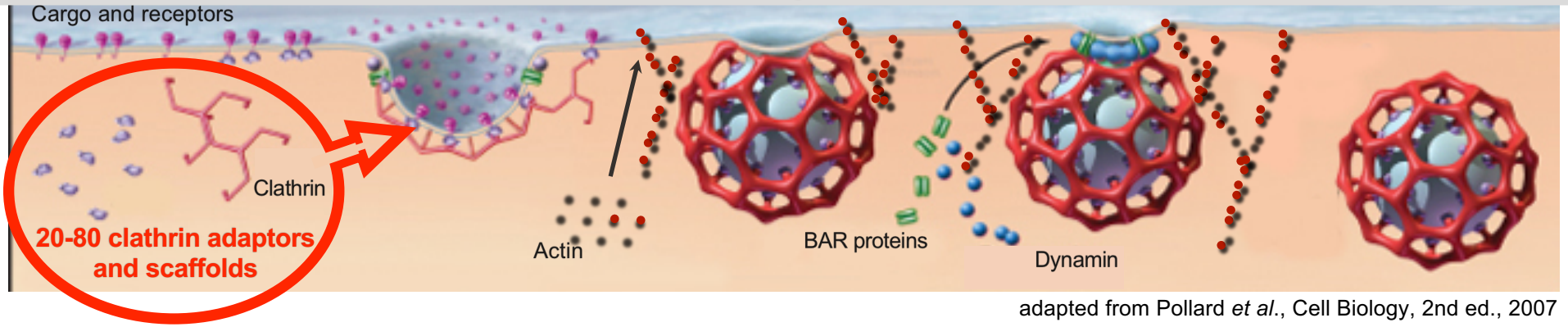


Icosahedron by Leonardo da Vinci (1452-1519)

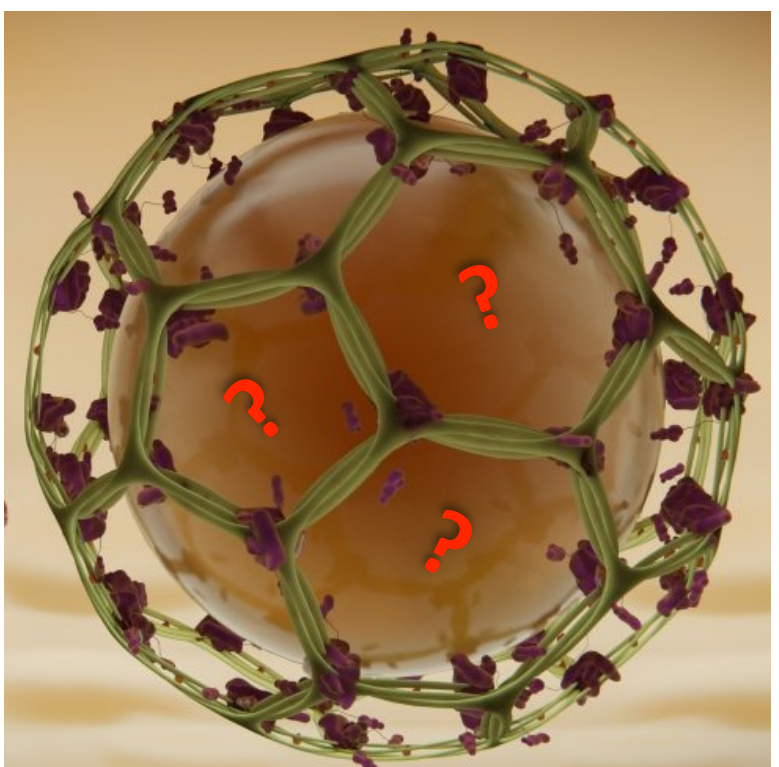


Fotin, et al. Nature. 2004

Functional organization of the central endocytic coat is not known



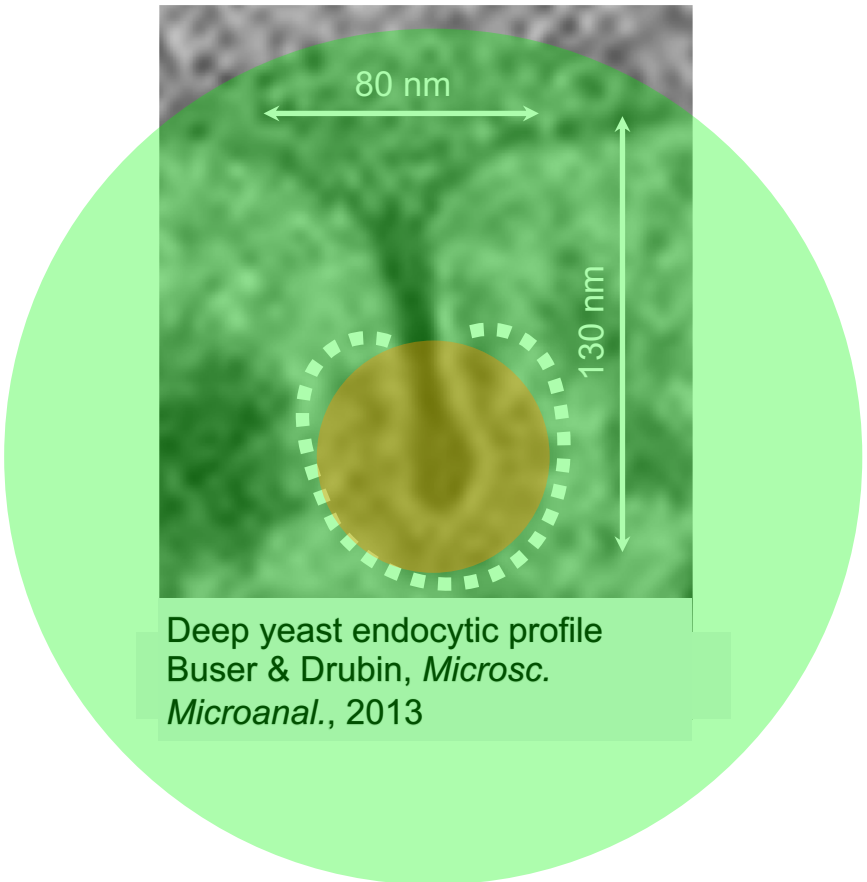
adapted from Pollard *et al.*, Cell Biology, 2nd ed., 2007



Clathrin-mediated endocytosis poster on Amazon
(by Stocktrek Images)

model of clathrin-coated vesicle by Prof. Margaret Robinson,
University of Cambridge

Mapping of the endocytic coat requires a high spatiotemporal resolution



Deep yeast endocytic profile
Buser & Drubin, *Microsc. Microanal.*, 2013

Methods, we currently use:

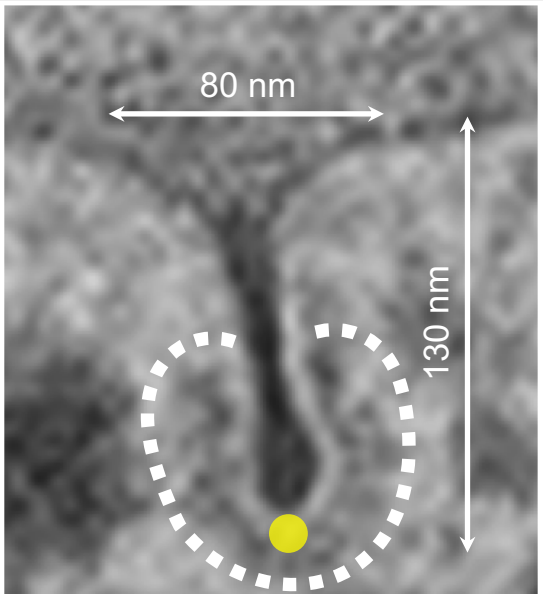
immunoelectron microscopy
Idrissi *et al*, 2008, 2012

correlative light-electron microscopy
Kukulski *et al*, 2012
Avinoam *et al*, 2015
Sochacki *et al*, 2017

live-cell imaging of centroid positions
Picco *et al*, 2015, 2018

single molecule localization microscopy
Mund *et al*, 2018

Mapping of the endocytic coat requires a high spatiotemporal resolution



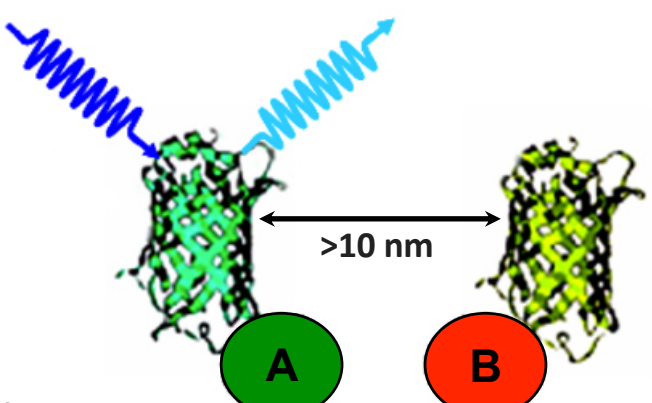
Deep yeast endocytic profile
Buser & Drubin, *Microsc. Microanal.*, 2013

Method, we can use:

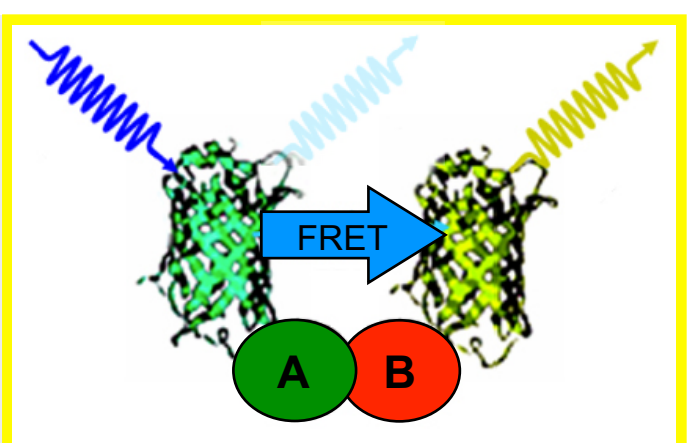
FRET-based protein-protein proximity mapping

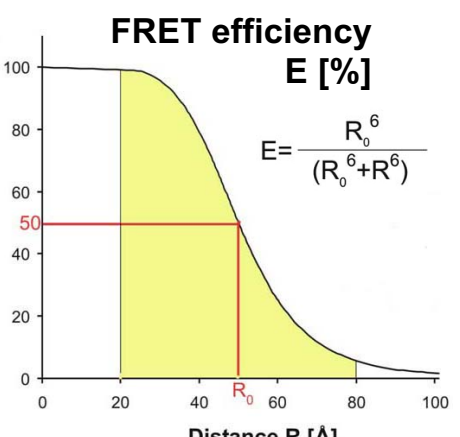
Förster resonance energy transfer (FRET) occurs between fluorophores separated by less then ~ 10 nm

approx. 20 coat-associated proteins (each in dozens of copies in semi-equimolar ratio) localize in a "FRET accessible" area with a very good signal/noise ratio!

