Marine Biological Laboratories
Workshop in Molecular Evolution

Adaptive protein evolution: Introduction

Belinda Chang

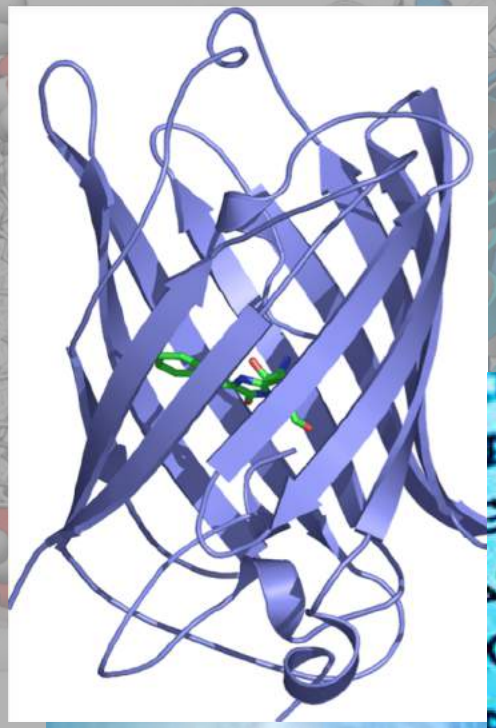
Department of Ecology & Evolutionary Biology
Department of Cell & Systems Biology
University of Toronto

How do protein sequences evolve?

Can we identify evolutionary patterns of selection associated with adaptive shifts in protein function?

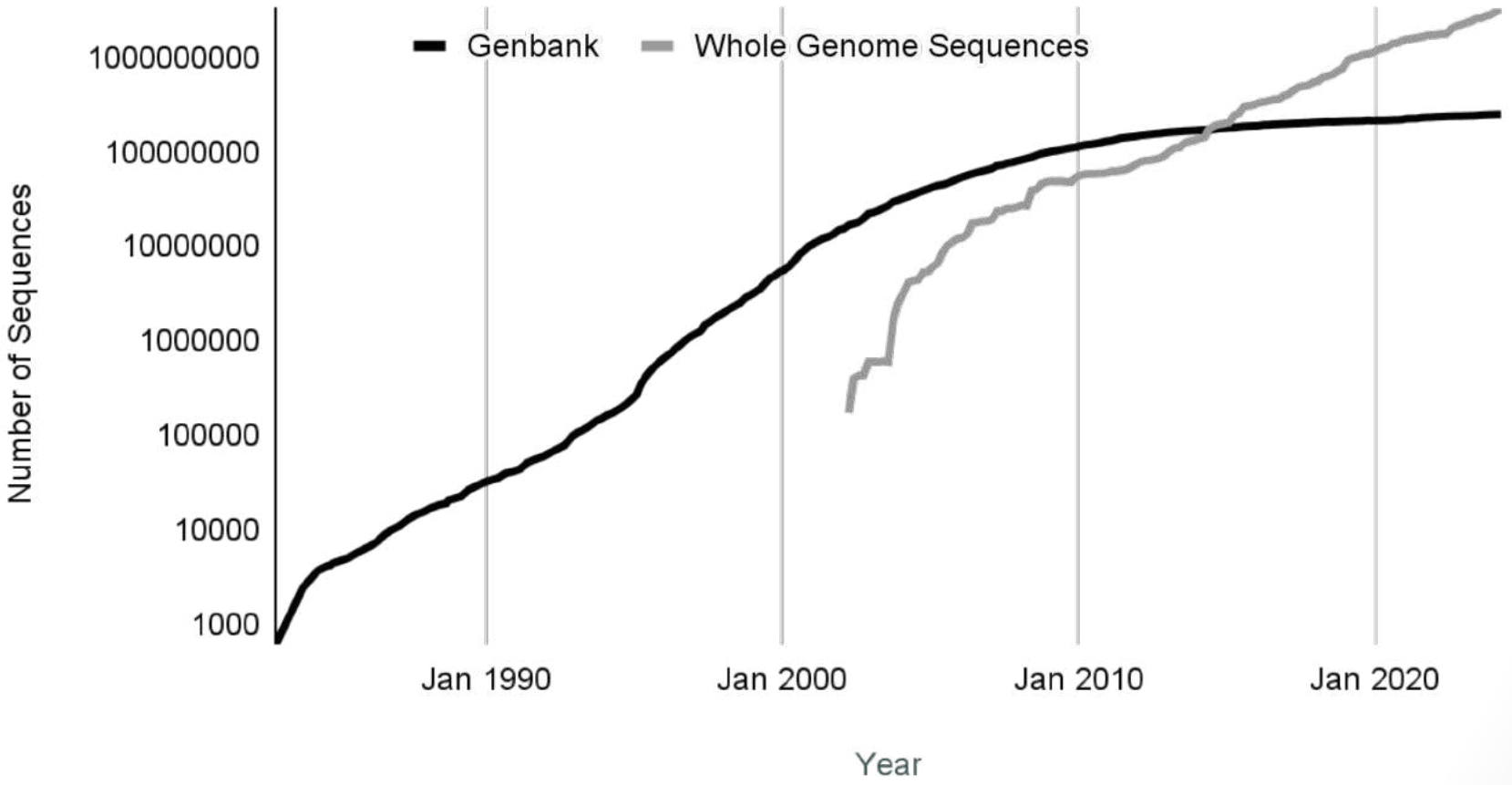
Can we identify the underlying mechanisms associated with adaptive shifts?

Evolution of protein function



$$Q_{ij} = \begin{cases} 0 \\ \pi_j \\ \kappa\pi_j \\ \omega\pi_j \\ \omega\kappa\pi_j \end{cases}$$

Rapid accumulation of sequence data



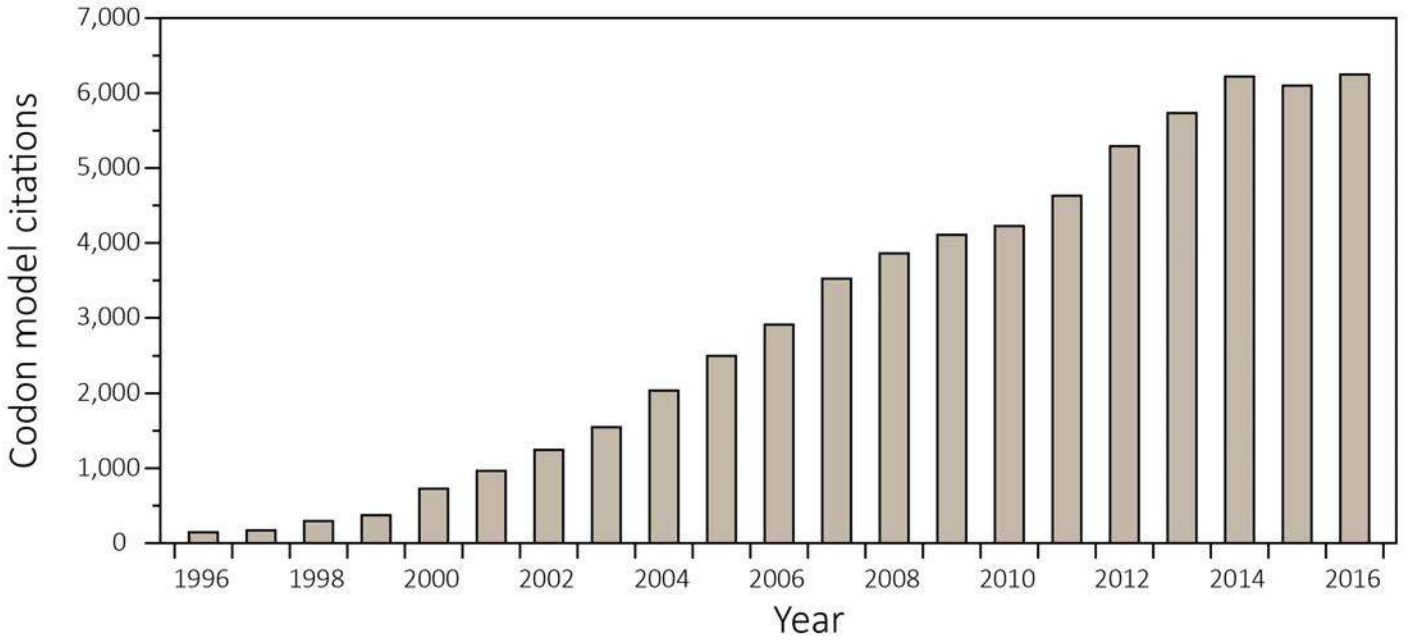
Comparative sequencing can be used to address questions at many different levels

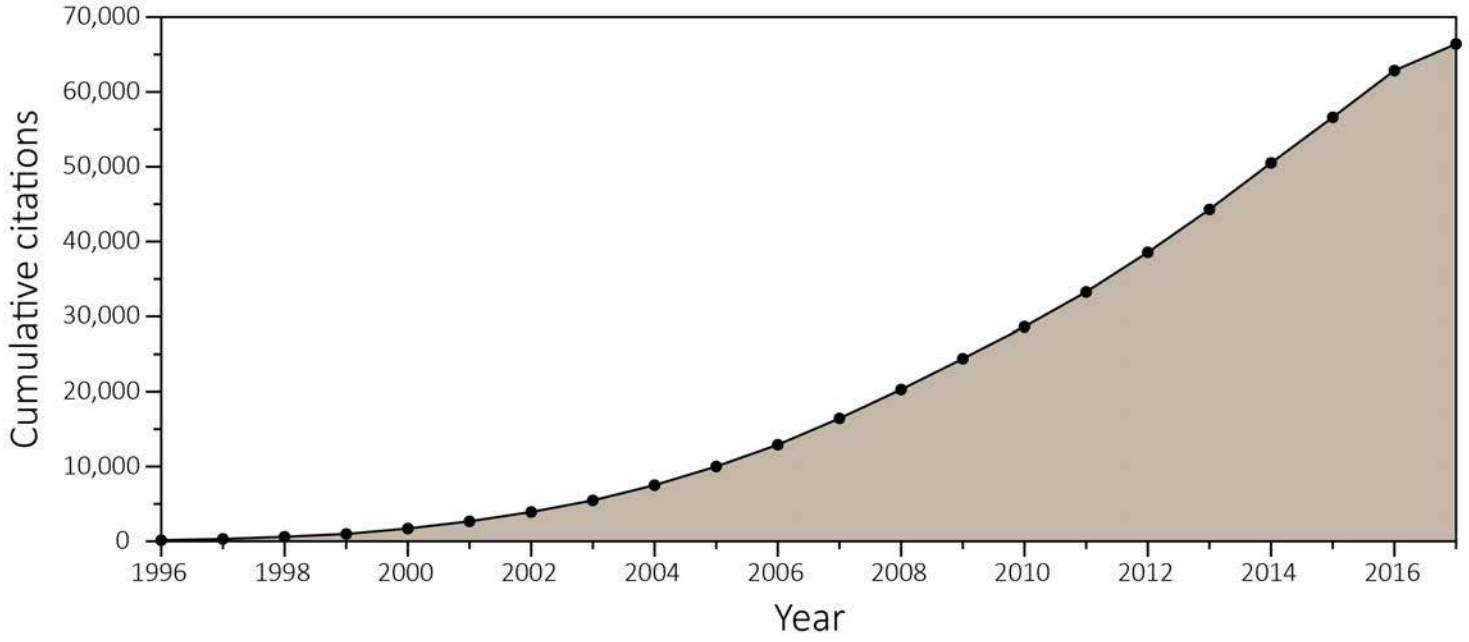
- Evolution of organisms, systematics
- Evolution of genomes
- Evolution of gene regulation
- Evolution of proteins

Evolution of protein-coding genes: Comparative sequence analysis

- Phylogenetically based methods
- Models of evolution (nucleotide, amino acid, codon)
- Hypothesis testing of theories of selection
- dN/dS as a measure of the strength of selection

Increased use of codon models

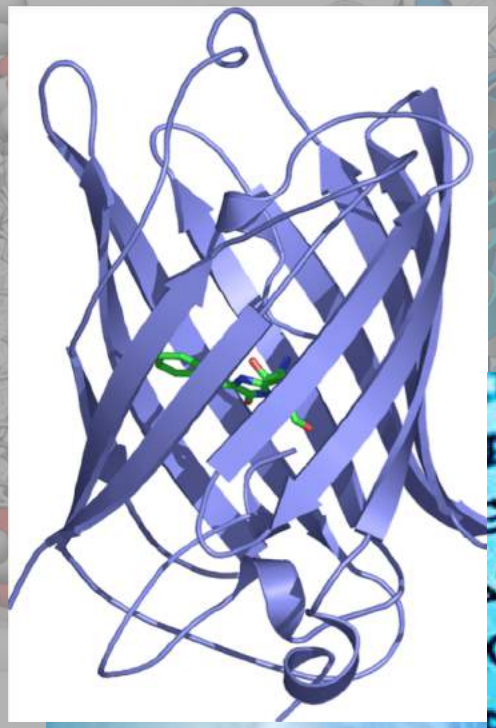




Codon-based testing for positive selection: Why so popular? Hypothesis testing!