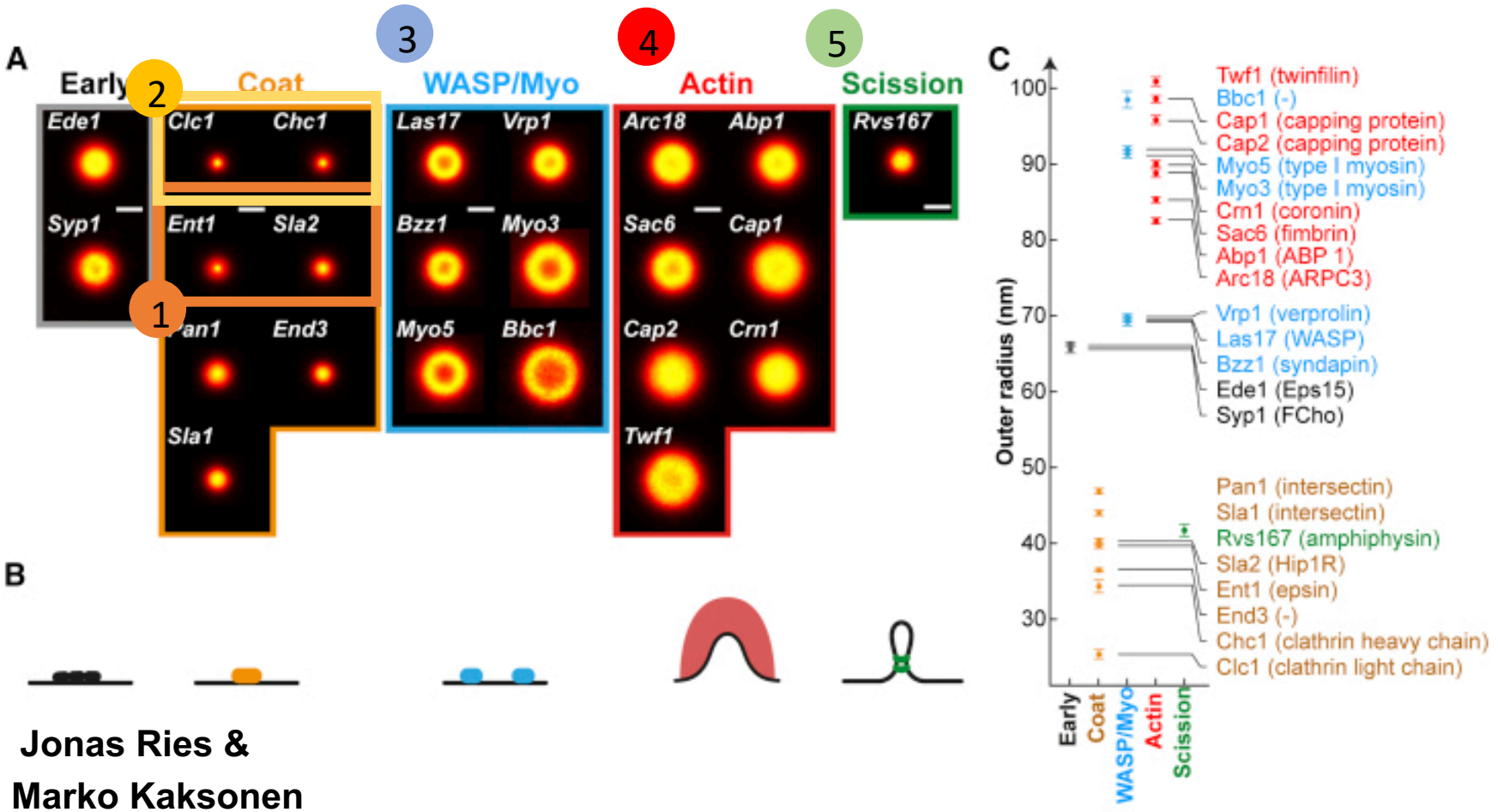


from Bhat *et al*, 2006





Superresolution microscopy on the endocytic site



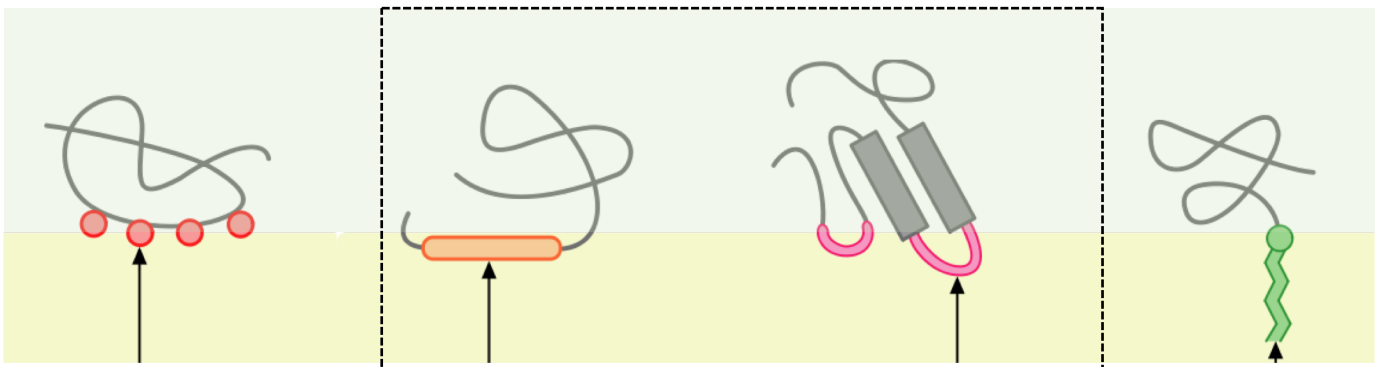
Jonas Ries &
Marko Kaksonen

Mund M. et al. Cell. 2018 Aug 9;174(4):884-896.e17

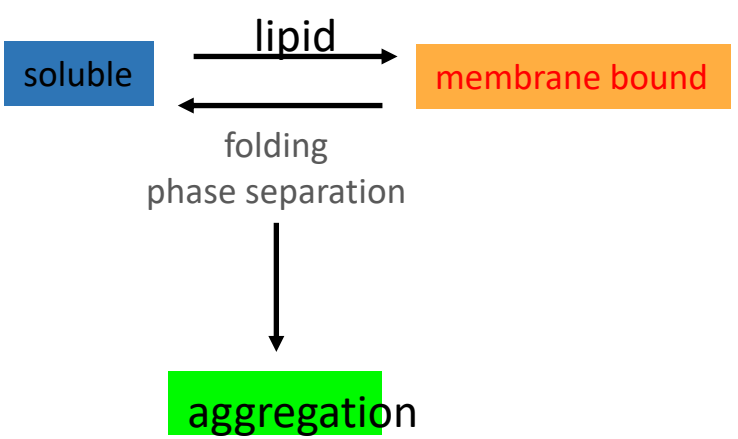
ENTH and ANTH are membrane targeting domains

(Lipid Clamps) PIP2

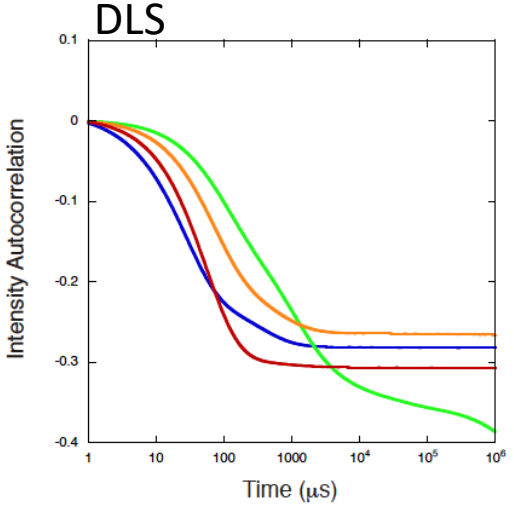
amphitropic domains

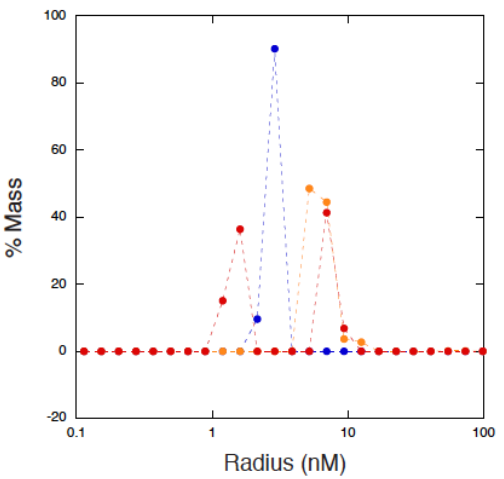


electrostatic interactions | Inducible helix-0 | hydrophobic loops | covalently bound lipid



DLS

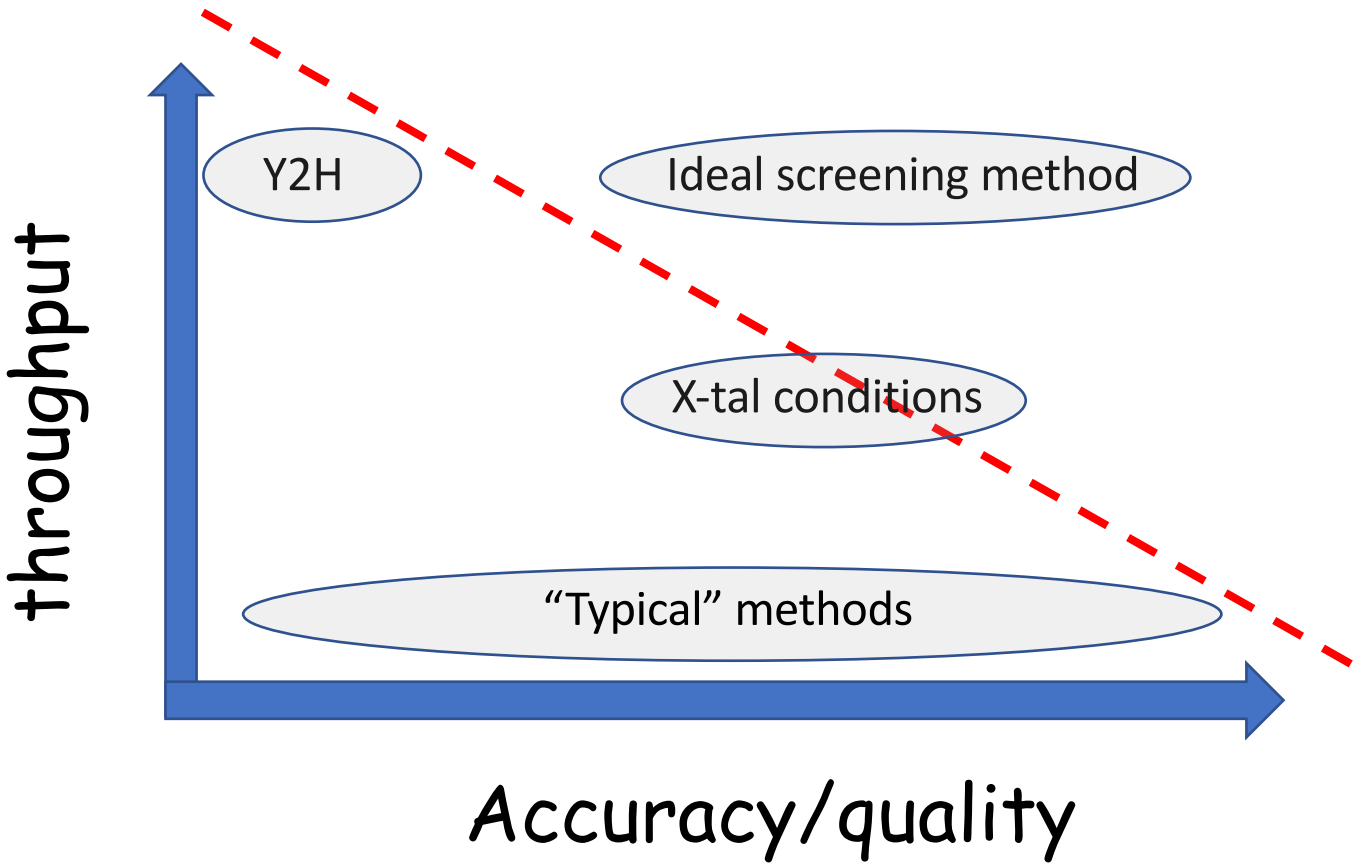




Protein-Protein interaction

Do these proteins form a complex?

Screening



Small molecules
Binding
X-tal conditions
Stability
Grow conditions
Inhibitors
Expression
Etc etc etc

Screening for binding

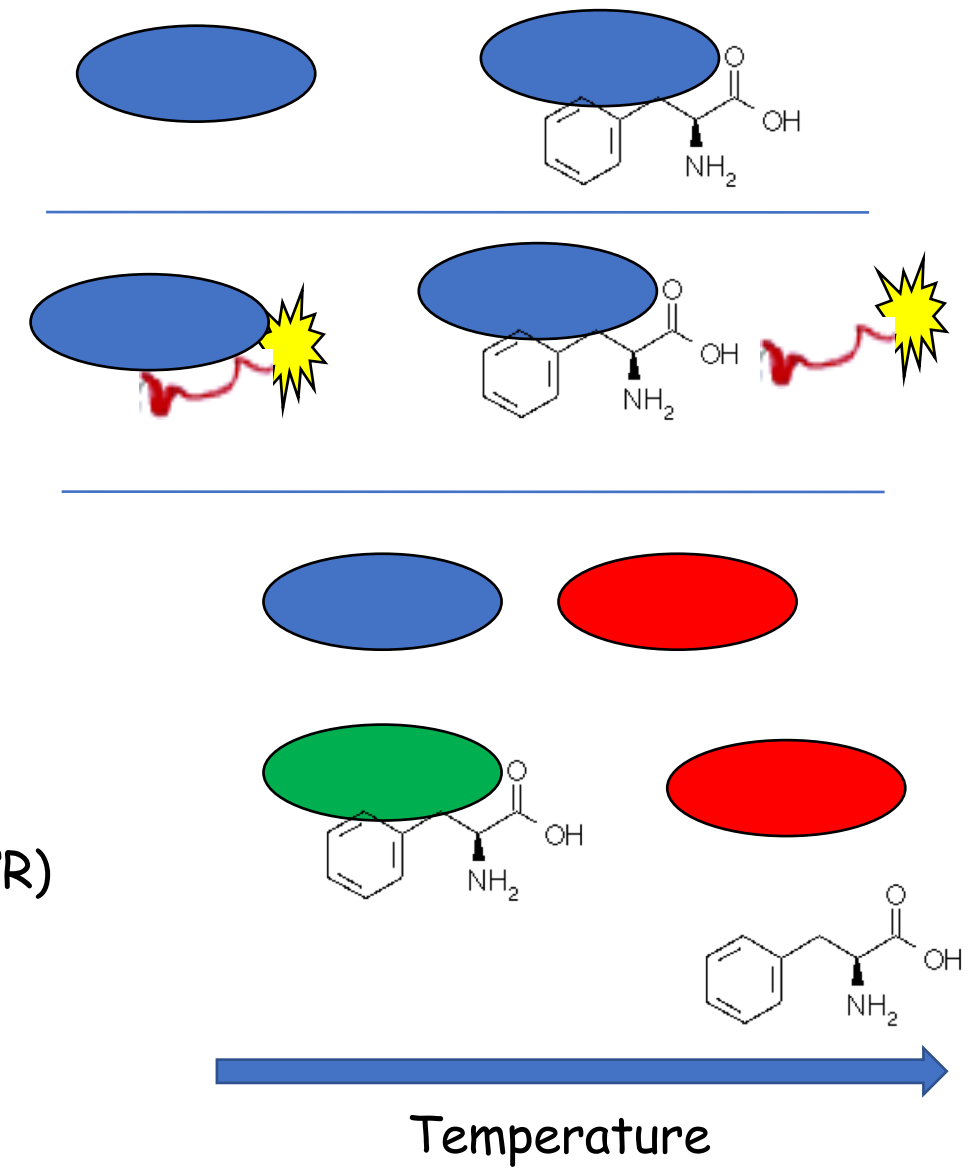
Analytical centrifugation

Fluorescence anisotropy – competition

Stability measurements/Mass-action law

DSC

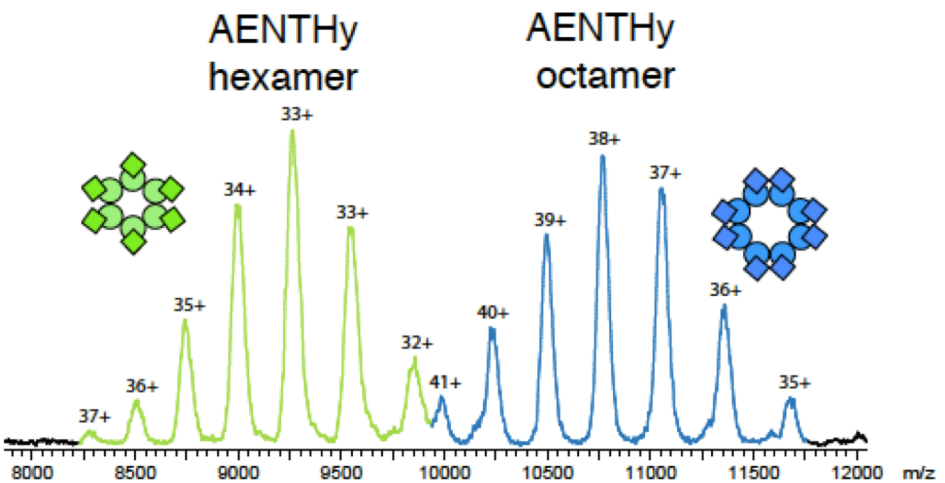
Differential Scanning Fluorimetry
(DSF, ThermoFluor, SYPRO Orange, qPCR)

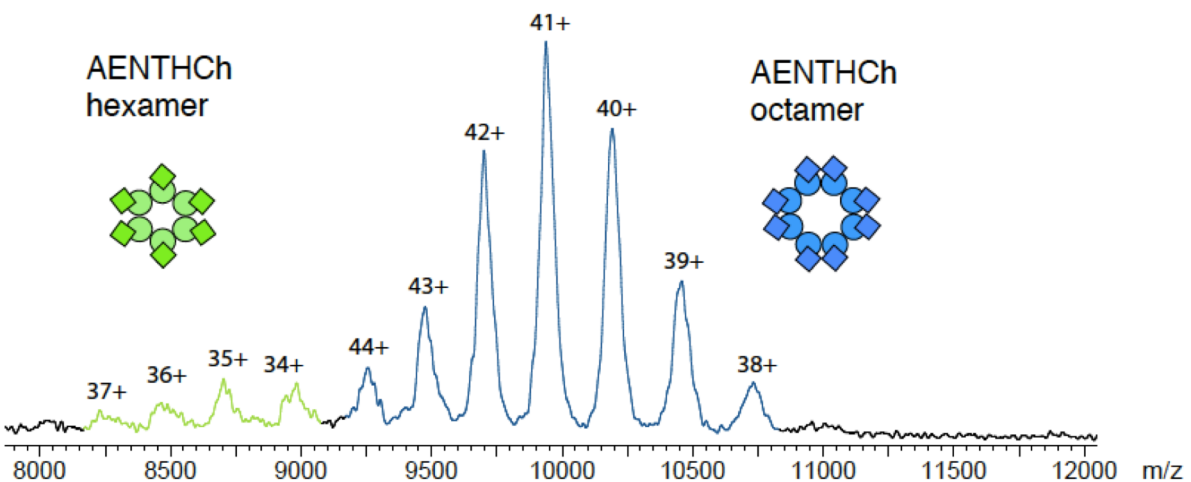


Summary

- Balance between quality and throughput
- Established system
  - Soluble
  - Stable
  - Characterised
- Reliable protein production (>1g)
- Model ligand
- Source of potential ligands to test
- Ways of designing next diversity set

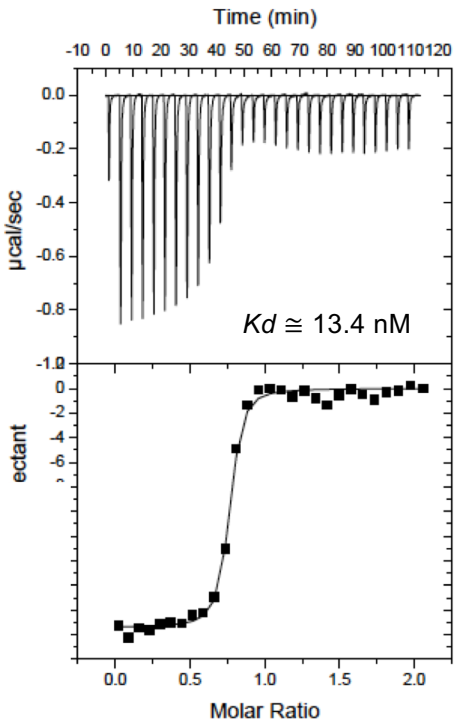
An Hexameric and an Octameric AENTH





NMS

ITC



SAXS
Observed *Mr* (Da) [from *I(0)*]
342 KDa ± 30 KDa

Expected Mr Hexamer: 313 KDa
Expected Mr Octamer: 417 KDa
Expected Average: 365 KDa

Question

-Are there any structural differences could explain the divergence for the biological functions of HIP1R/Sla2 vs CALM proteins?

The ANTH domains: Hip1R vs CALM

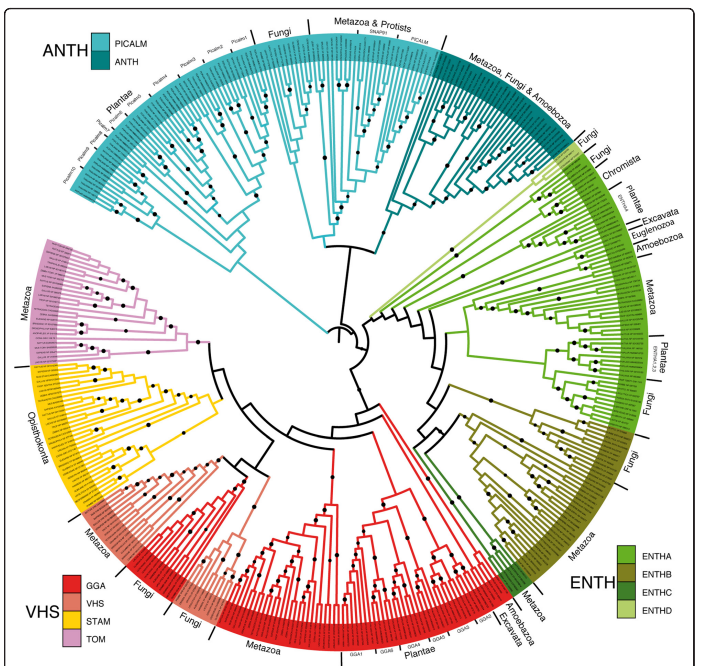
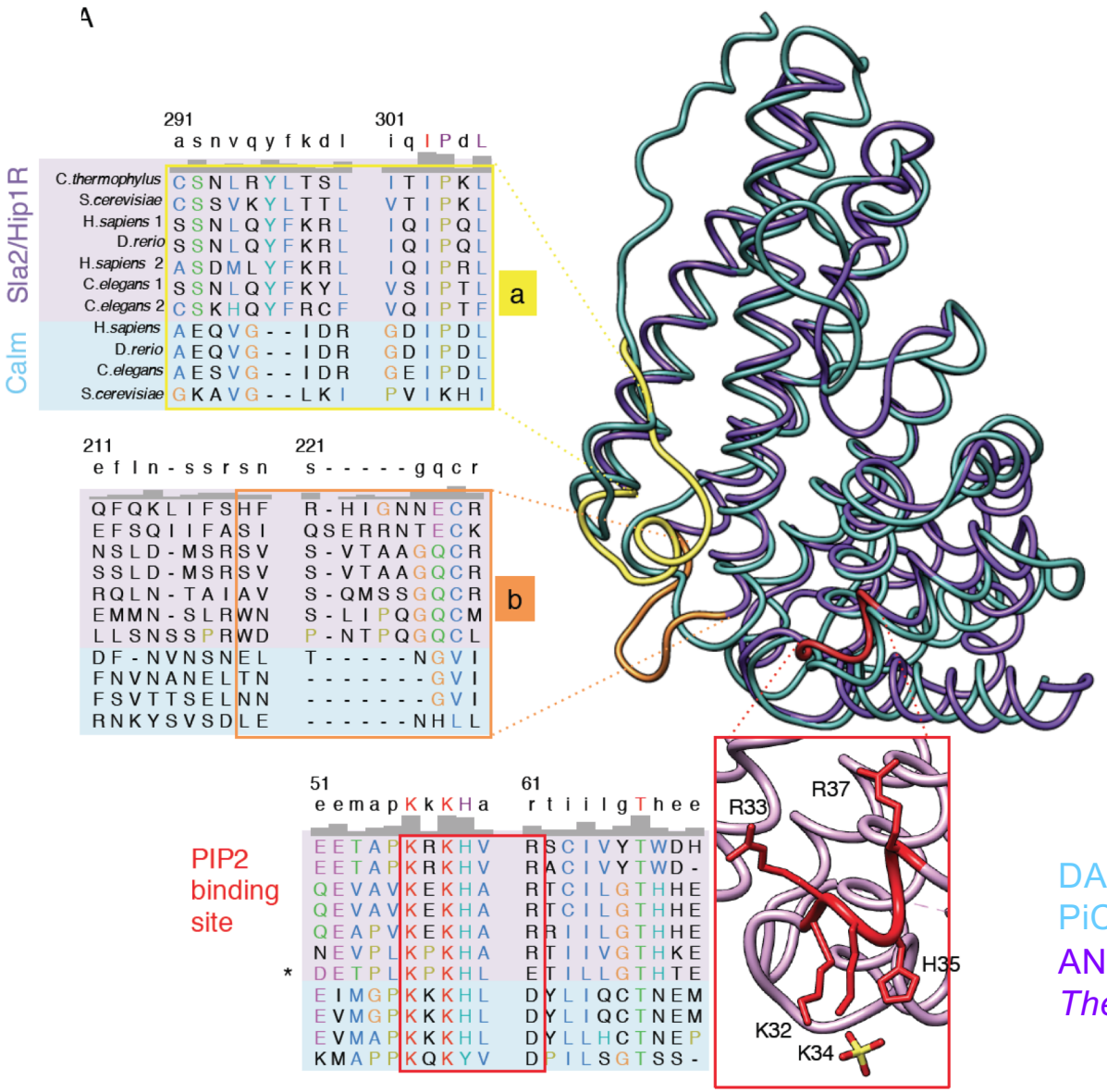


Figure 1 Phylogeny of the proteins with an ENTH, ANTH or VHS domain. Unrooted tree displaying the grouping of most ANTH, ENTH and VHS proteins from 42 fully sequenced organisms. The tree was generated using the highly conserved α-helices 2 to 7 of the N-terminal domain. (●) indicates branches with at least 70% confidence after 500 bootstrap calculations. Protein subfamilies are indicated by different colors.

De Craene et al. BMC Genomics 2012, 13:297. http://www.biomedcentral.com/1471-2164/13/297

Structural Differences between Calm and ANTH subfamilies

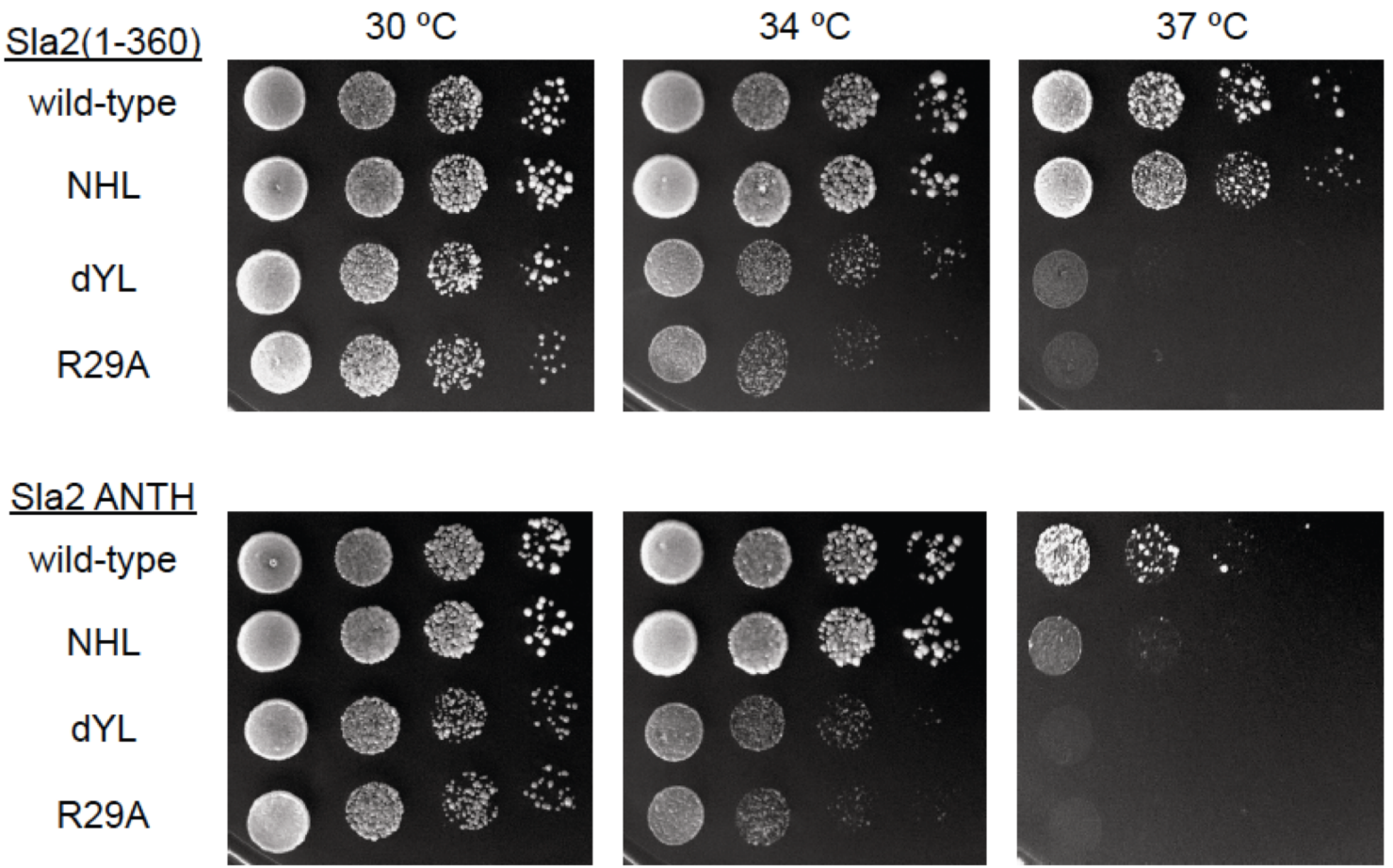


RNA *Chaetomium thermophilum:*
Peer Bork
Thomas Bock

DALI: AP180 (1HX8)
PiCALM (3ZYM)
ANTH from *Chaetomium Thermophilum* (1.8 Å)

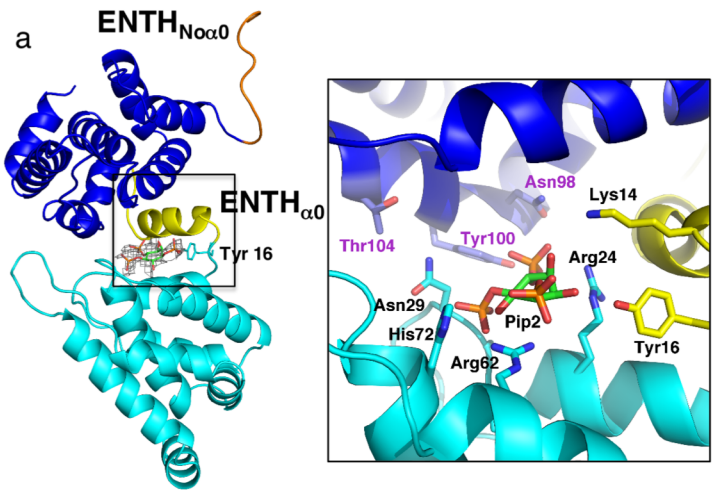
Garcia-Alai *et al.,* Nature Commun, 2018 vol (1) pp. 328

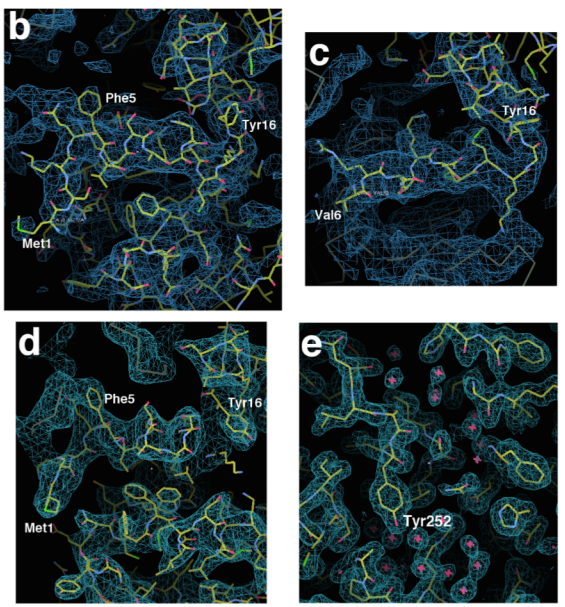
Structural Differences between Calm and ANTH subfamilies