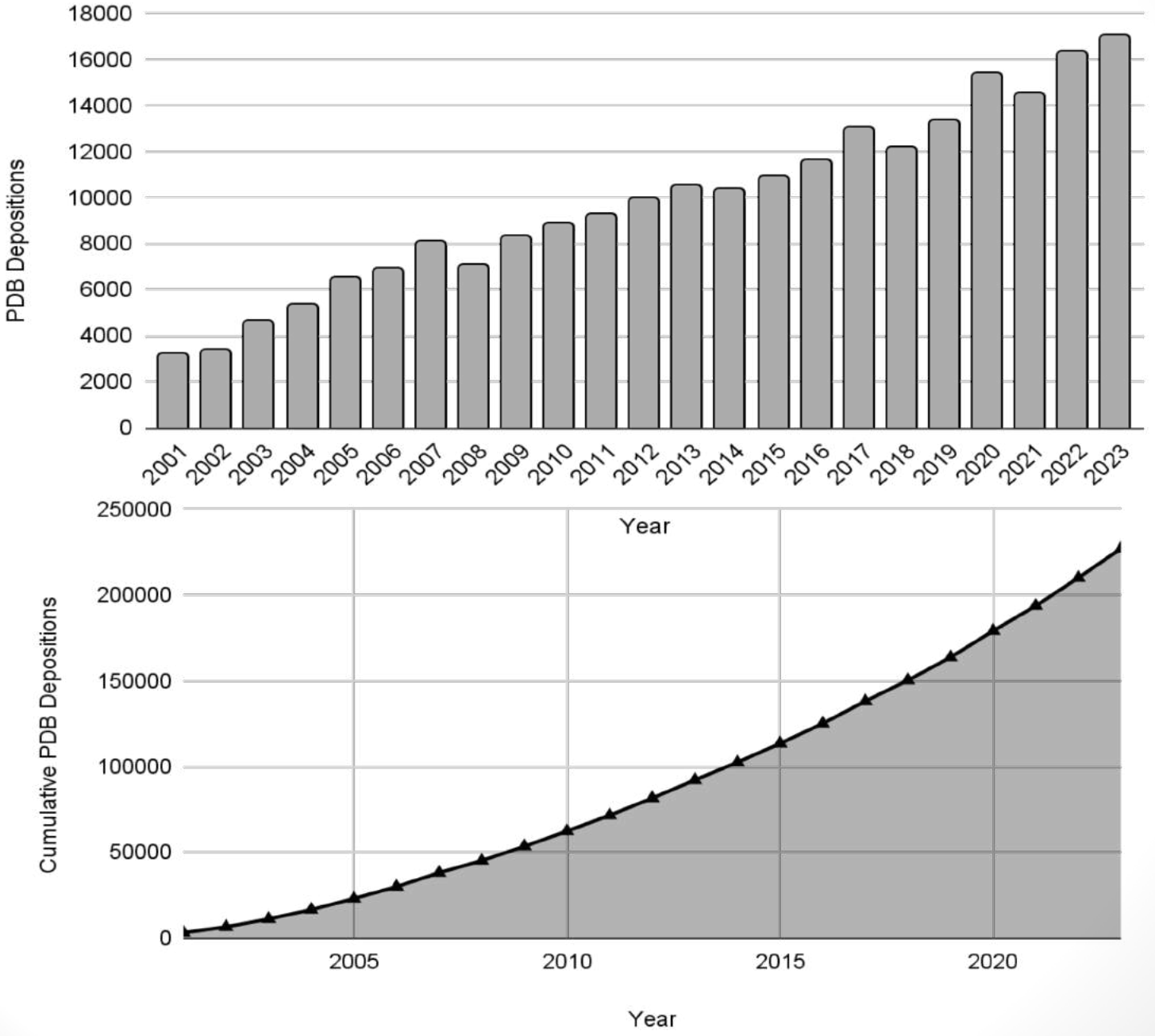
- WHEN selection occurred in evolution
  - Episodic, pervasive, lineage-specific selection
- WHICH proteins were targets of selection
  - Physiology: sensory, metabolic, developmental
- WHICH regions of the protein
  - Mechanisms underlying evolution of function
- Hints as to WHY selection occurred

Evolution of protein function



$$Q_{ij} = \begin{cases} 0 \\ \pi_j \\ \kappa\pi_j \\ \omega\pi_j \\ \omega\kappa\pi_j \end{cases}$$

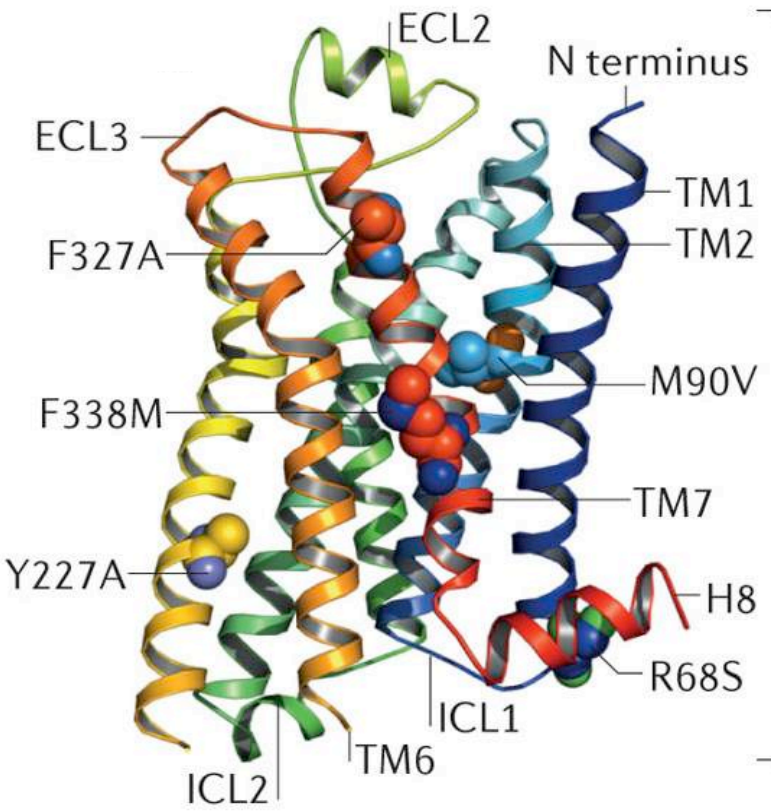
Rapid accumulation of protein structures

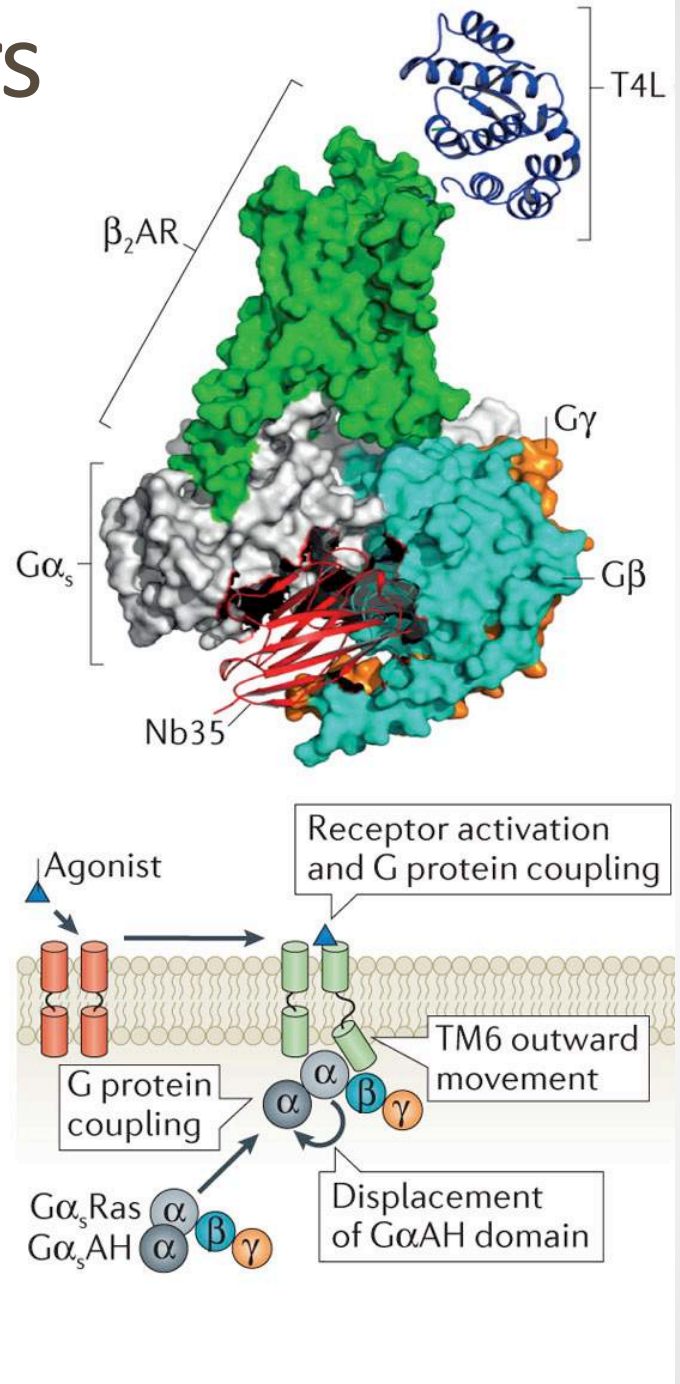


Rapid accumulation of protein structures

- Driven by interest in high-throughput crystallography
- Advances in protein structure determination methods
- Programs such as the Protein Structure Initiative
- Targeted difficult to crystallize proteins such as membrane proteins and large macromolecular assemblies

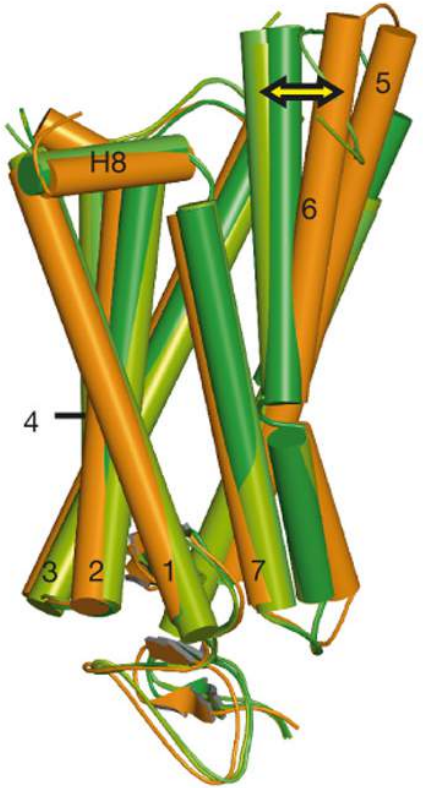
G protein-coupled receptors

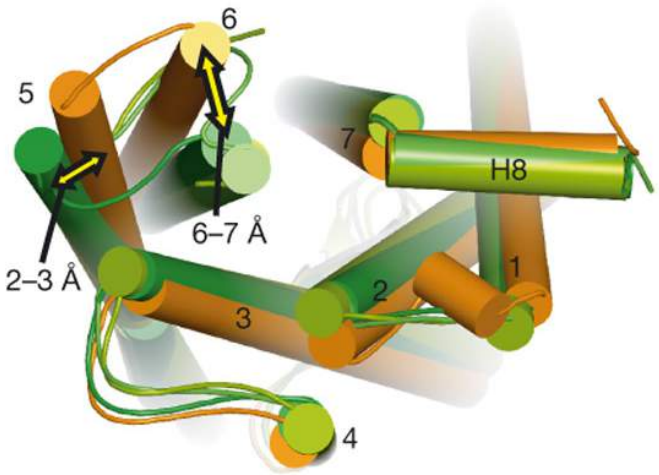


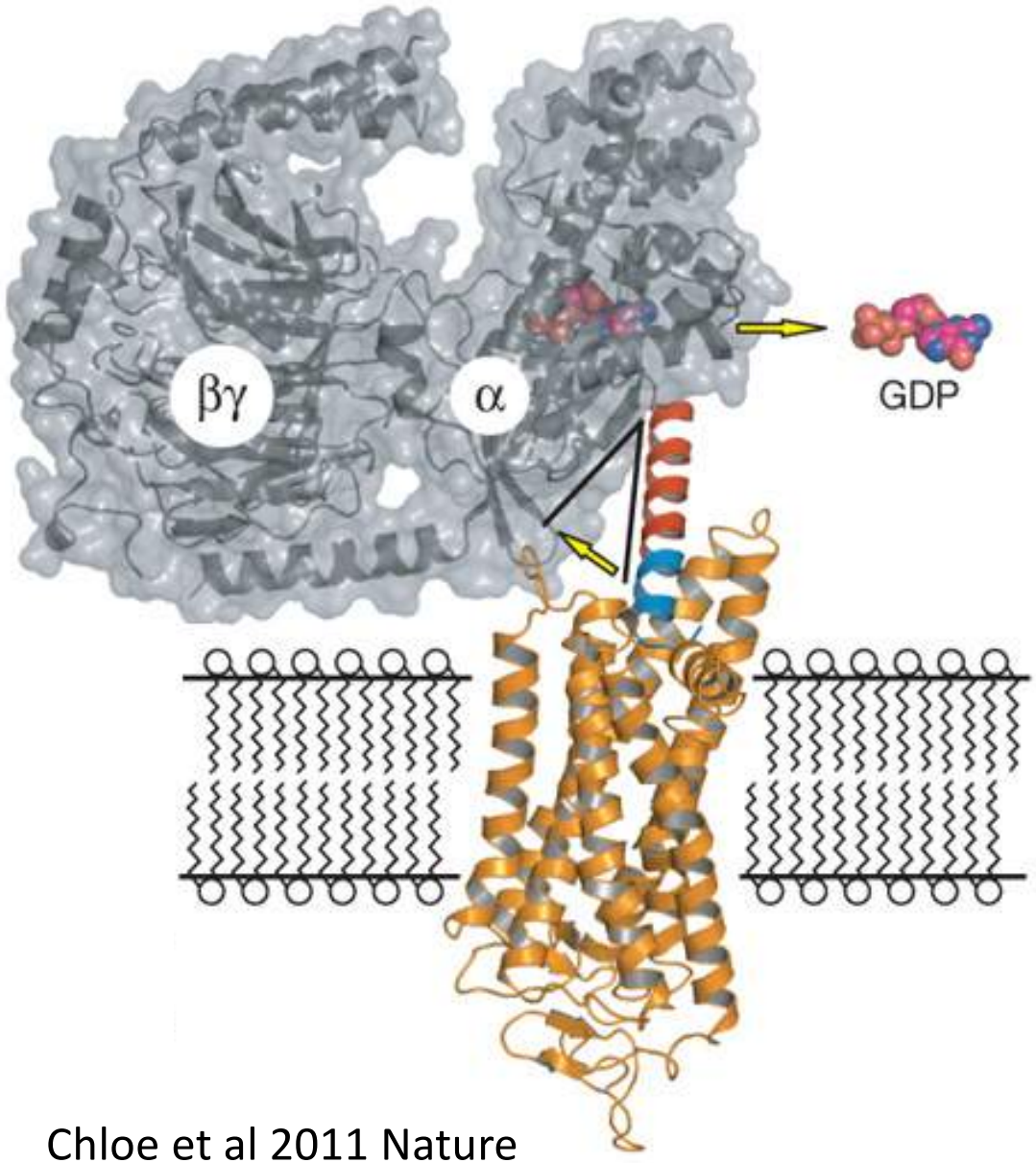


- Largest family of TM signaling proteins
- Extremely difficult to crystallize

Conformational changes upon activation

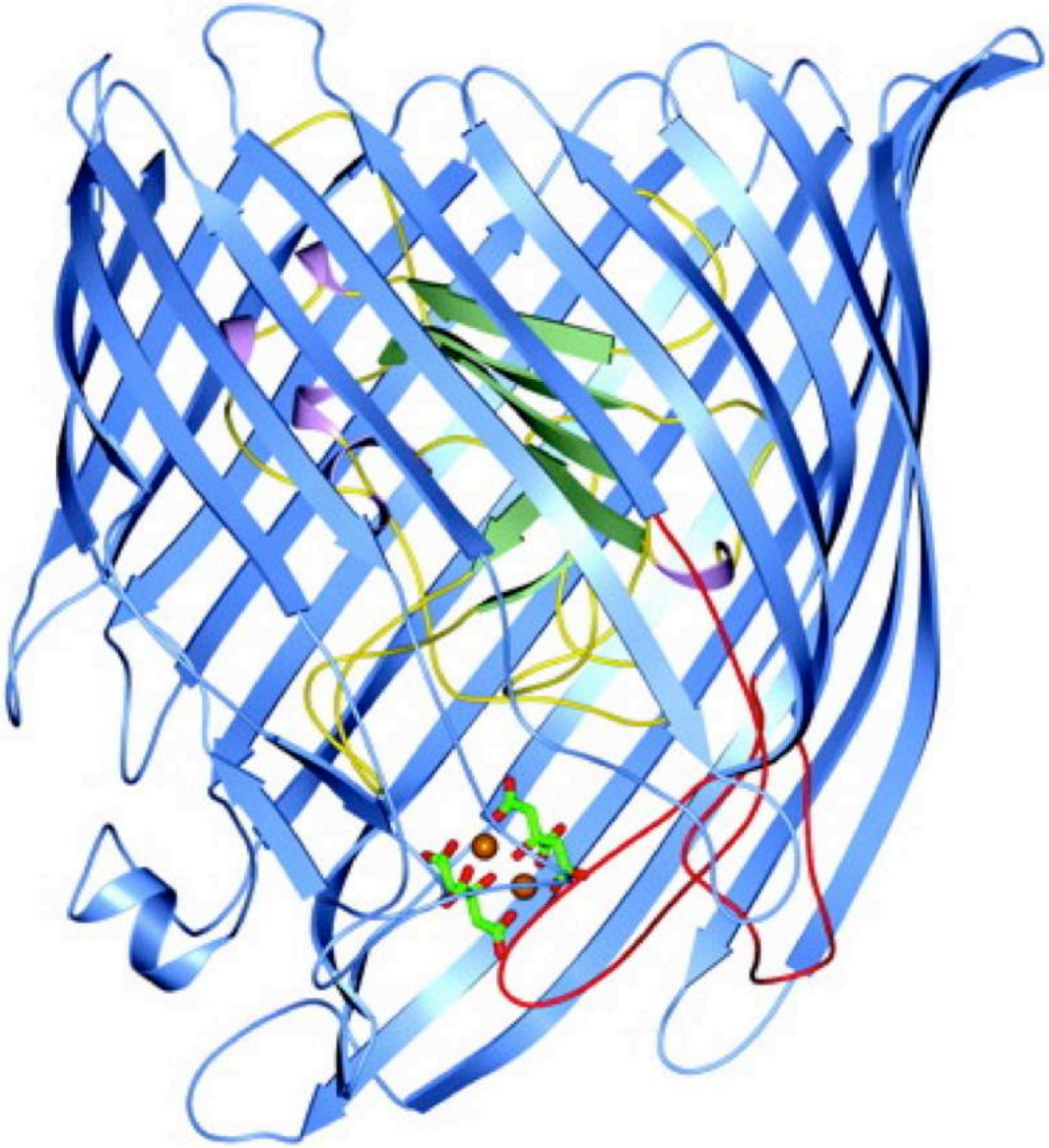




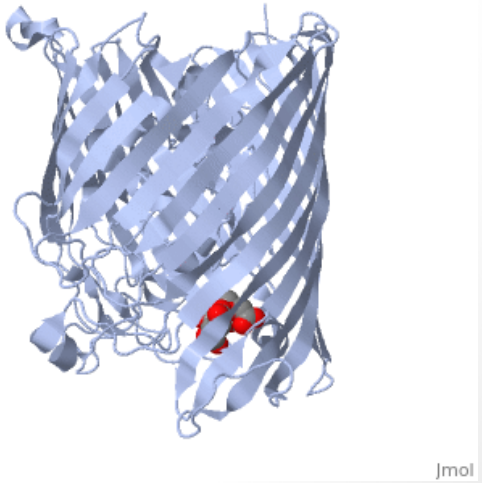


Chloe et al 2011 Nature

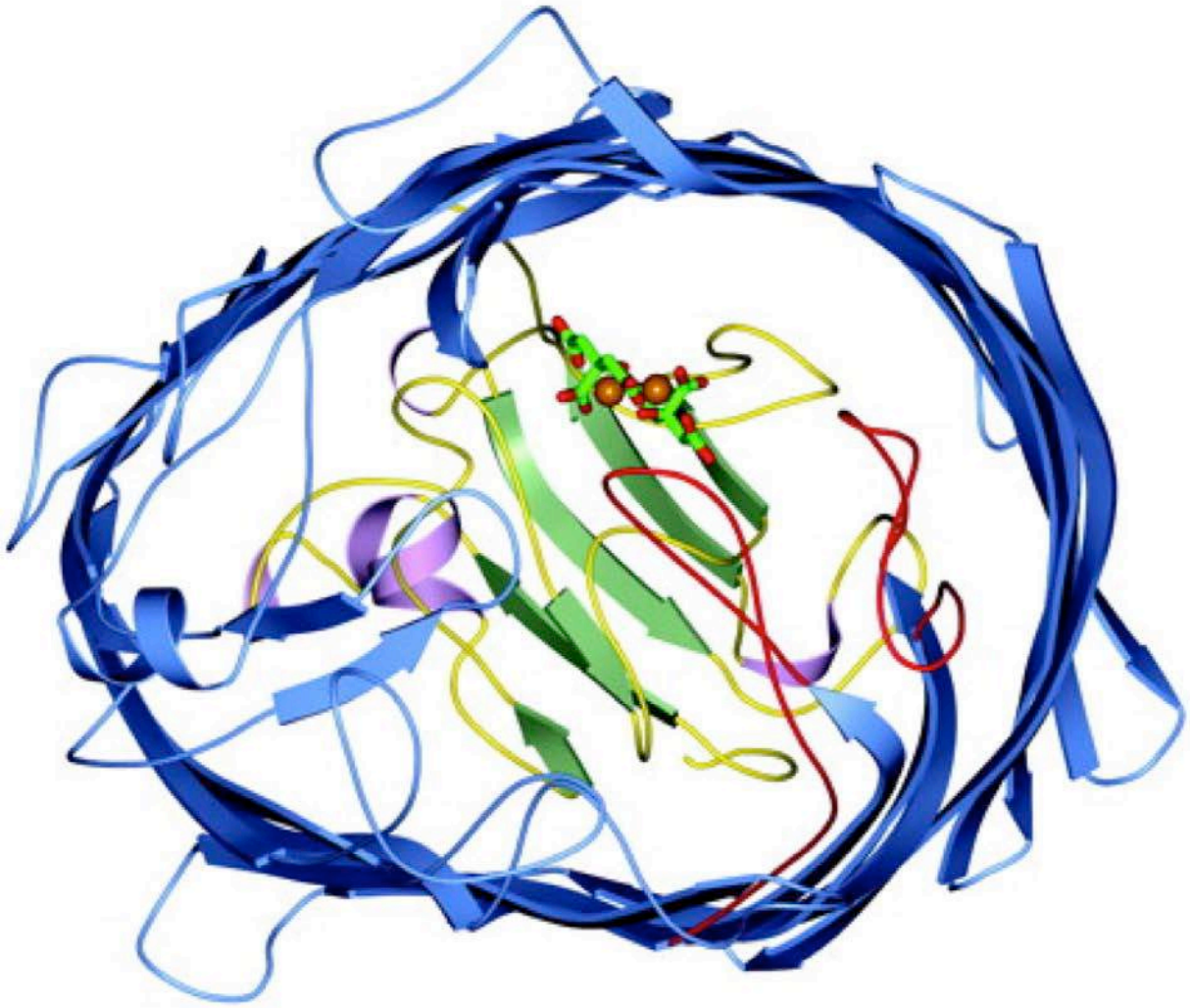
Largest monomeric TM protein: FecA



- Bacterial ion transporter
- Large transmembrane protein
- 22 beta strands

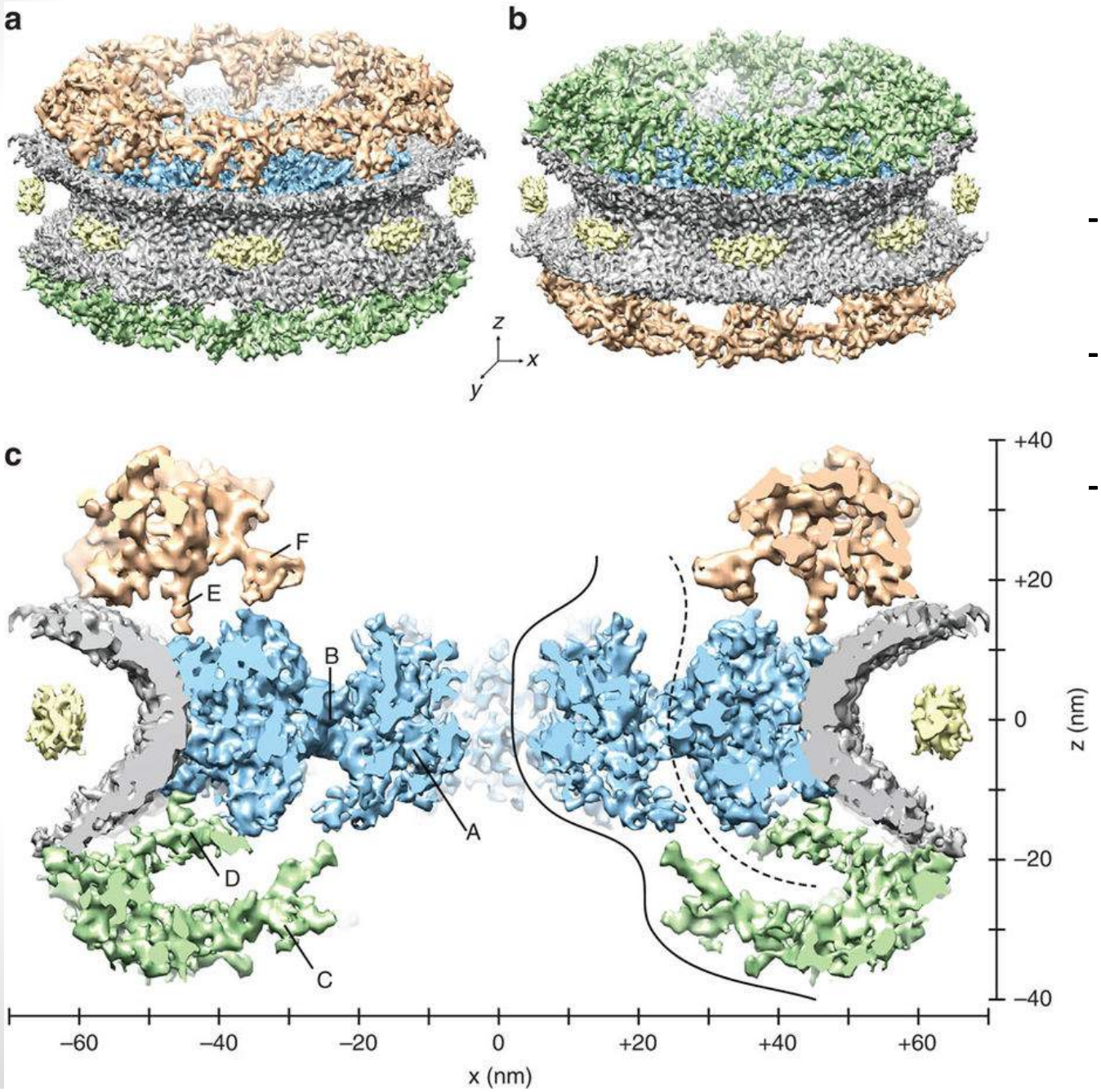


Largest monomeric TM protein: FecA



Bacterial ion transporter
Large transmembrane protein
22 beta strands

Nuclear pore complex

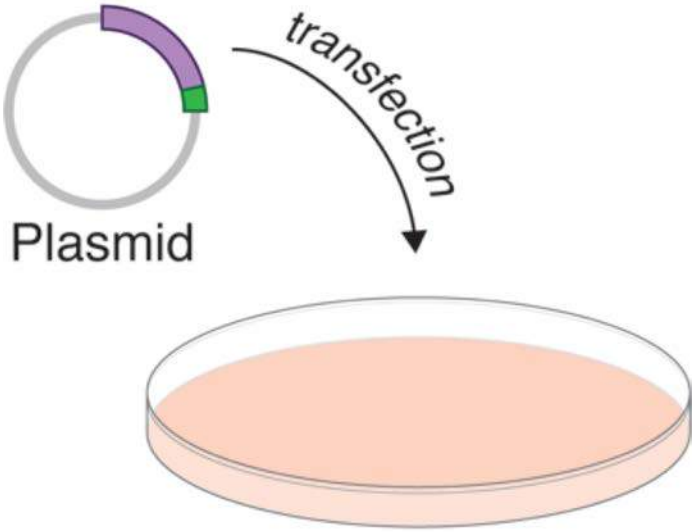


- Largest membrane bound structure
- About 30 different proteins
- Diameter of 98 nm, 50 MDa

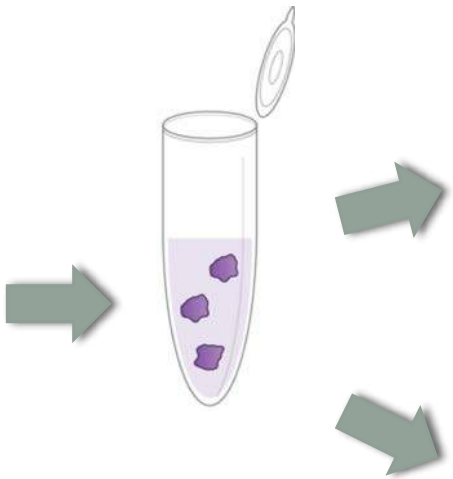
Recent advances in protein structure studies

- Difficulties of working with proteins
- Required the development of expression methods to obtain large amounts of properly folded protein
- Mostly X-ray crystallography, but also NMR, and more recently cryo-electron microscopy
- Homology modeling, molecular dynamics
- Structure predictions

In vitro protein expression methodologies