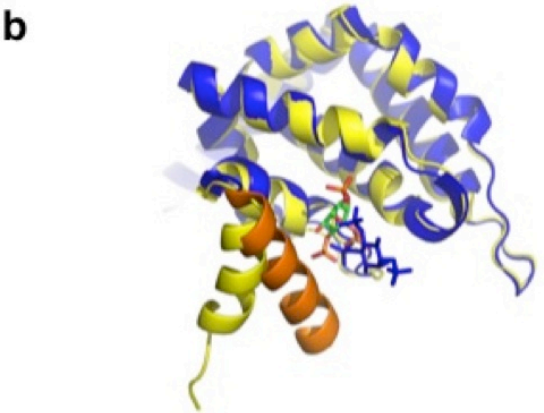


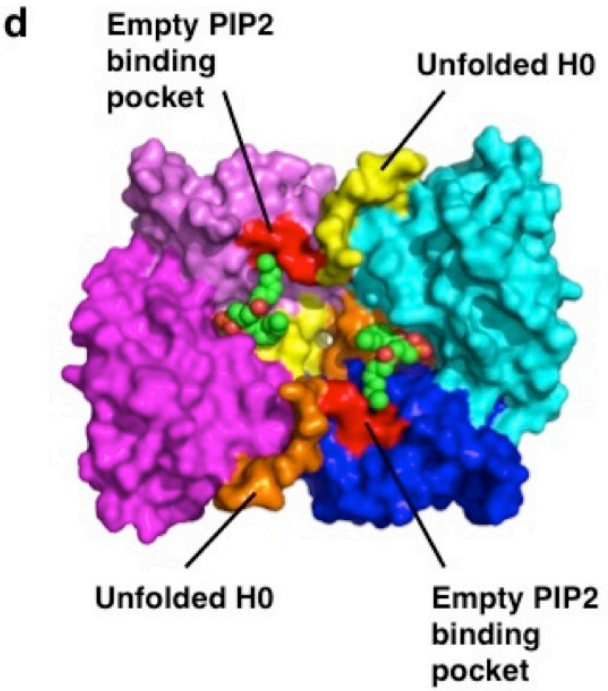
-Growth defects of Sla2 ΔYL and ΔNHL mutant strains. Ten-fold serial dilutions of *sla2*D strains expressing indicated proteins were incubated on SC-Ura plates for 1.5–2 days at 30C, 35C, and 37C.

Crystal structure of the ENTH2/PIP2 complex reveals an allosteric-binding mechanism



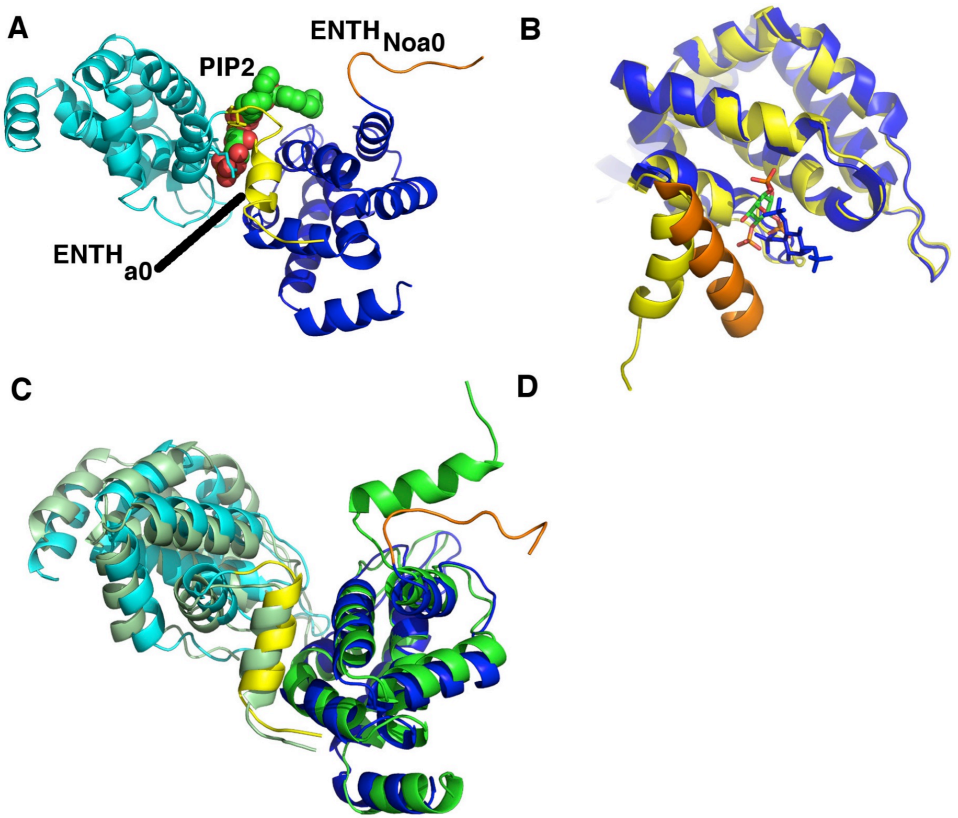


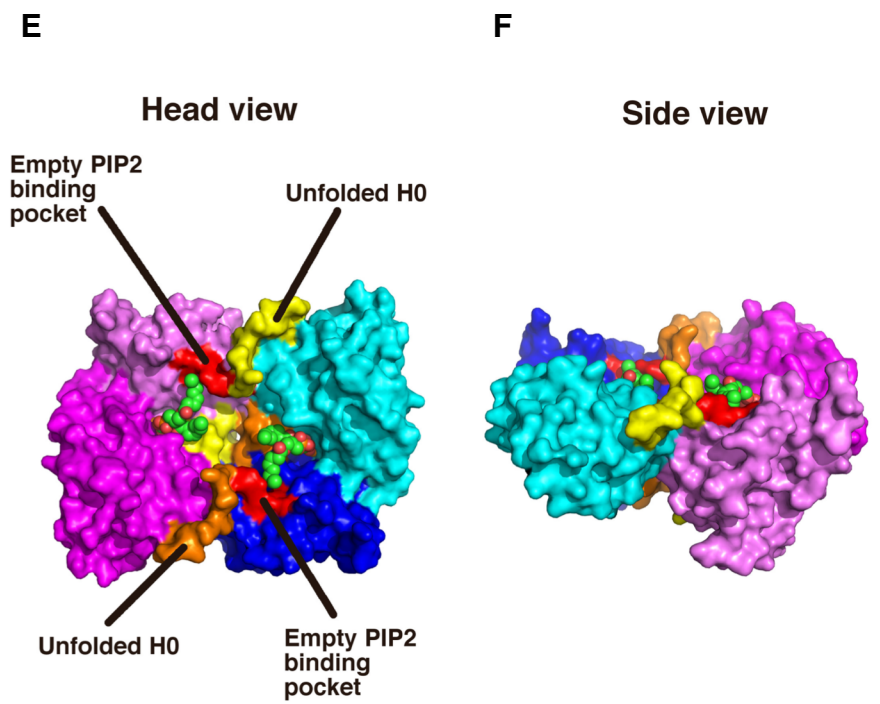




Epsin forms assemblies through phospholipid interfaces

Crystal structure of Epsin ENTH bound to PIP2

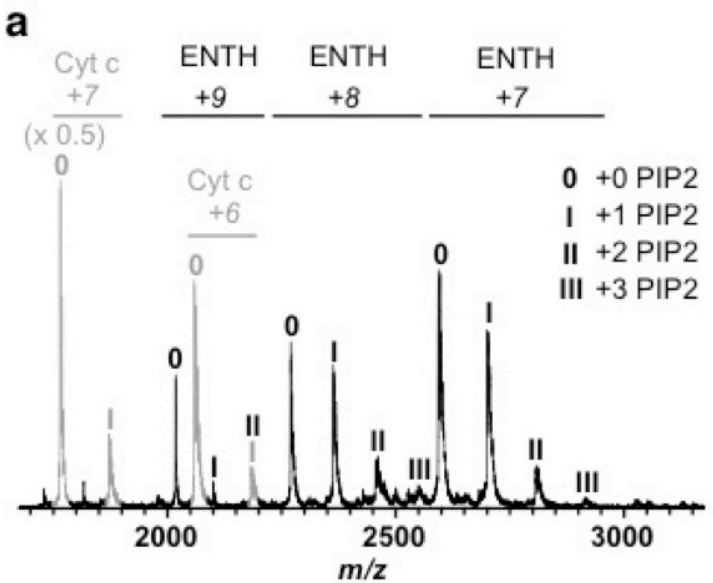


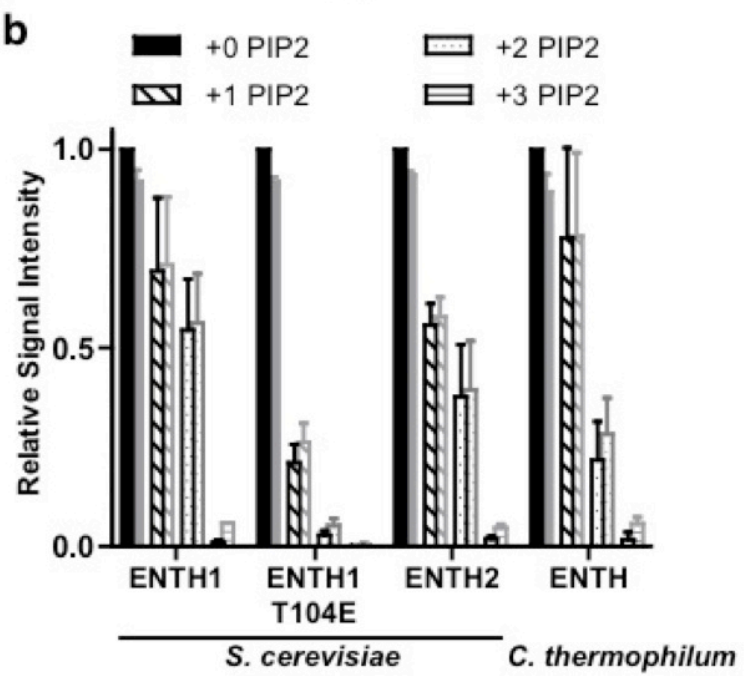


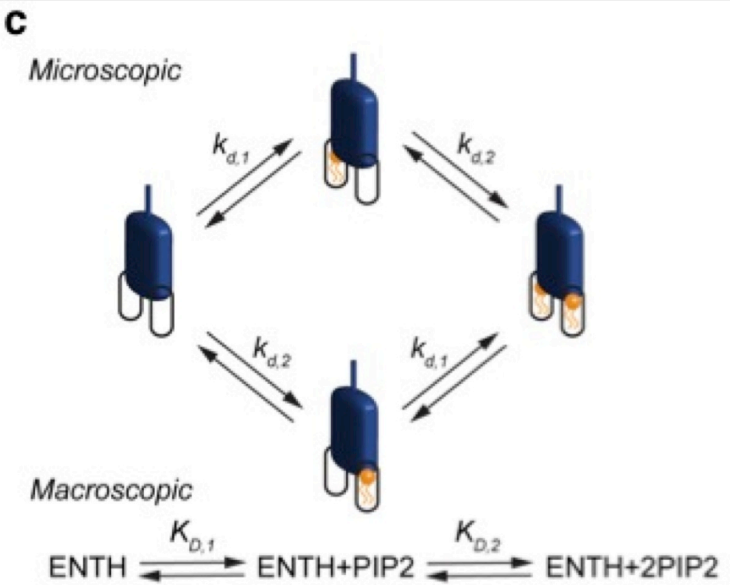
Surface presentation of the ENTH/PIP2 complex showing a tetrameric assembly

- Two building blocks in cyan/blue and magenta/violet
- Tyr 16, Arg 24, Arg 62 and His 72 form the empty PIP2 binding pocket

Cooperative binding of PIP2 to the ENTH domain of epsin







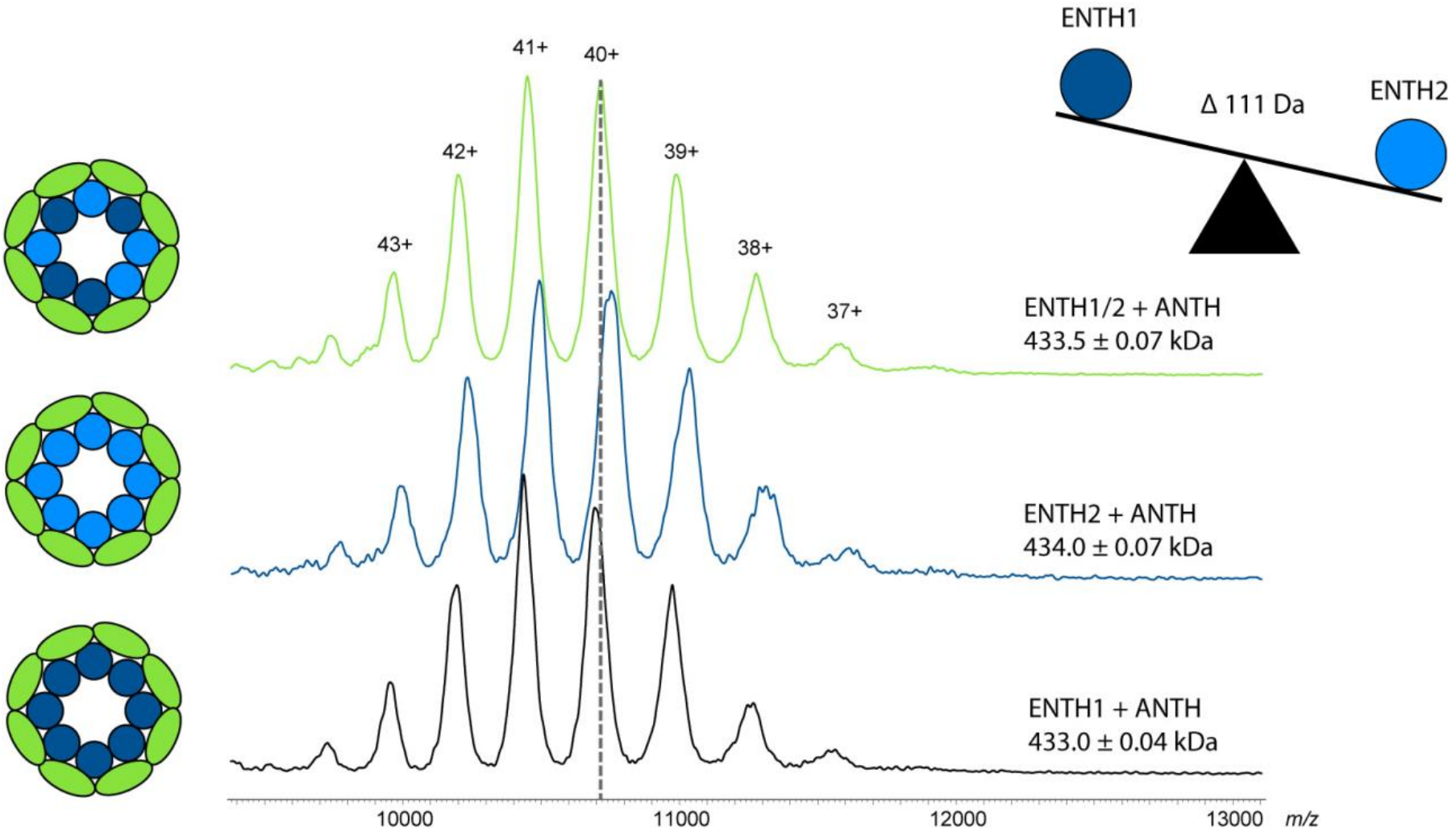
≈ 100 uM

- Relative peak intensities were used to determine the ratio of lipid-bound to non-bound
- Cooperativity of the two binding sites was assessed by reviewing the mathematical relation:

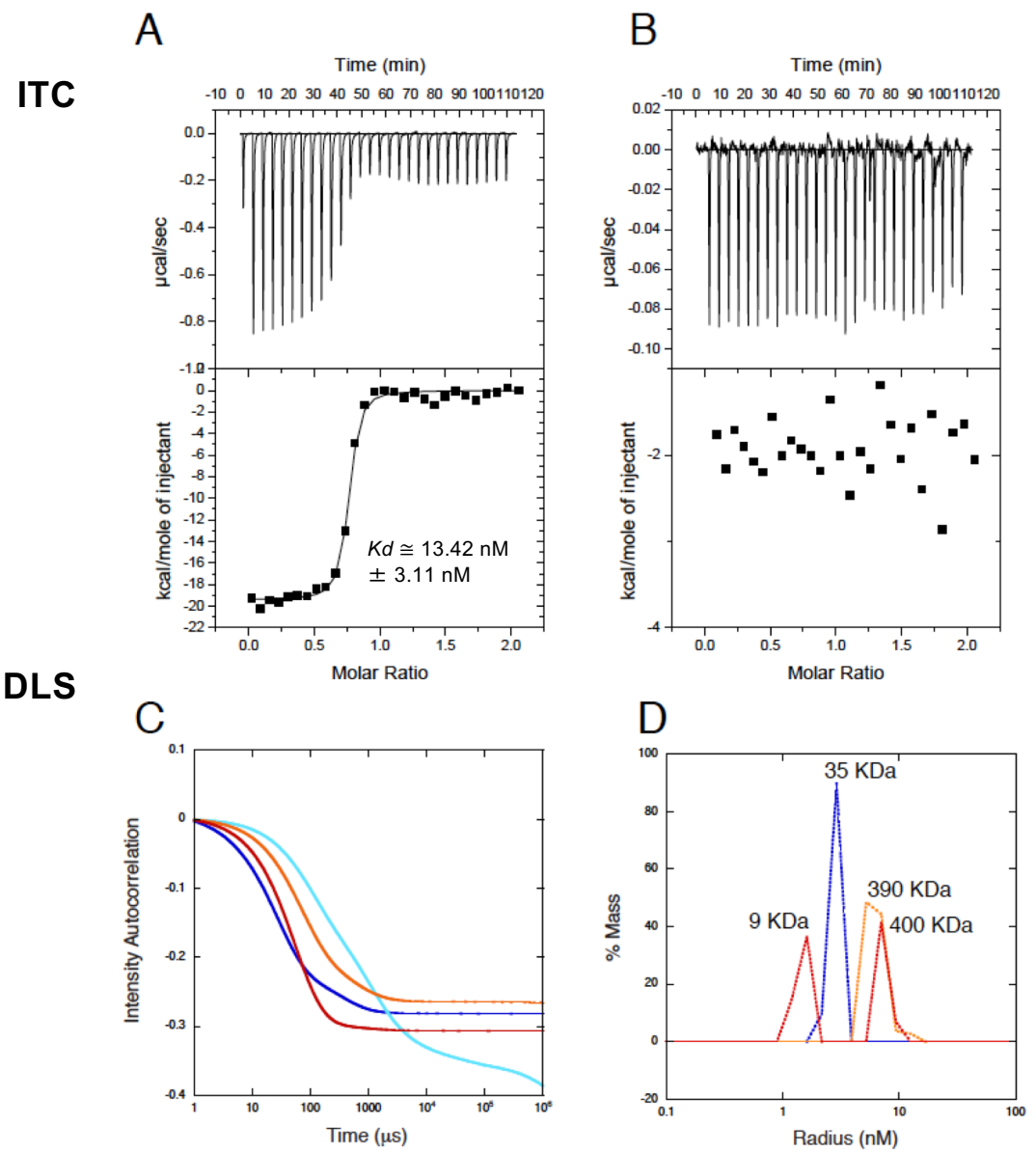
$$K_{D,1} = \frac{k_{d,1} \times k_{d,2}}{k_{d,1} + k_{d,2}}$$

$$K_{D,2} = k_{d,1} + k_{d,2}$$

Complexes with ENTH1 and ENTH2

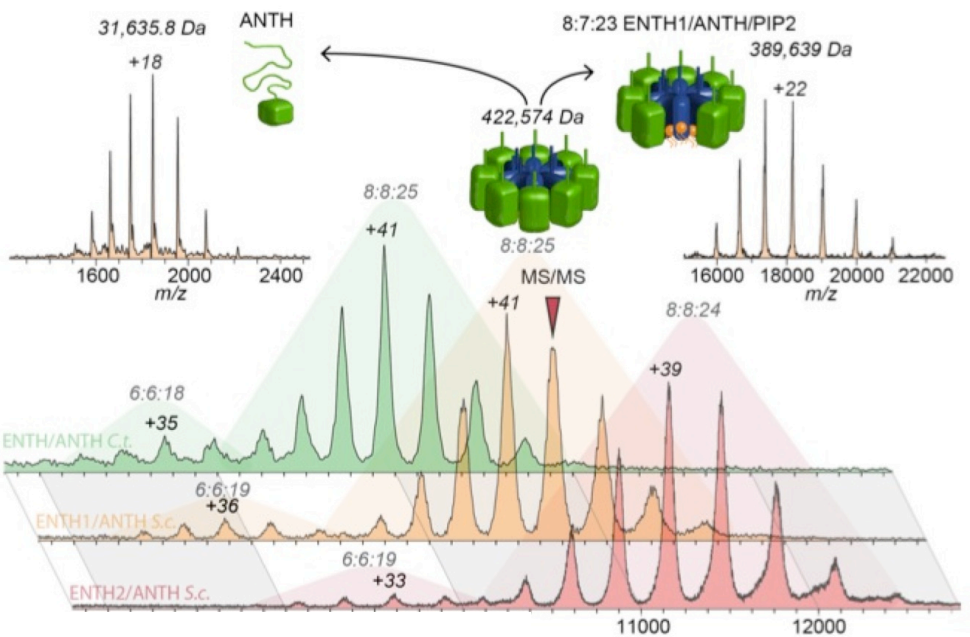


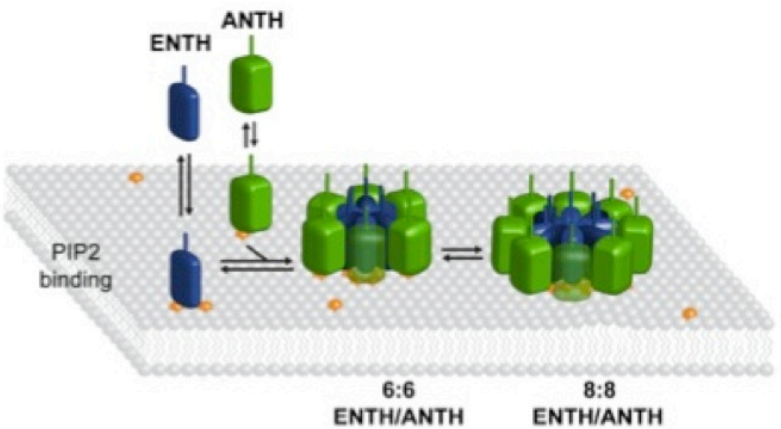
PIP2 availability as the regulatory mechanism for AENTH assembly

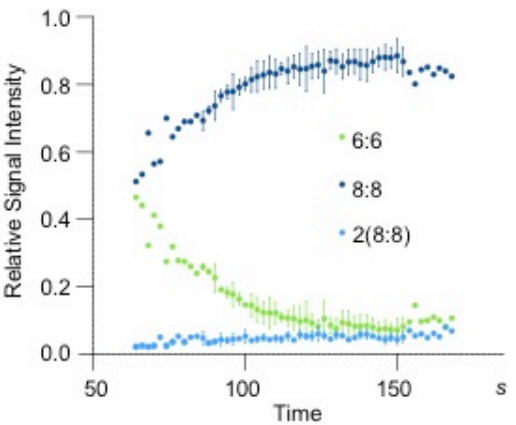


[PIP2]= 0 µM, blue; 80 µM, light-blue; 200 µM, orange and 400 µM, red

Ordered assembly formation of fungal ENTH and Sla2 ANTH

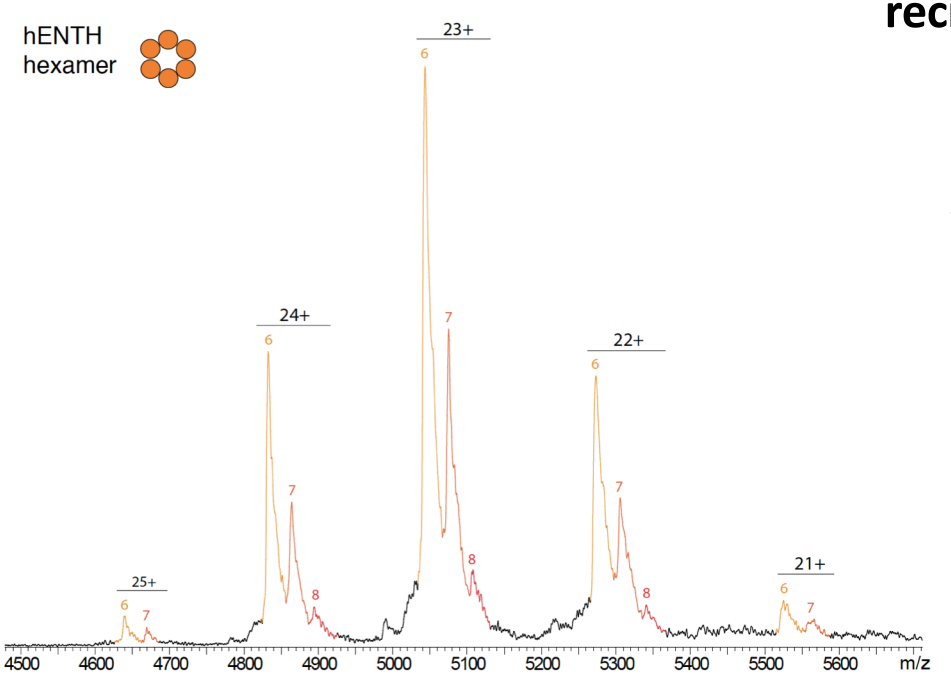


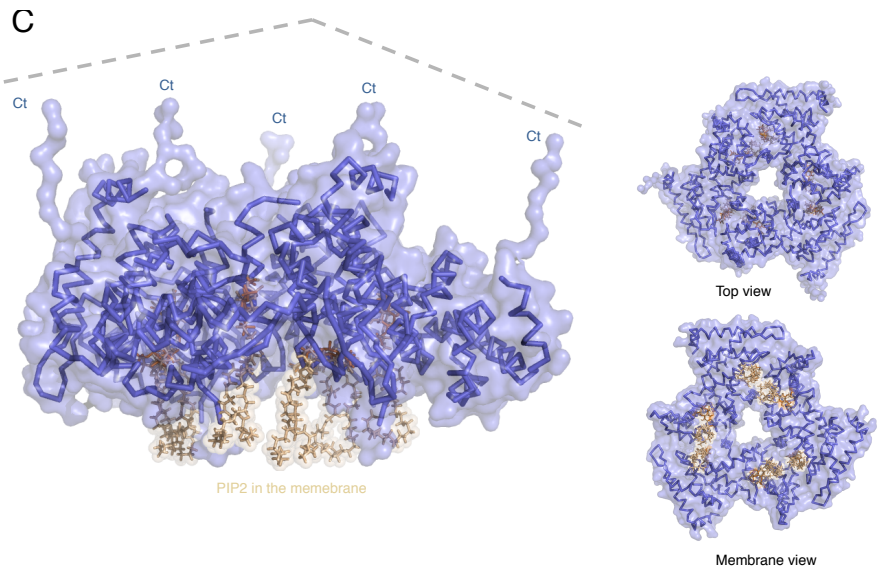




The human ENTH core

- this could explain the **epsin dependent Hip1R recruitment** observed *in vivo…*