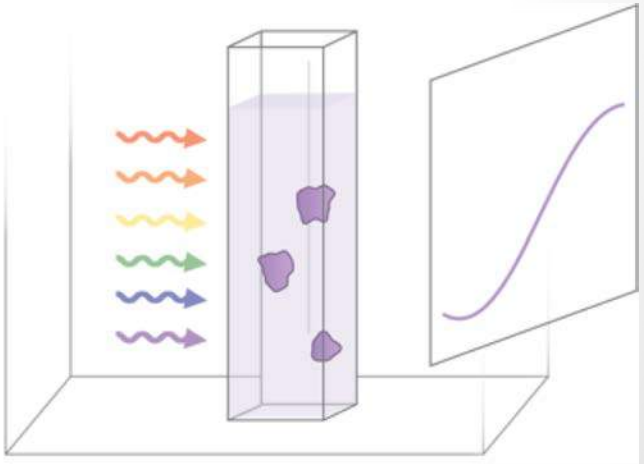


Grow cells containing gene of interest



Purify protein

Protein structure



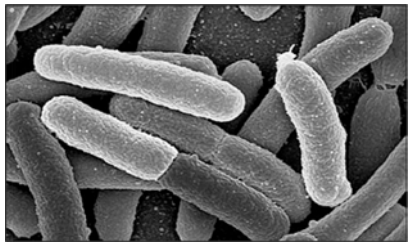
Functional assays

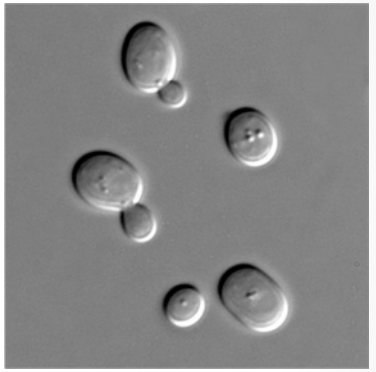
In vitro expression vs. purification from tissue

- Many proteins only present in small amounts in tissue
- Purity of sample may be an issue with complex tissues
- Purification from tissue samples does not allow for site-directed mutagenesis studies
- In vitro expression allows for testing of evolutionary hypotheses of protein structure and function

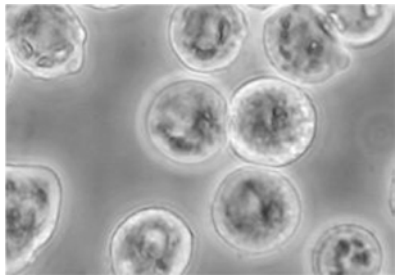
In vitro protein expression methodologies

- Bacteria: E. coli

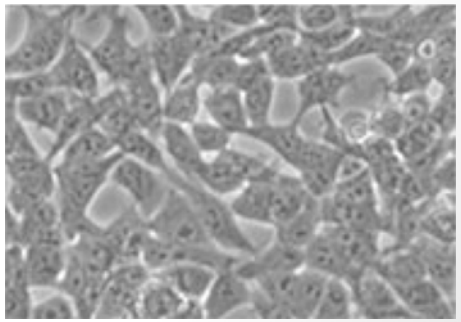




- Yeast cells: S. cerevisiae

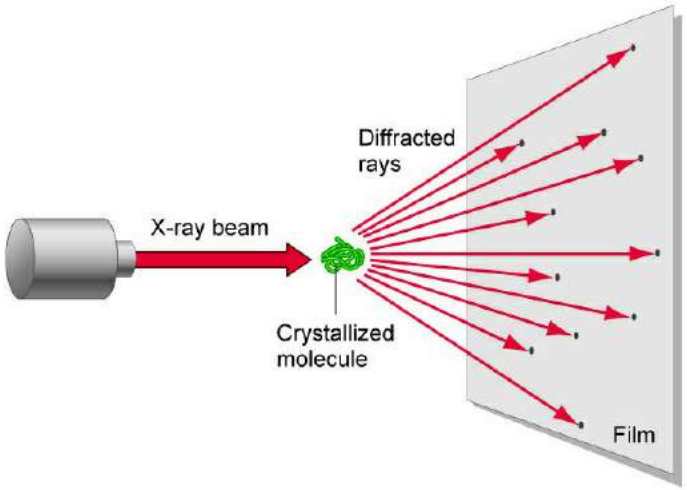


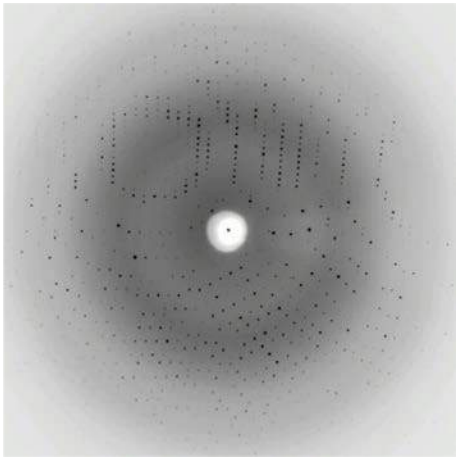
- Insect cells: SF9

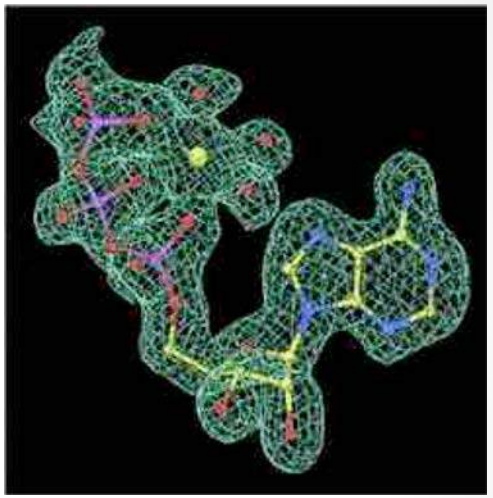


- Mammalian cell culture: HEK293

Protein structure methodologies
X-ray crystallography







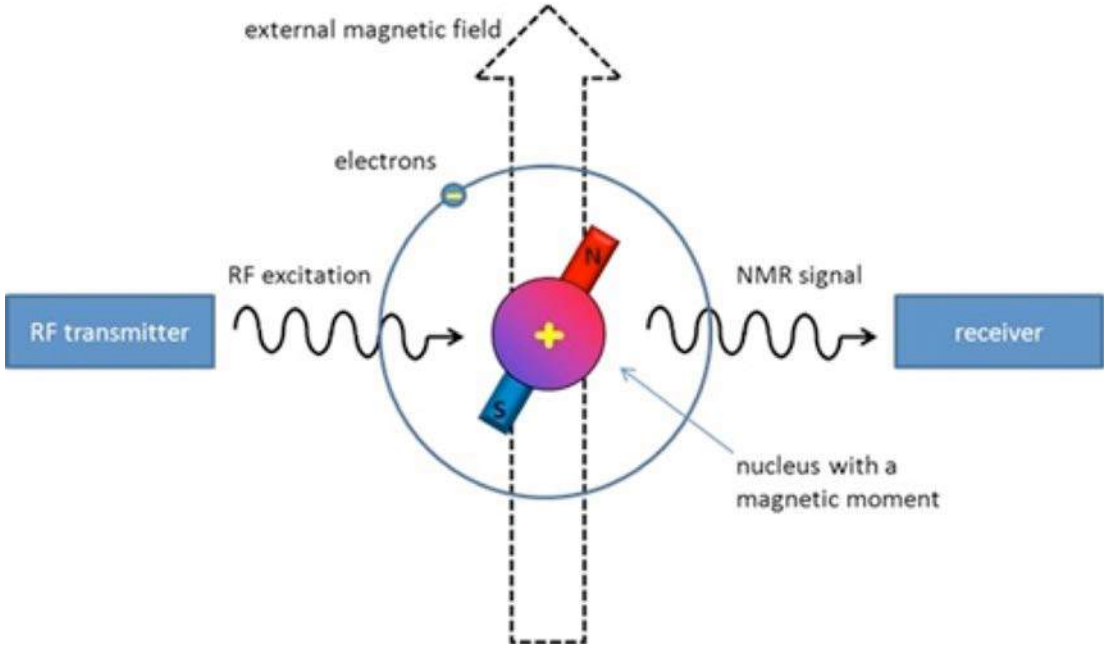
Synchotron

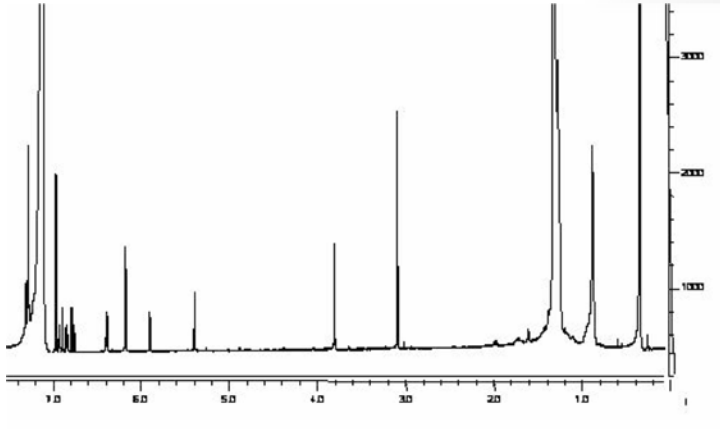
Diffraction pattern

Electron density map

- Multiple conformations, flexible regions often unresolved
- Crystallization conditions not found in nature
- Serial femtosecond crystallography

Protein structure methodologies
NMR spectroscopy

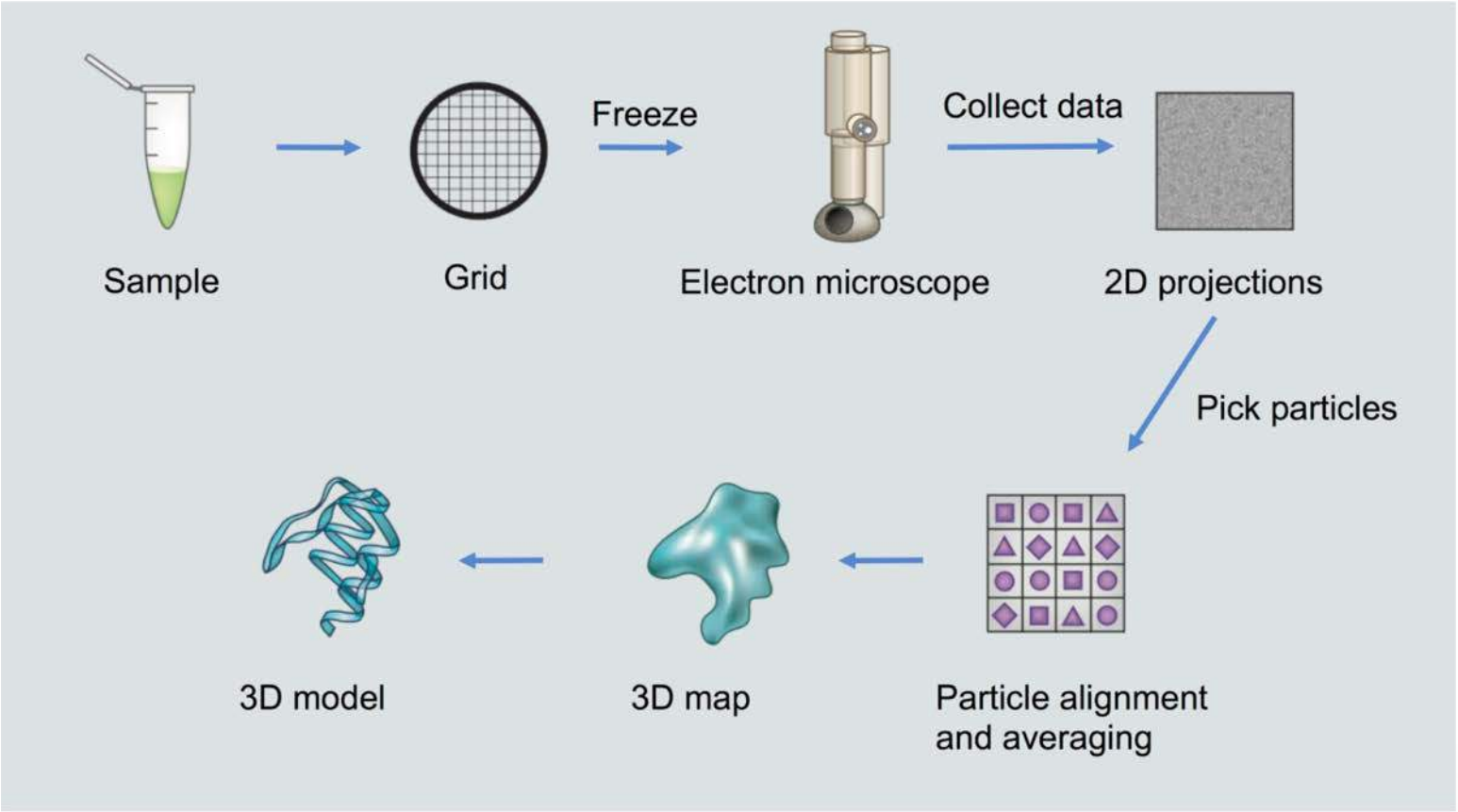




NMR spectra

- Advantage of measuring proteins in solution
- Great for studying flexible proteins
- Limited to small proteins

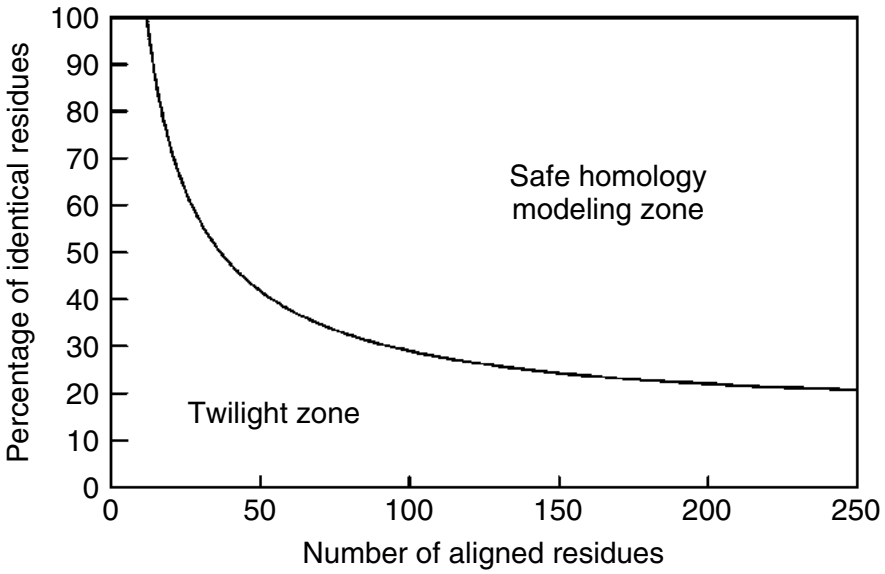
Protein structure methodologies
Cryo-electron microscopy



- Advances in direct detection, sample prep, and instrumentation have achieved high resolution for larger protein complexes
- This technique offers high resolution of larger proteins in a native state
- Requires highly specialized instrumentation

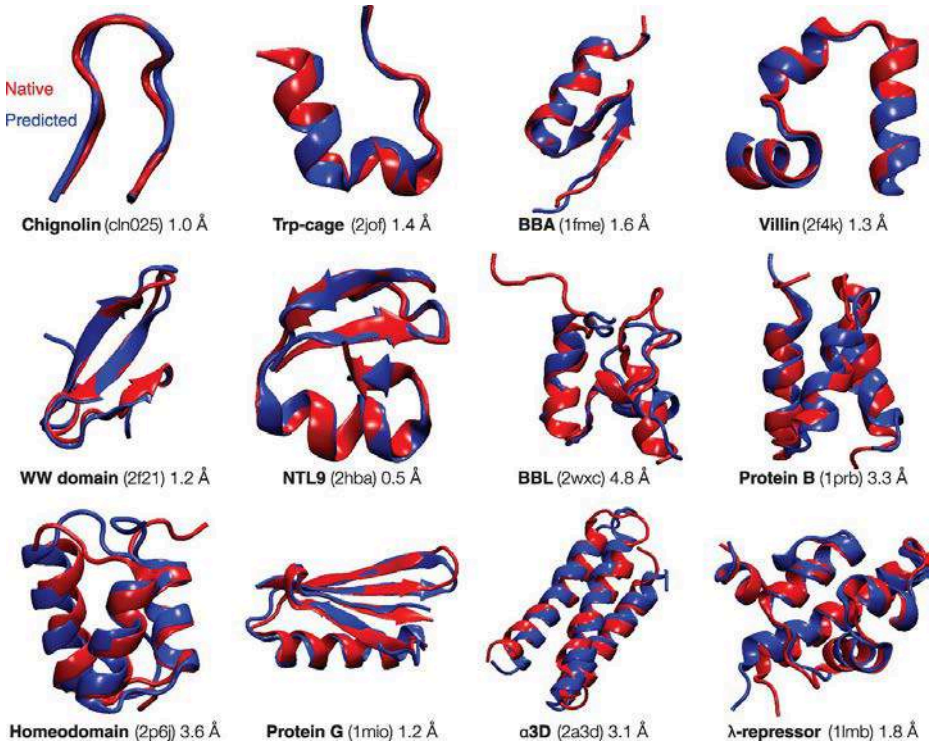
What about protein structure prediction?

Where the twilight zone is may depend on your of view

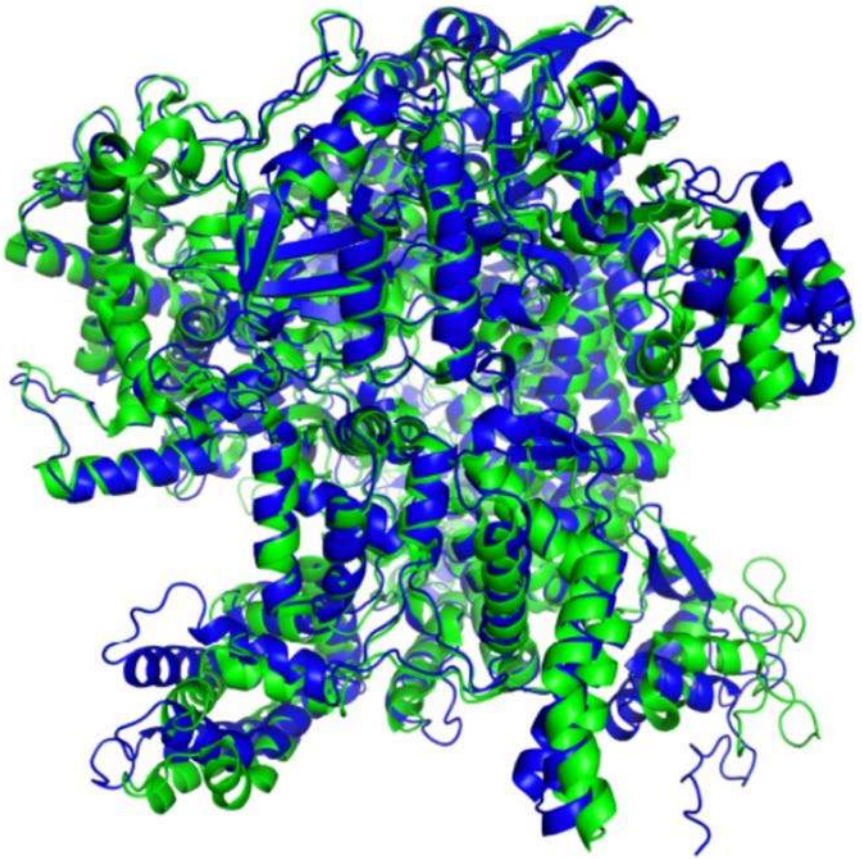


Krieger et al. 2003

The twilight zone for ab initio protein folding predictions



Dill & MacCallum, Science 2012



Jumper et al, Nature 2021

Protein structure prediction

- Homology modeling
  - MODELLER (https://salilab.org/modeller/)
  - Rosetta suite (http://robetta.bakerlab.org/)
  - SWISS-MODEL (https://swissmodel.expasy.org/)
- Molecular dynamics simulations
- Machine learning (AlphaFold)