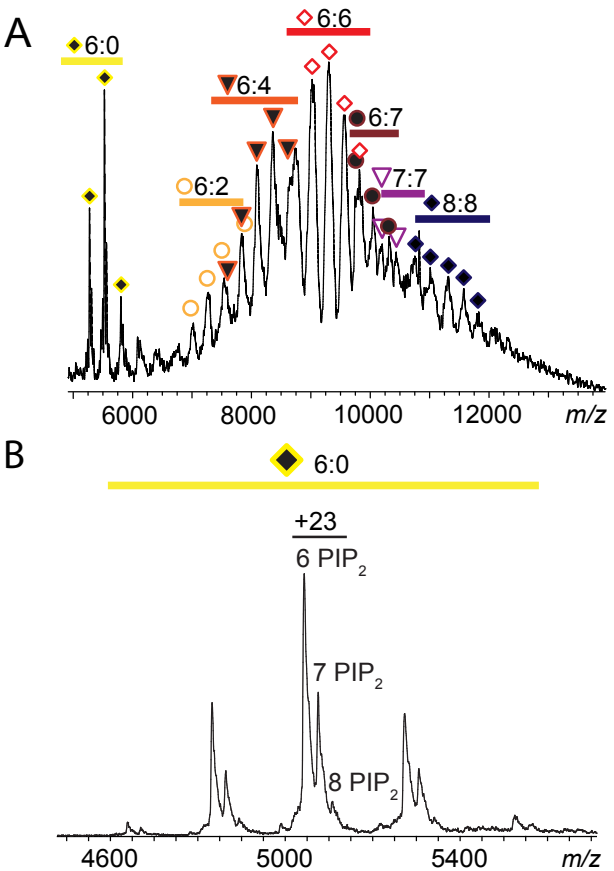


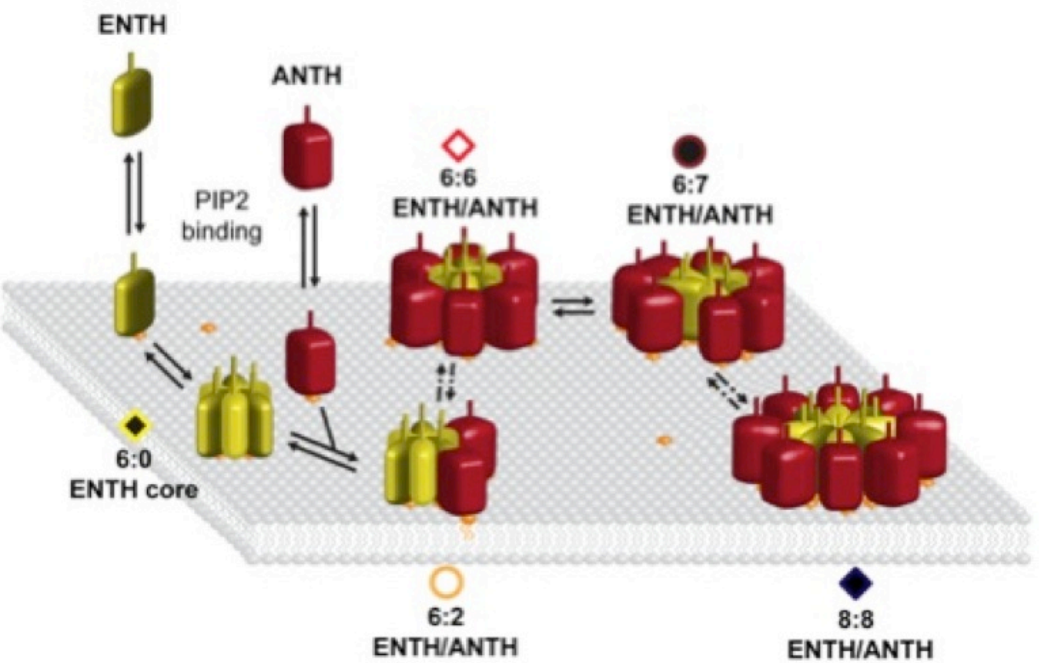
SAXS *ab-initio* model

- Is this ENTH core a precursor for further AENTH oligomerization?

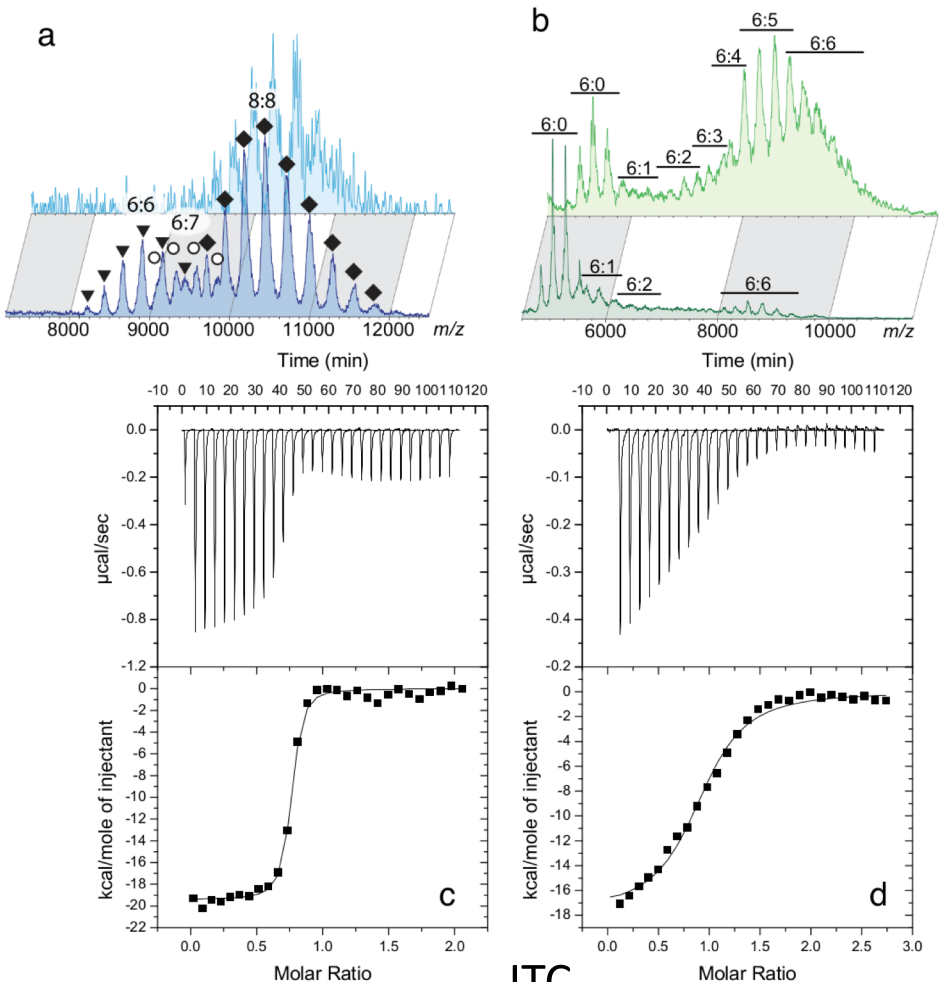
The mechanism of assembly

Biophysics, Crystallography, SAXS and NMS



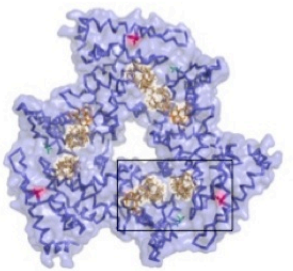


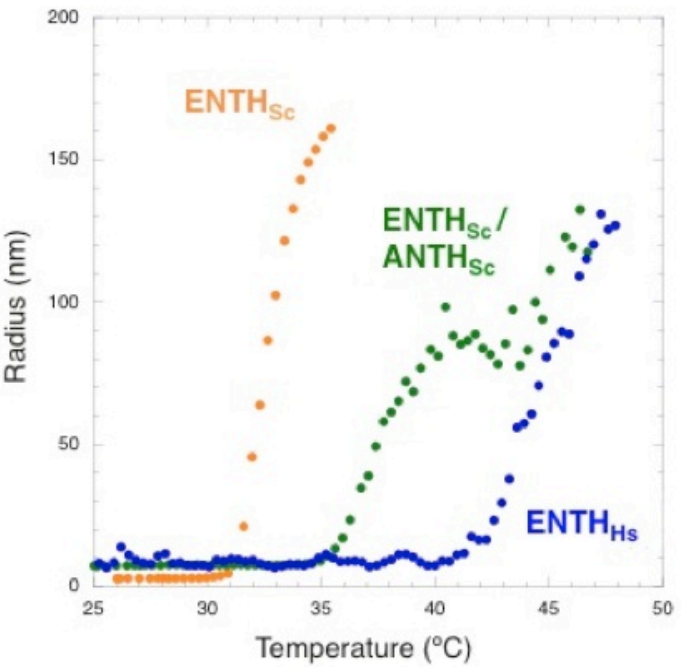
PIP2 driven cross-species interaction of ENTH and ANTHSla2 domains



ITC

Human vs. Yeast



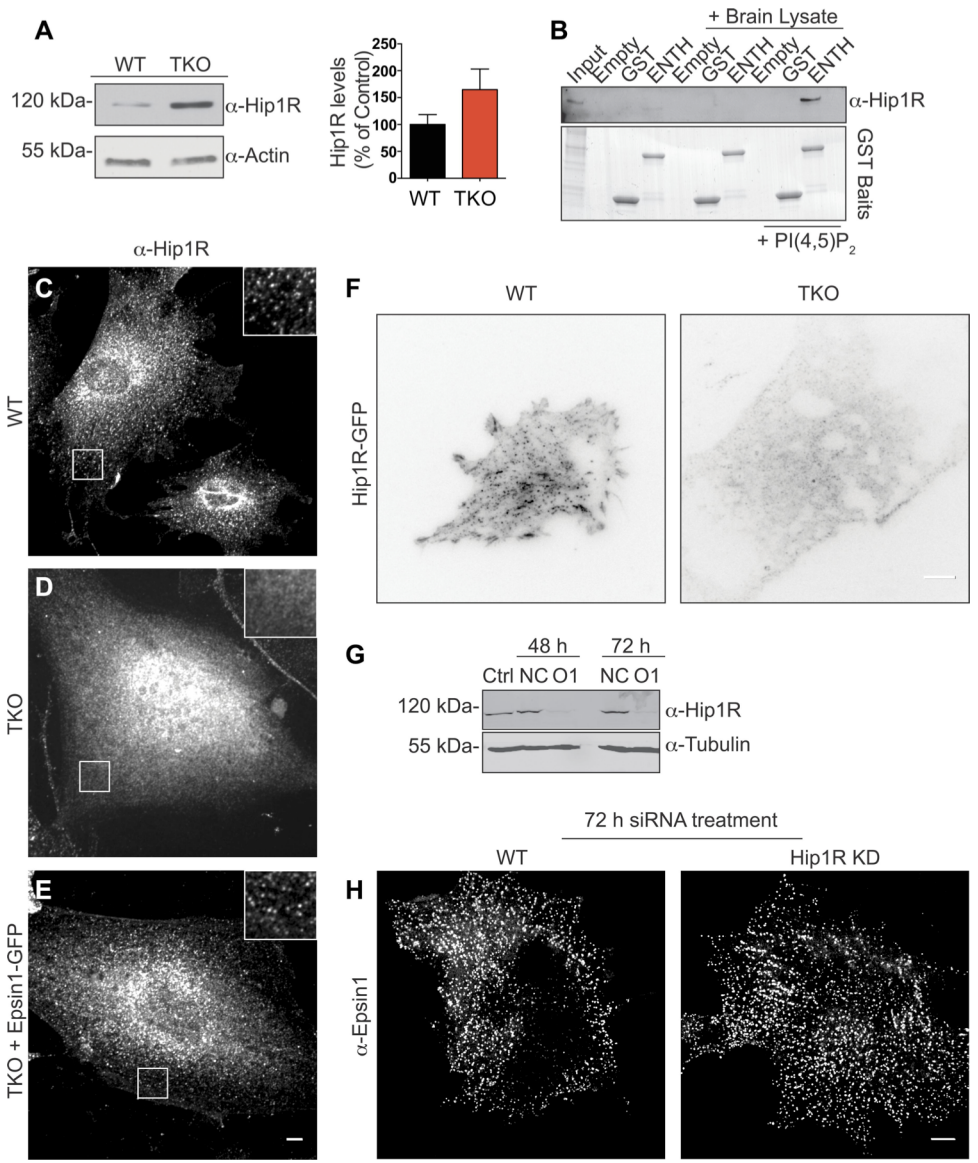


-Thermal denaturation
-Dynamic light scattering

Question

- Is this complex **evolutionary conserved** as a common feature crucial for the clathrin-dependent endocytic path, **or** was it selected as a mechanism occurring **only in yeast?**

The typical clathrin-coated pit-like punctate localization

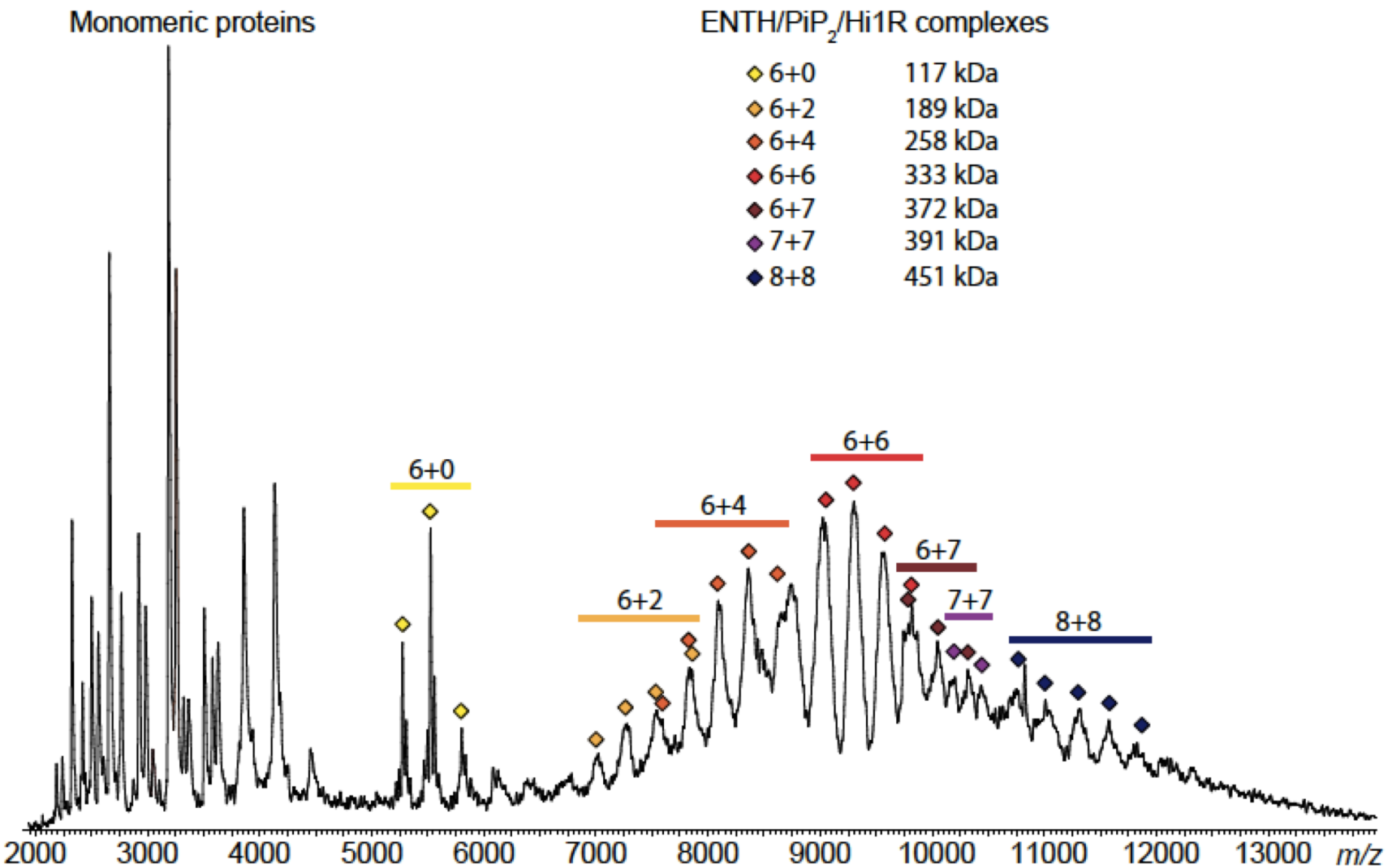


Pull-down experiments from rat brain homogenate, using ENTH domain of epsin 1 as bait revealed an enrichment of Hip1R in the affinity-purified material in the sample also containing PIP2

H) siRNA-mediated knockdown of Hip1R does not affect epsin localization in HeLa cells as shown by epsin immunofluorescence.