Molecular evolution:
Evolution of protein function

DATA

- Genomic sequencing
- Protein structures

TOOLS

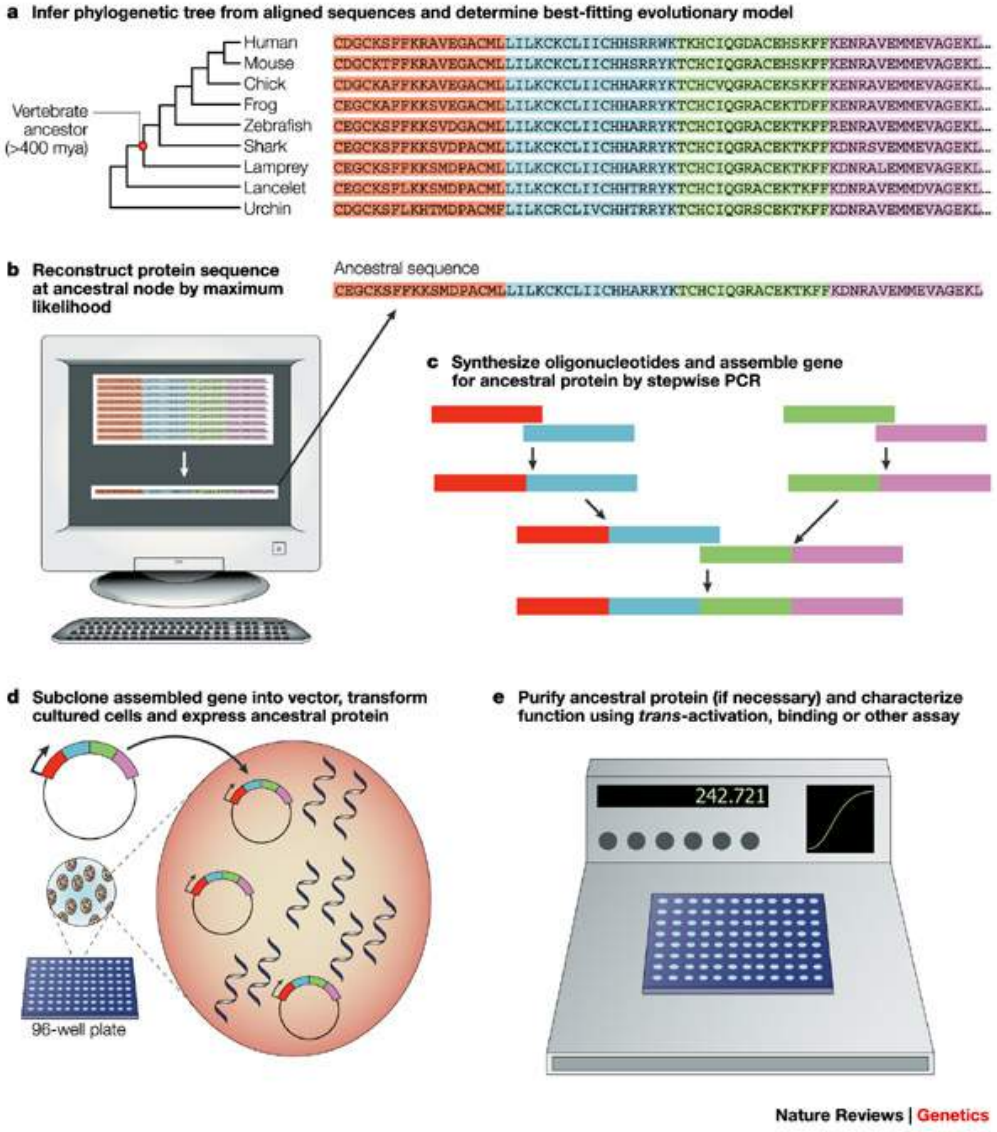
- Phylogenetic models of coding sequence evolution
- Experimental studies, mutagenesis and ancestral resurrection

Phylogenetic approaches to the study of protein structure and function

- Ancestral protein reconstruction
- Computational analyses of selection (dN/dS)

-> Combining computational with experimental approaches allows us to test hypotheses of selection in protein evolution

Resurrecting ancestral proteins



Thornton, 2004
Nat. Reviews Genet. (5):366

Ancestral reconstruction: considerations

- Most studies use ML/Bayes methods to infer ancestral sequence with highest probability, single point estimate
- Violations of model assumptions, e.g. shifts in equilibrium frequencies
- Uncertainty in tree topology
- Statistical bias towards states with highest equilibrium frequencies
- This may also result in functional bias towards more stable proteins (Goldstein et al. 2013)

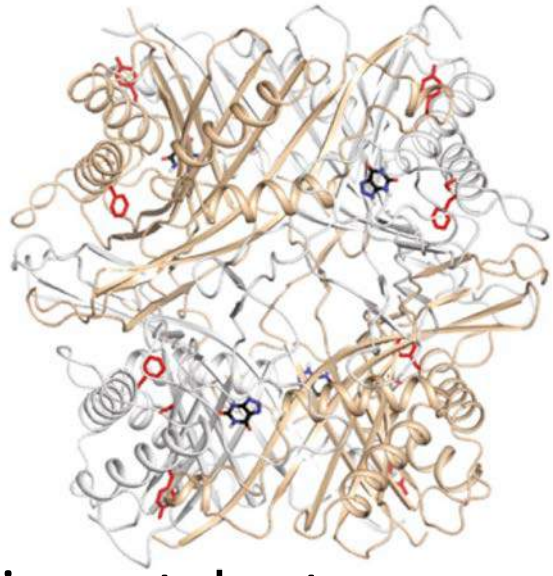
-> How to assess robustness of reconstruction in a functional context?

Assessing robustness of reconstruction in a functional context

- Alternate tree topologies, species tree topology
- Alternate approaches, models of evolution
- Sampling alternate ancestors from the posterior distribution (Pollack & Chang 2012)
- Sampling of near ancestor sequences (Bar-Rogovsky et al. 2015)
- Uncertainty in genotype does not necessarily reflect uncertainty in phenotype (Gaucher et al. 2008)

-> Need for experimental data to inform effects of uncertainty in reconstruction on function

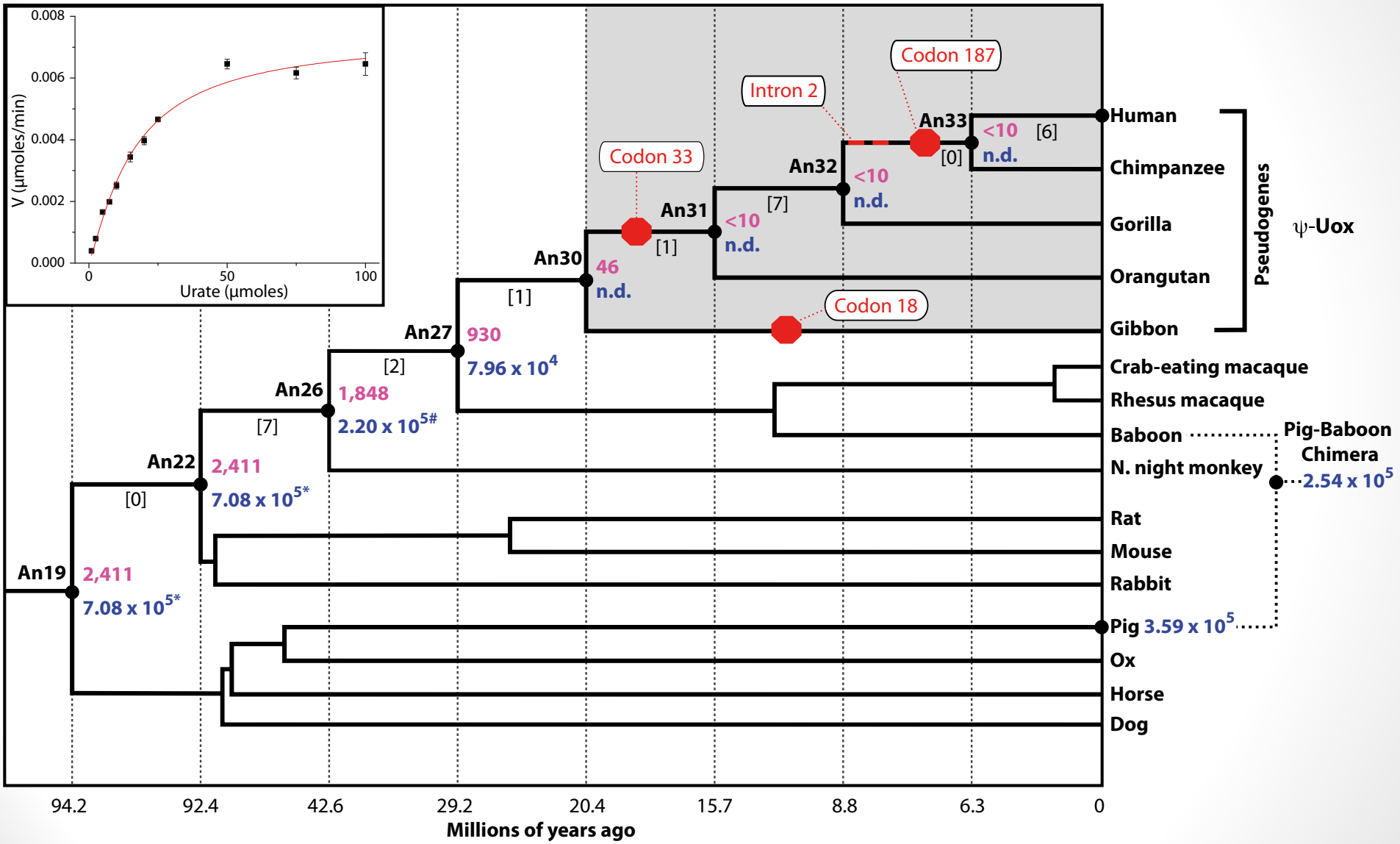
Uricase evolution in primates



- Key enzyme metabolizing uric acid in vertebrates
- Lost in some primates, including humans
- Prevalence of diseases such as gout, hypertension, obesity, cardiovascular disease
- Uricase knockouts in mice result in mortality in first 4 weeks

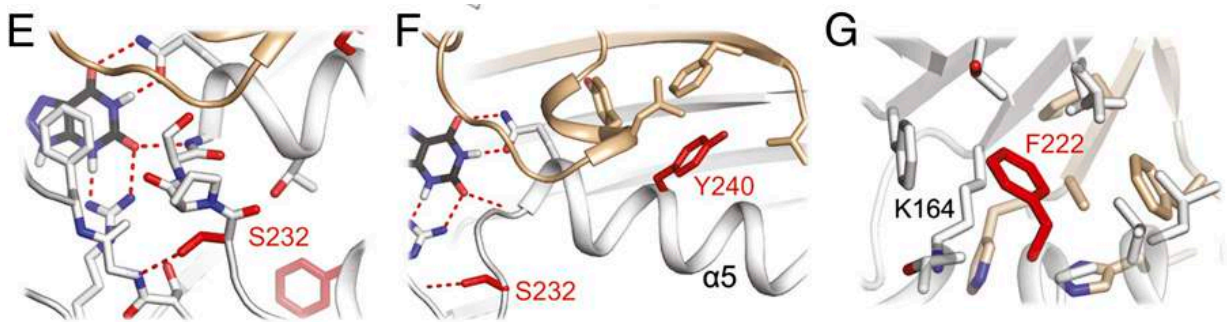
-> Kratzer *et al.* 2014 (PNAS) used experimentally recreated ancient uricases to determine exactly when, and how, uricase function was lost in primates.