Messa *et al*. Elife. 2014 Aug 13;3:e03311.

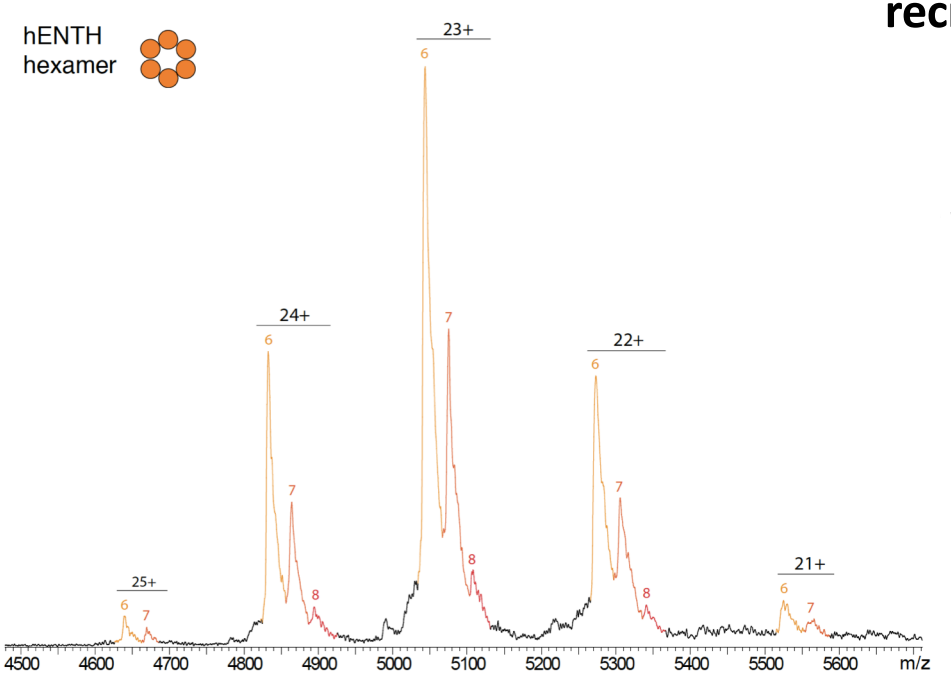
The AENTH human complex

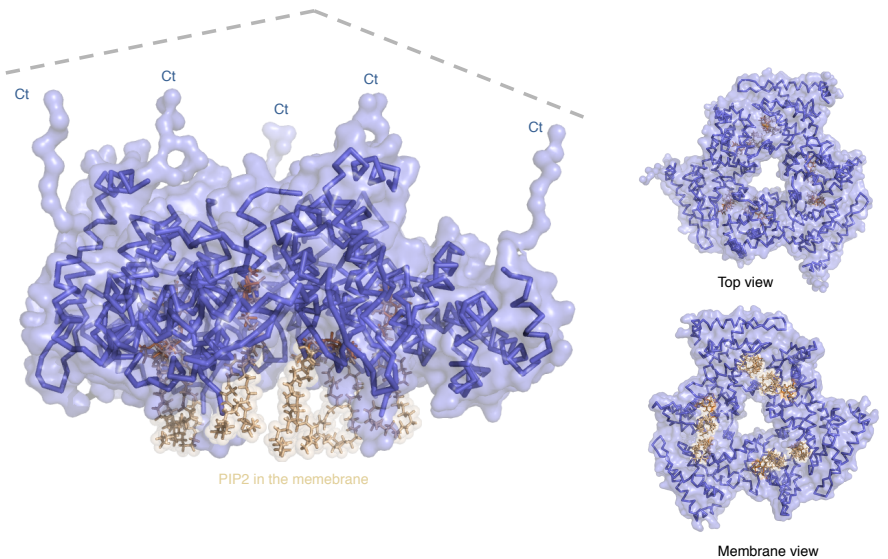


NMS

The human ENTH core

- this could explain the **epsin dependent Hip1R recruitment** observed *in vivo…*

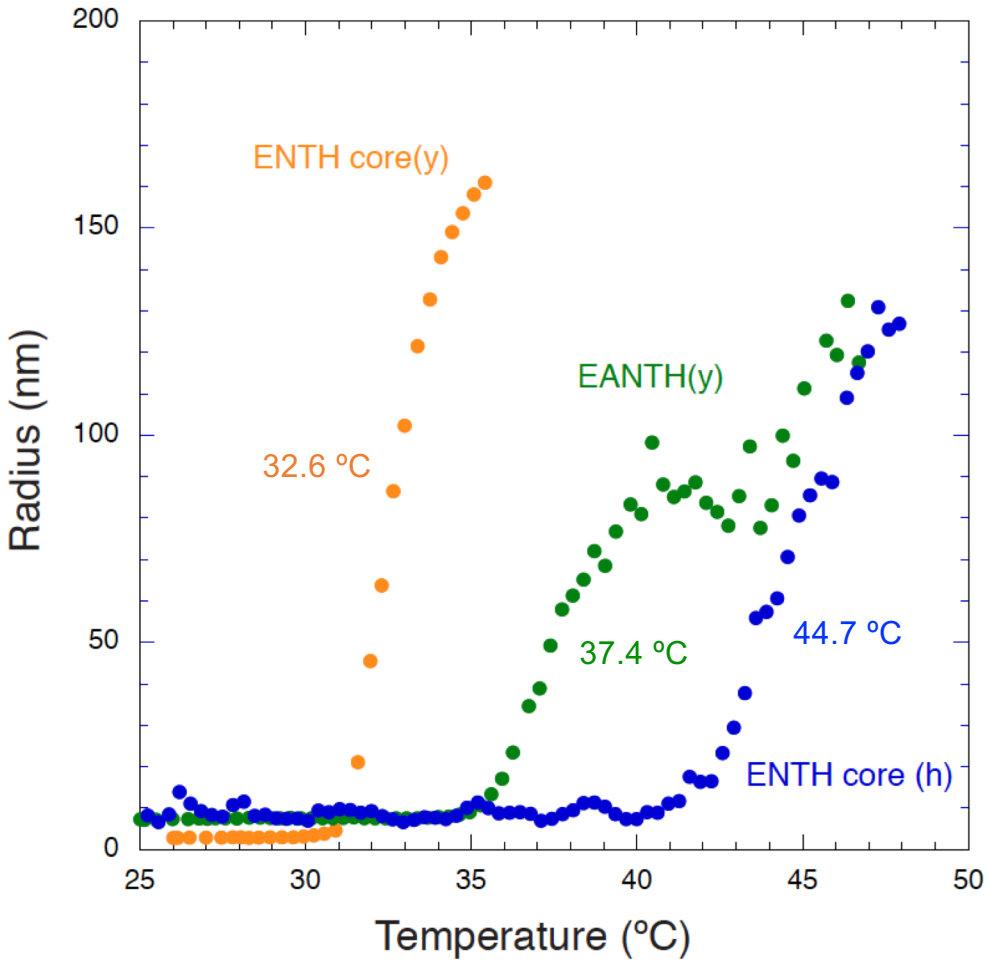




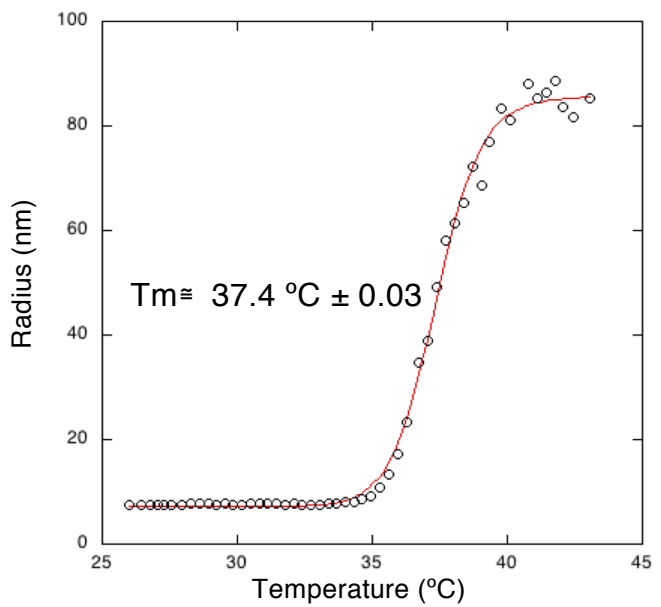
SAXS *ab-initio* model

- Is this ENTH core a precursor for further AENTH oligomerization?

human ENTH core is stable on its own

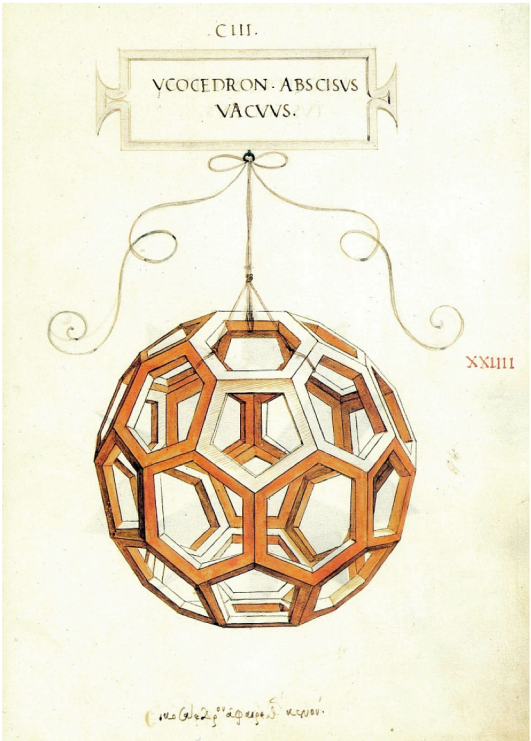


The AENTH complex is more stable than the ENTH-PIP2

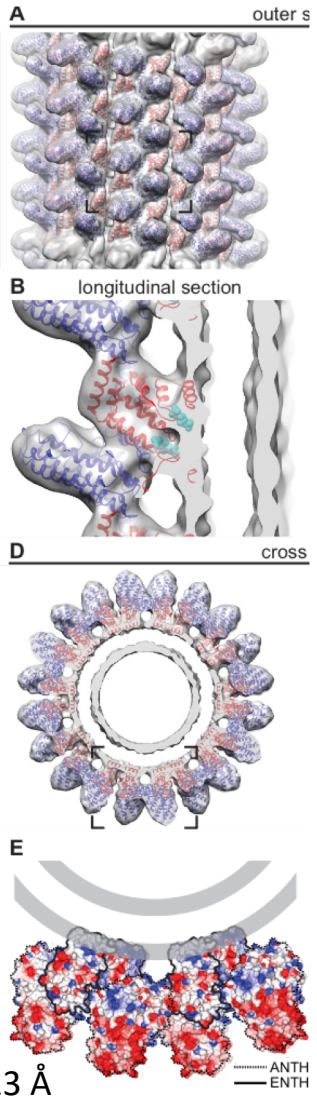


Increase in Rh as a function of temperature monitored by DLS

The puzzle?



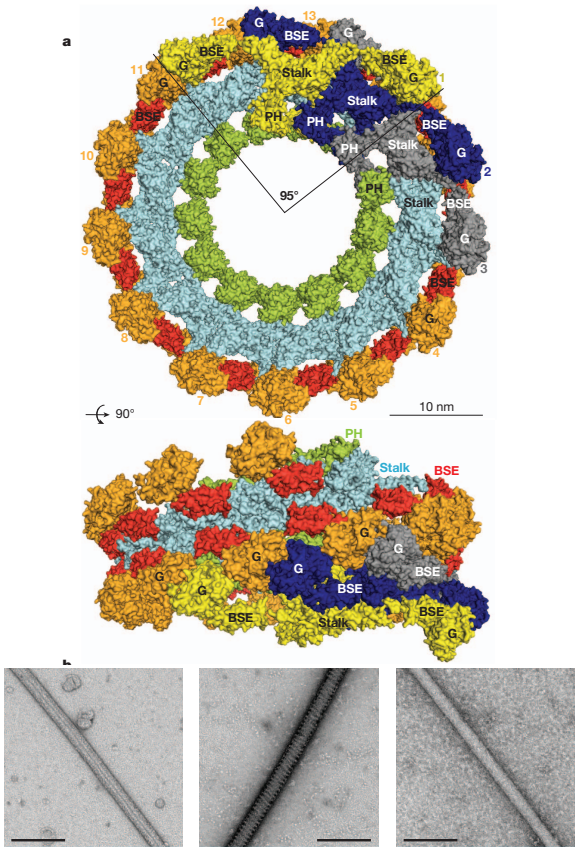
CLC by Leonardo da Vinci?