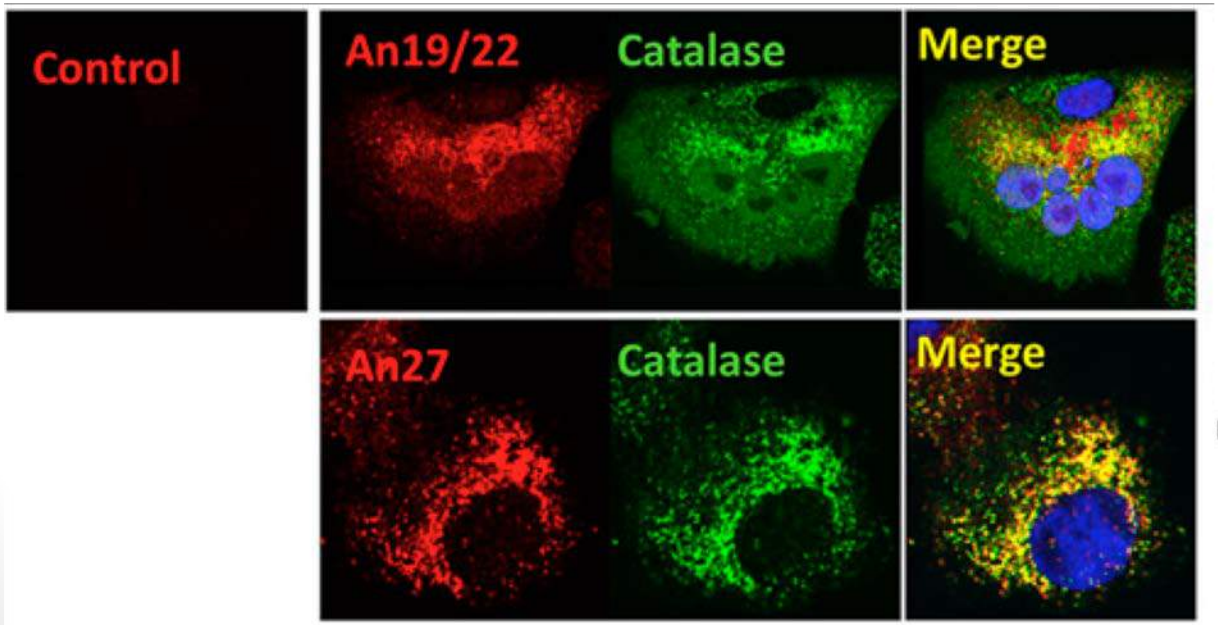
Uricase evolution in primates

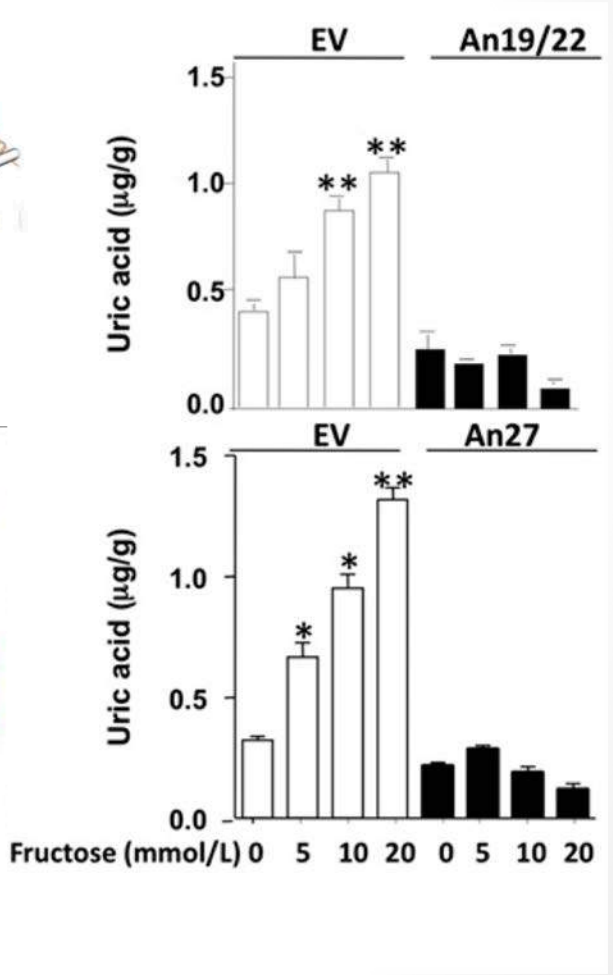


Kratzer et al., 2014, PNAS

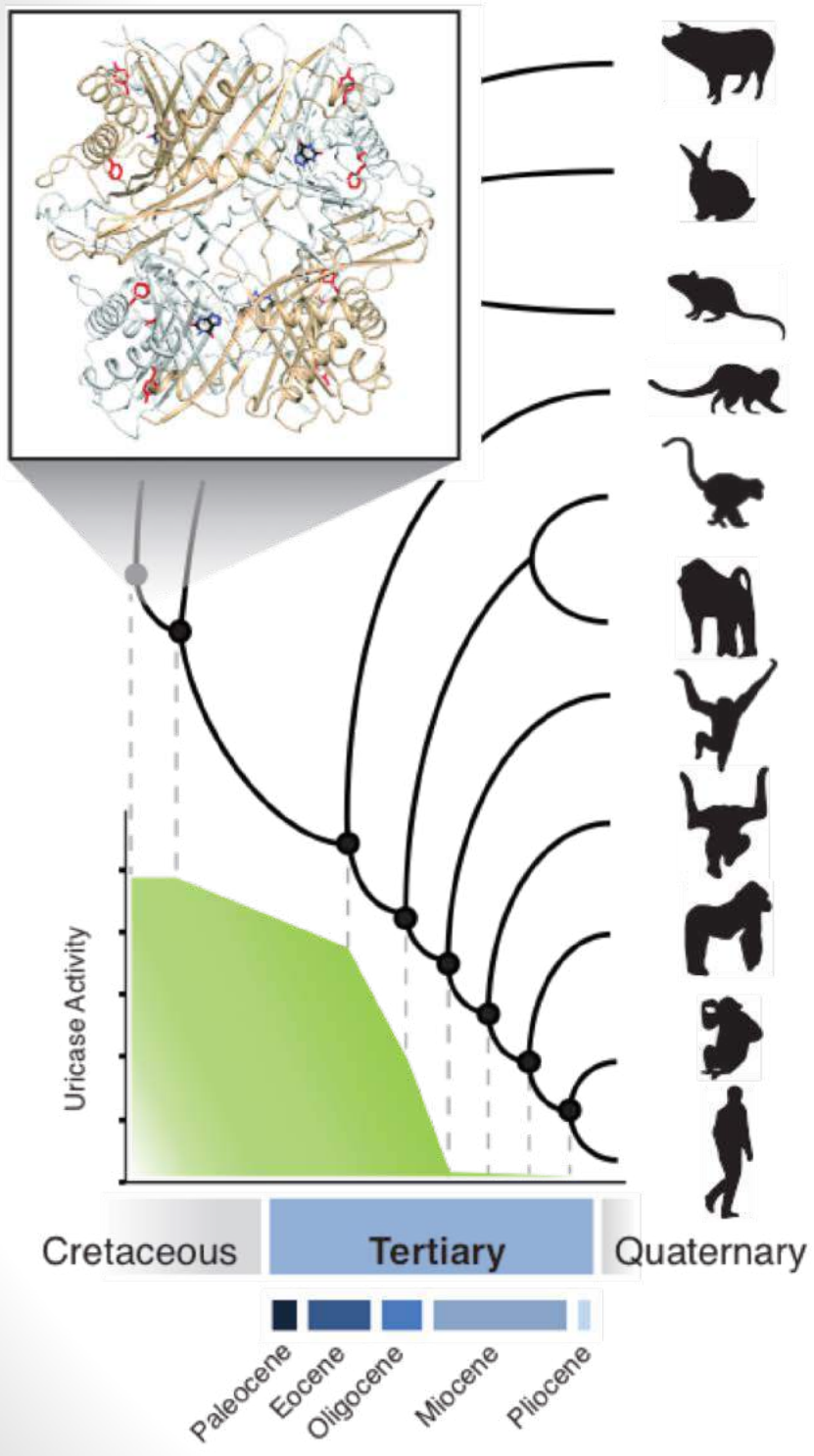
Uricase evolution in primates







Kratzer et al., 2014, PNAS



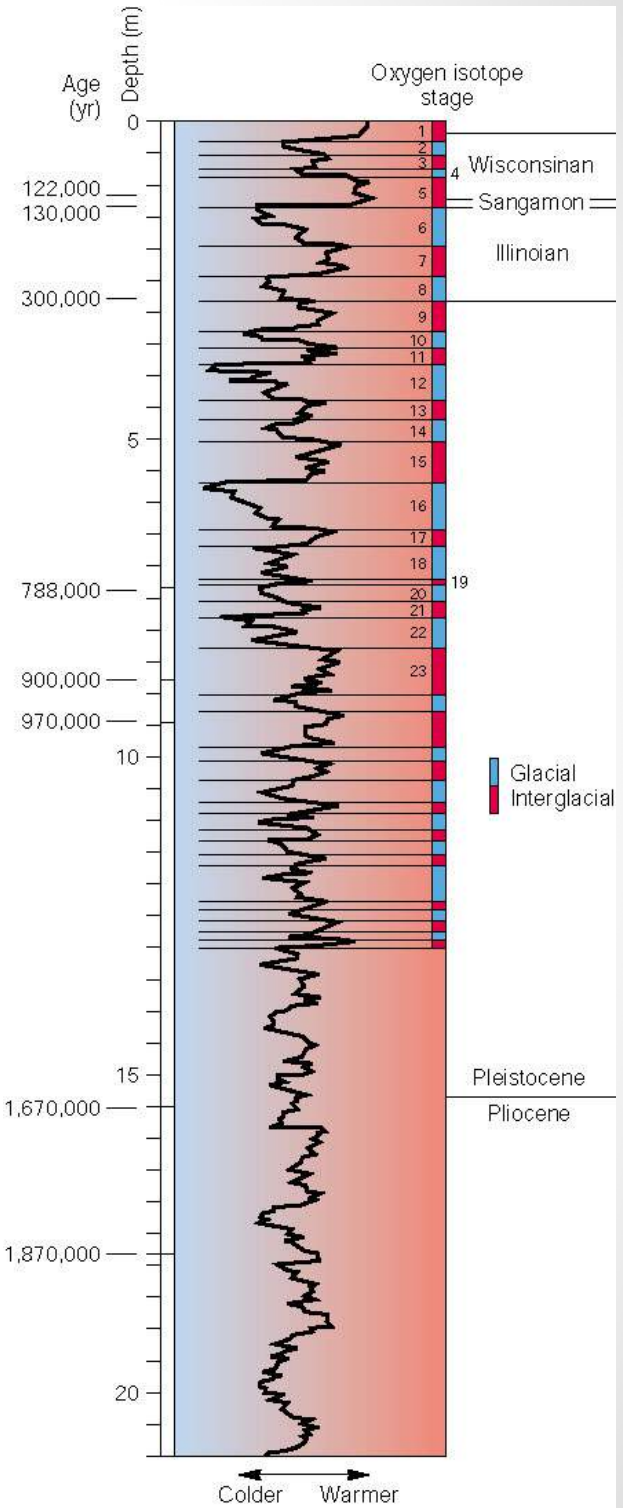
Uricase evolution in primates

Why was uricase lost in the great apes?