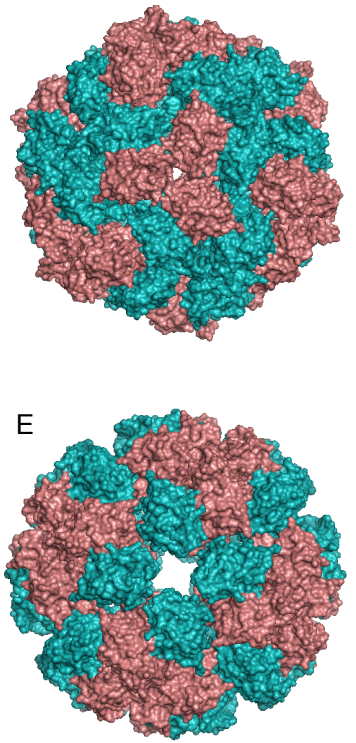
13 Å

Skruzny *et al.* Dev. Cell, 2015

assembly in GUVs



Ford et al. Nature 2011
Faelber et al. Nature 2011



3.8 Å

Our ENTH2 +PIP2
F432 symmetry

probably not…

cryo-EM on GUVs with PIP2

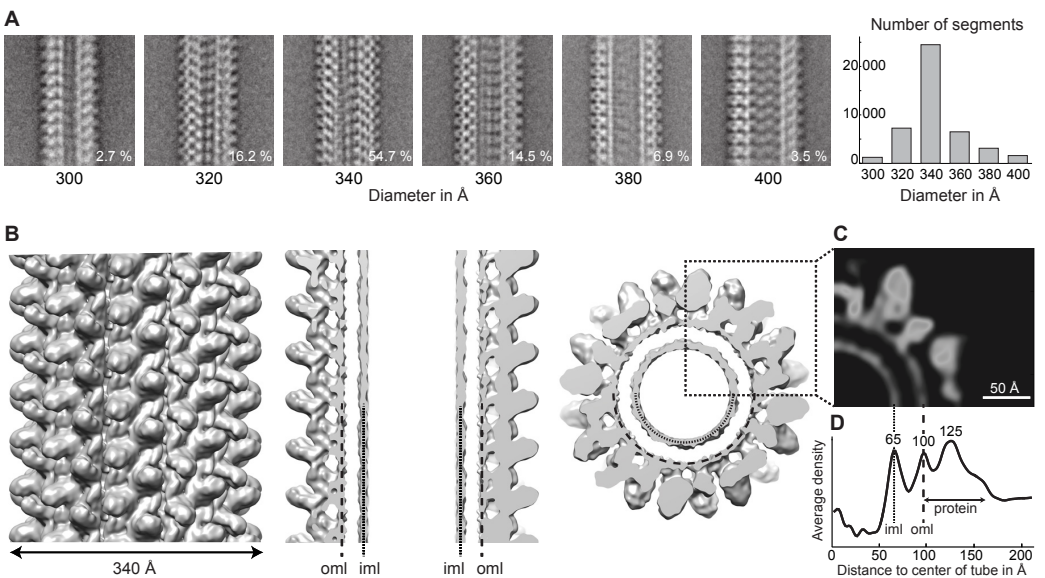


Figure 3

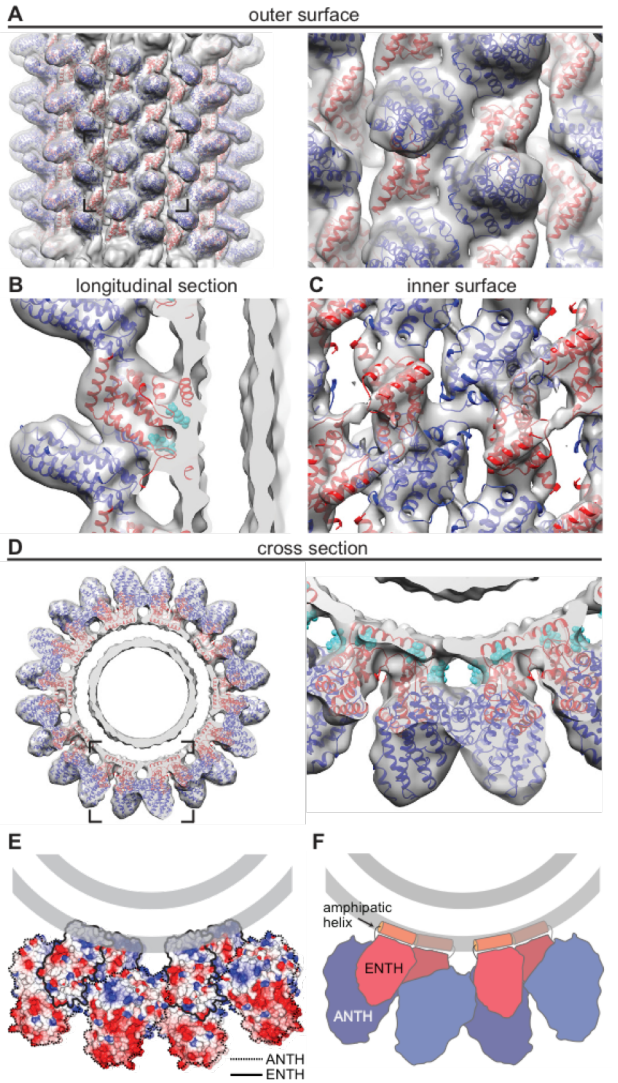
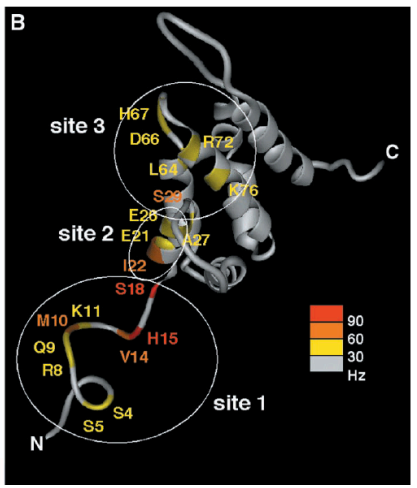


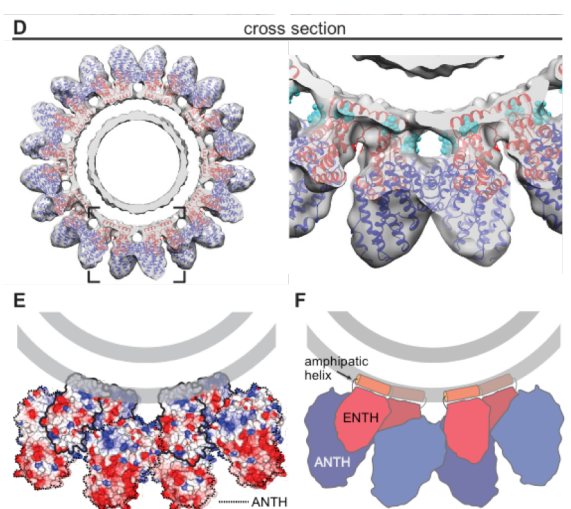
Figure 4

Skruzny, Desfosses, ..., and Sachse, Kaksonen. Developmental Cell, 2015

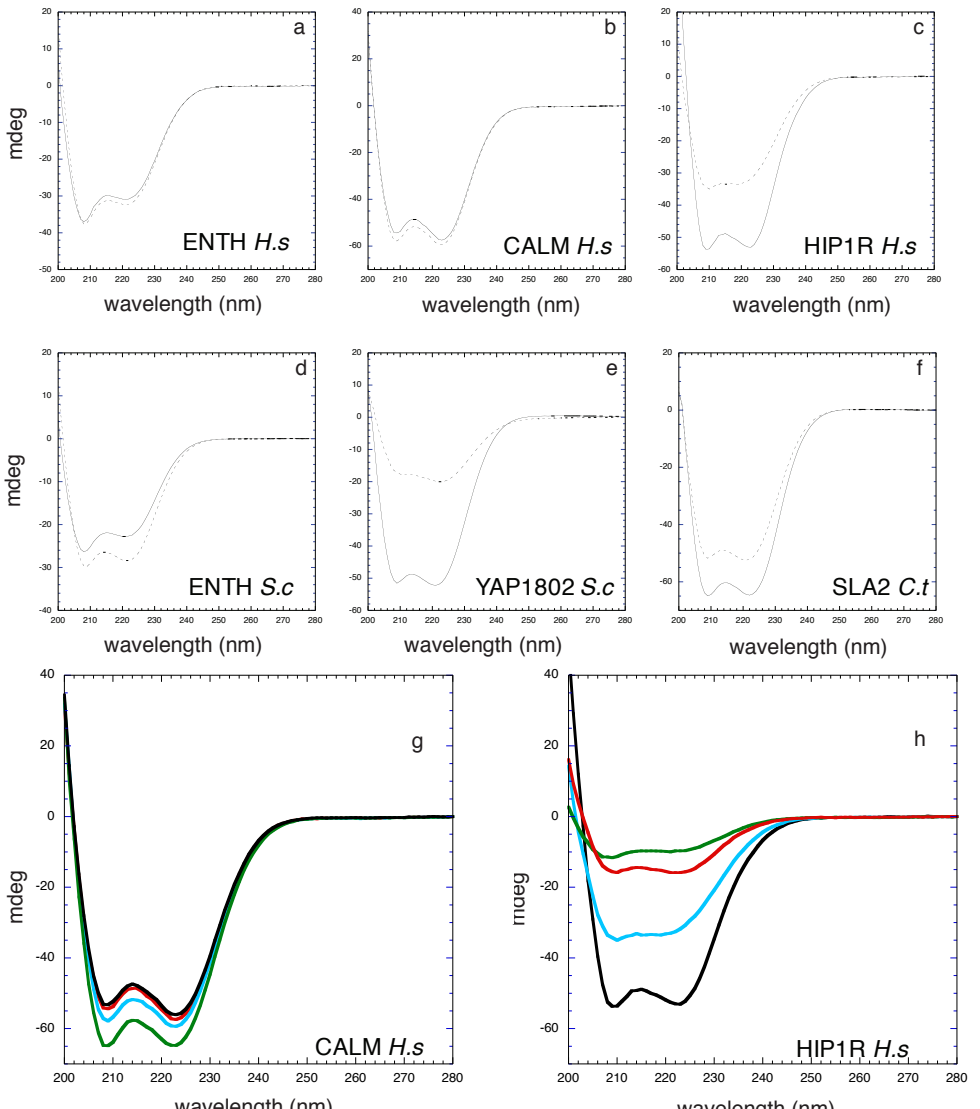
Far-UV CD spectra of ANTH and ENTH domains



Itoh *et al*., Science, 2001 vol 291 pp1047



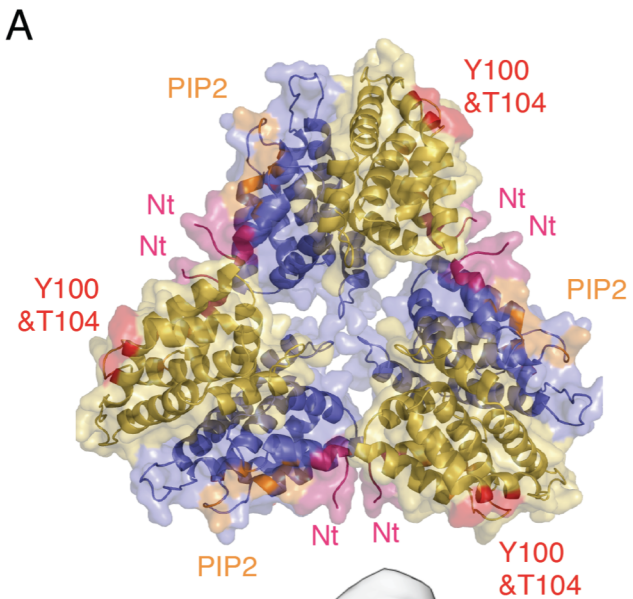
Skruzny et *al*., Developmental Cell, 2015

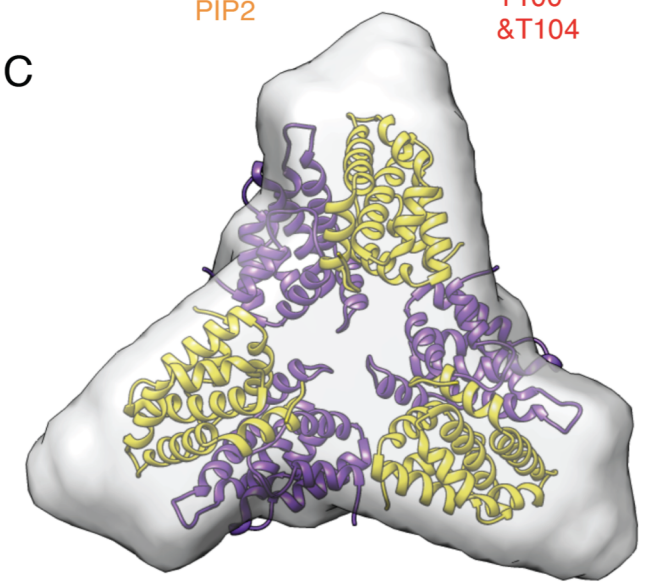


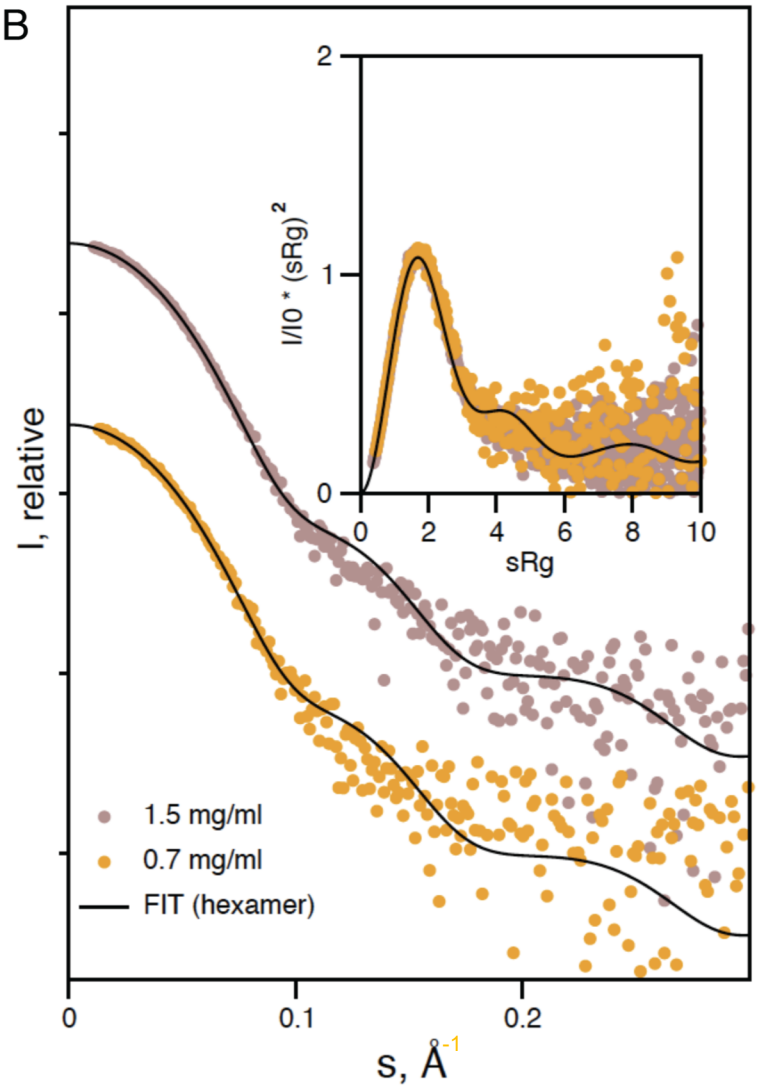
Garcia-Alai *et al.,* Nature Commun, 2018 vol (1) pp. 328

The ENTH core is conserved from yeast to humans

yENTH2 X
(3.8 Å)







SAXS
-The human ENTH core in solution

Summary

- **ENTH** could adopt **different oligomeric states** when binding PIP2 **Crystal structures & DLS**
- We showed the **structural differences between** the **CALM** and **Hip1R** subfamilies of ANTH domains **CD**
- We show the **EANTH** complex occurs in **yeast**, **thermophiles** and **humans** **ITC & NMS**

Thank you!

Acknowledgements:

IMPs Project

Vadim Kotov
Thomas C. Marlovits
(UKE, CSSB)
Kim Bartels
Christian Löw
(EMBL, CSSB)
Katharina Veith
Inokentijs Josts
Henning Tidow
(Uni HH)
Udaya Subhramanyam
Jörg Labahn
(Jülich, CSSB)
Isabel Moraes
(National Physical Laboratory, UK)

The Team

Stephan Niebling Angelica Struve Lucas DeFelipe

Javier Lizarrondo Christian Günther

EMBL DFG Deutsche Forschungsgemeinschaft iNEXT