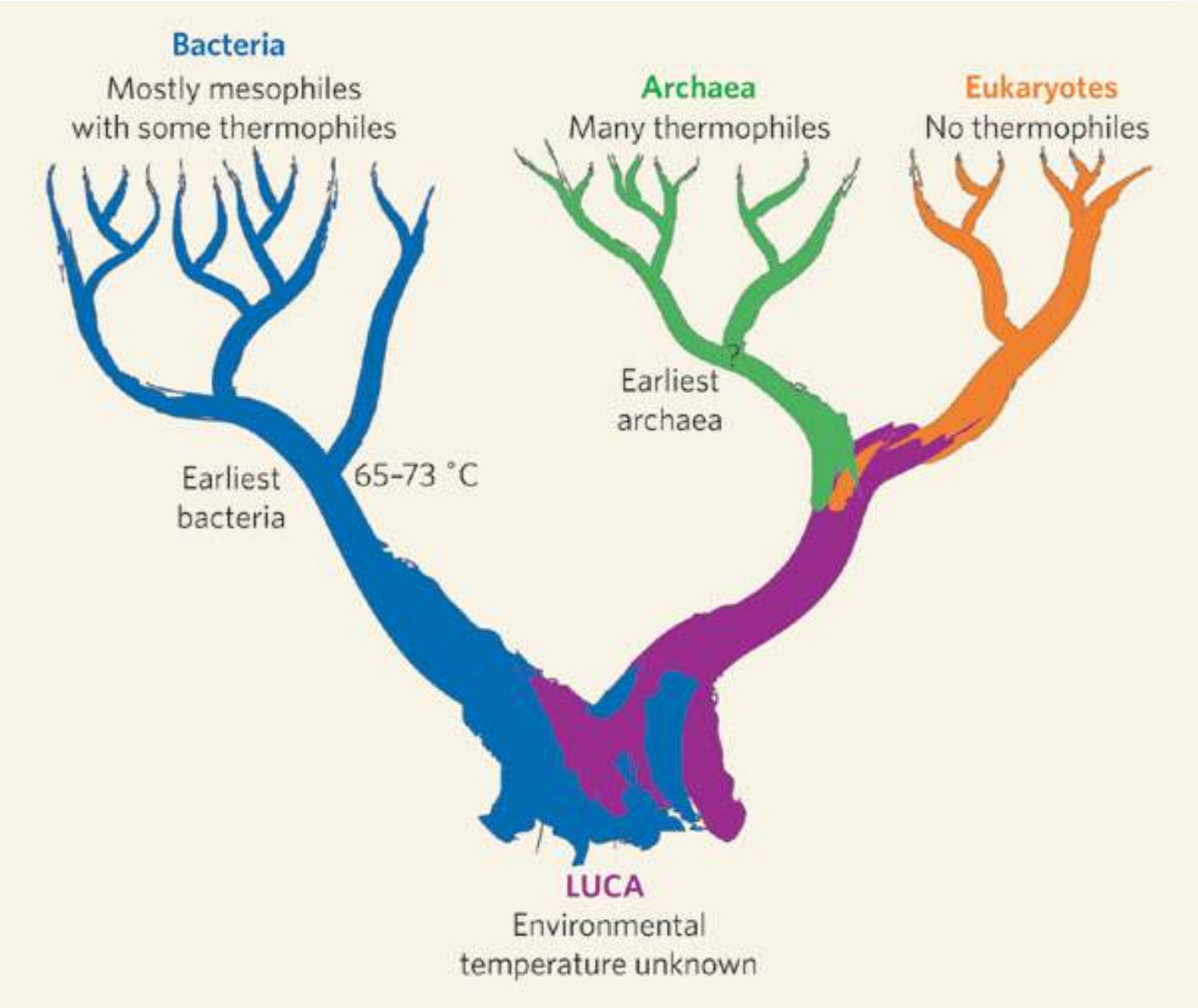
Thrifty genes vs. "drifty" genes

Kratzer et al., 2014, PNAS

Oceanic paleoenvironments

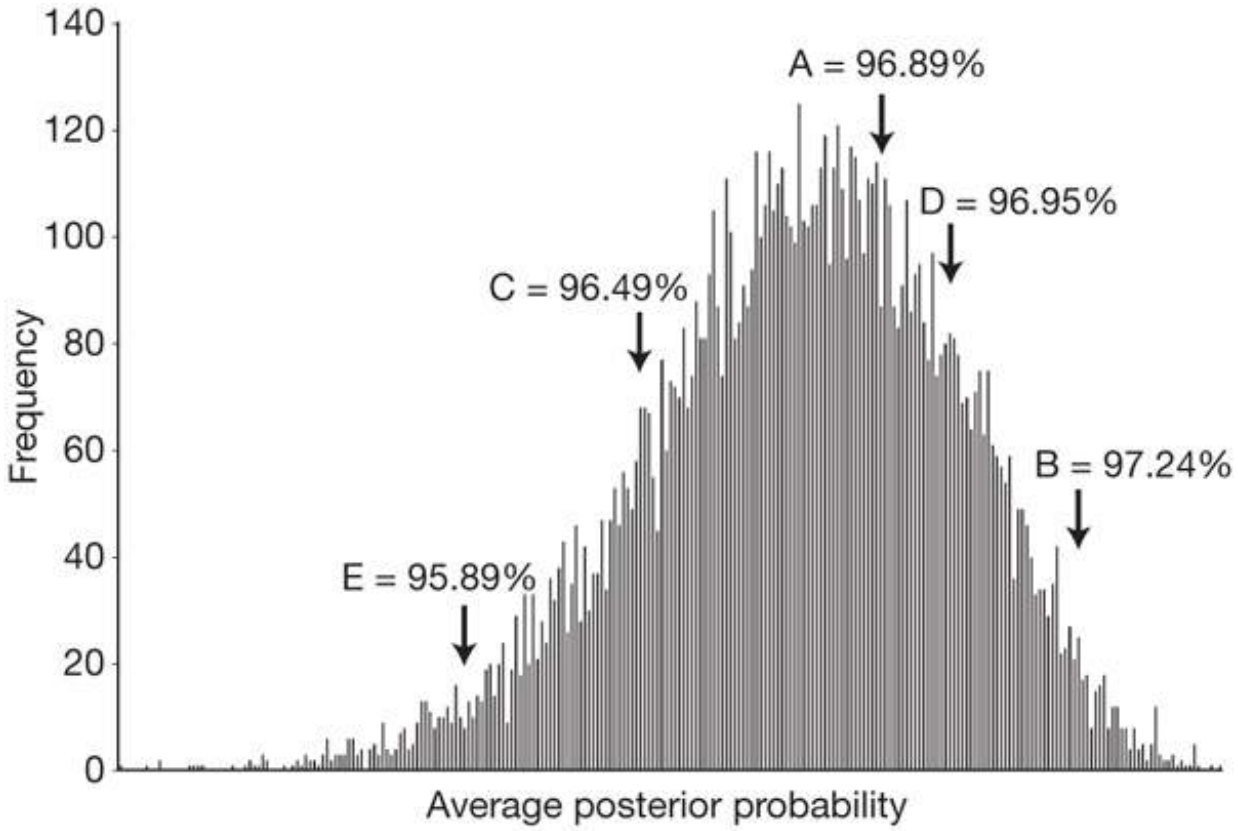


Paleoenvironments (EF-Tu)



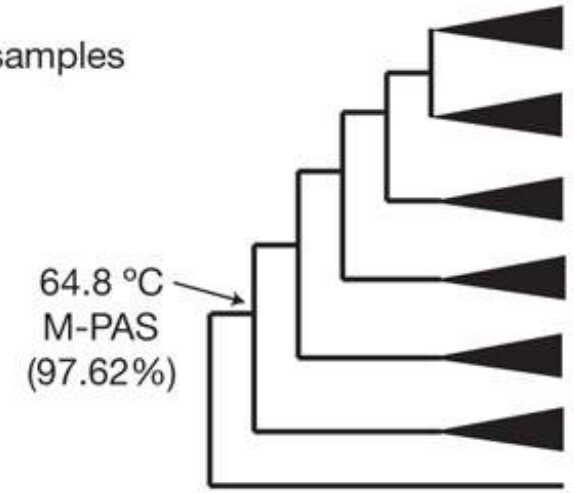
Paleoenvironments (EF-Tu)

- Resurrected proteins can provide clues about the temps at which ancient organisms lived
- EF-Tu, an elongation factor crucial for protein synthesis in all cells throughout evolutionary history
- Present day organisms have EF-Tu's which are highly correlated to temp at which organisms live
- Express resurrected gene in E coli, measure thermostability (T_m) of proteins using CD
- Bacterial ancestors appear to be thermophilic (60-80 deg C)

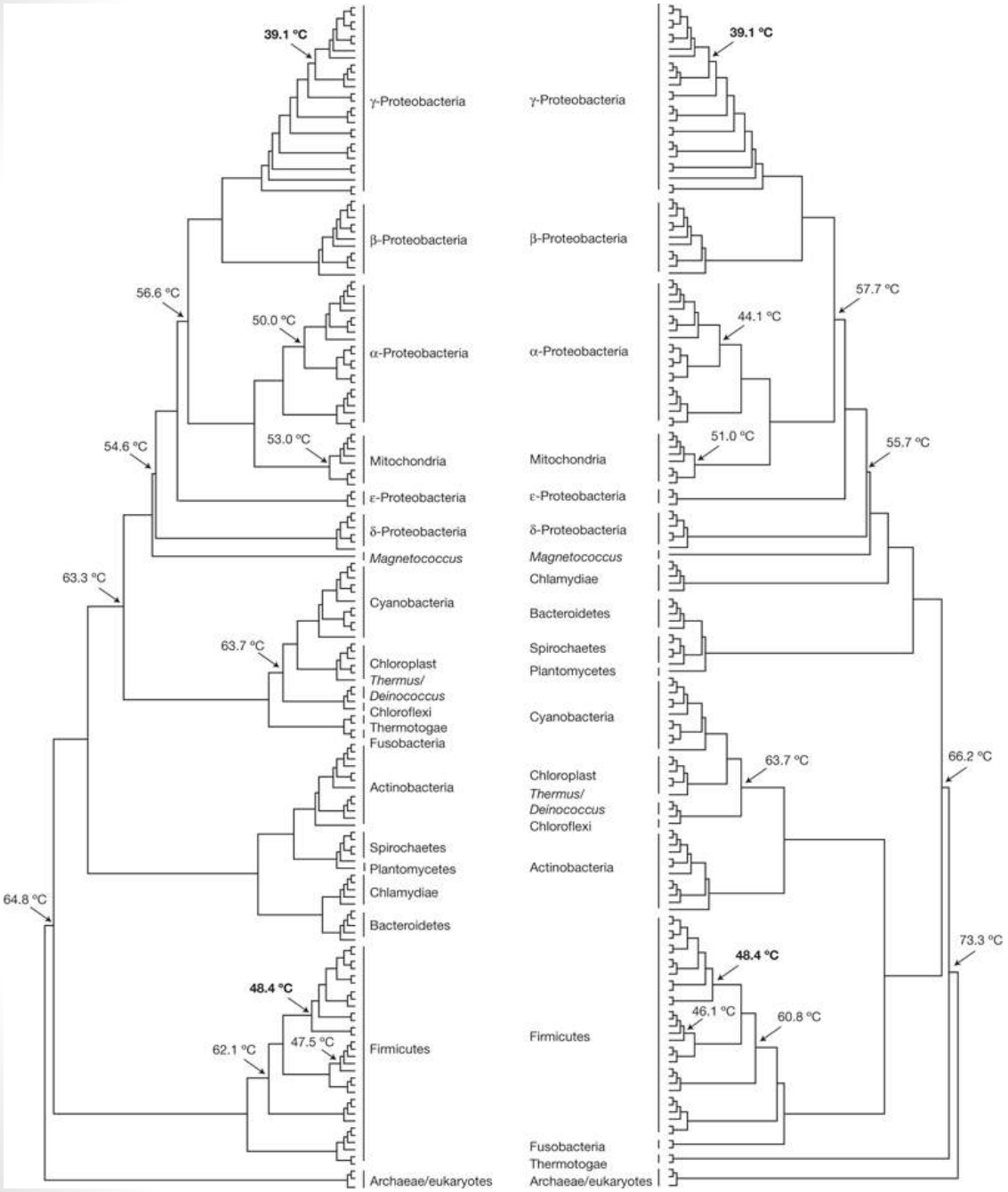


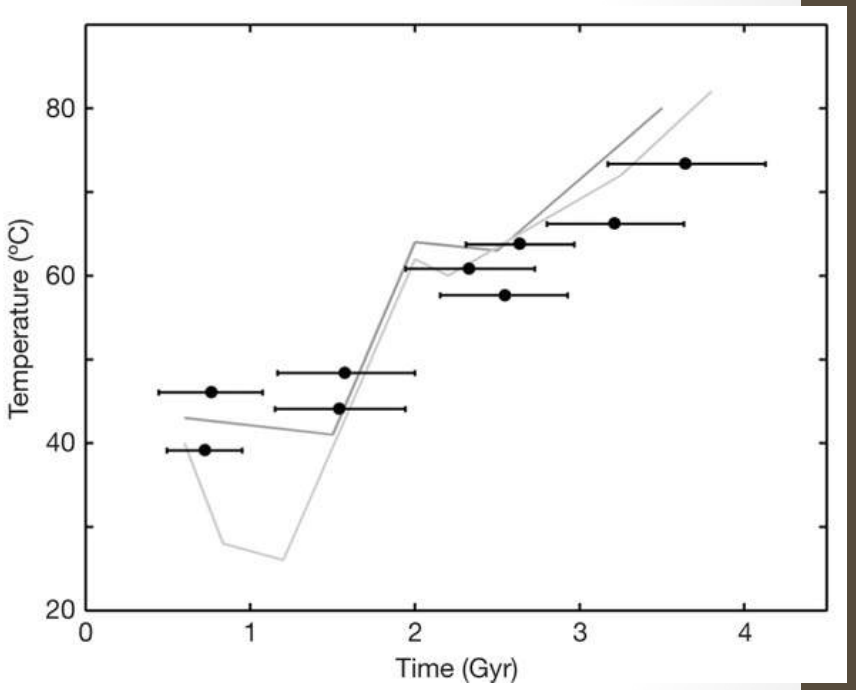
T_m values for weighted random samples from the posterior distribution:
A = 66.3 °C
B = 62.2 °C
C = 64.6 °C
D = 60.5 °C
E = 60.0 °C

Ancestral π_{eq} T_m = 61.4 °C



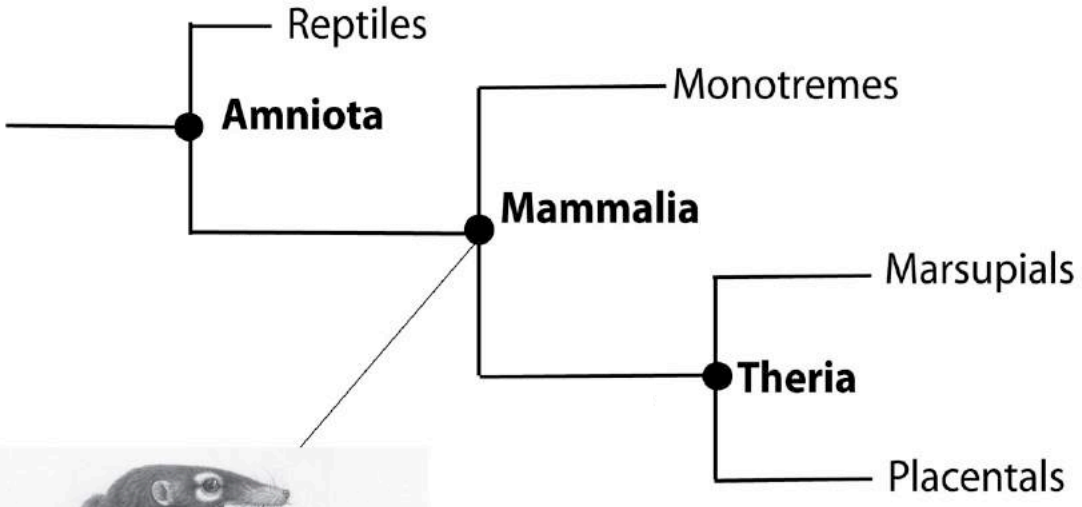
Gaucher et al., 2008
Nature (451):704

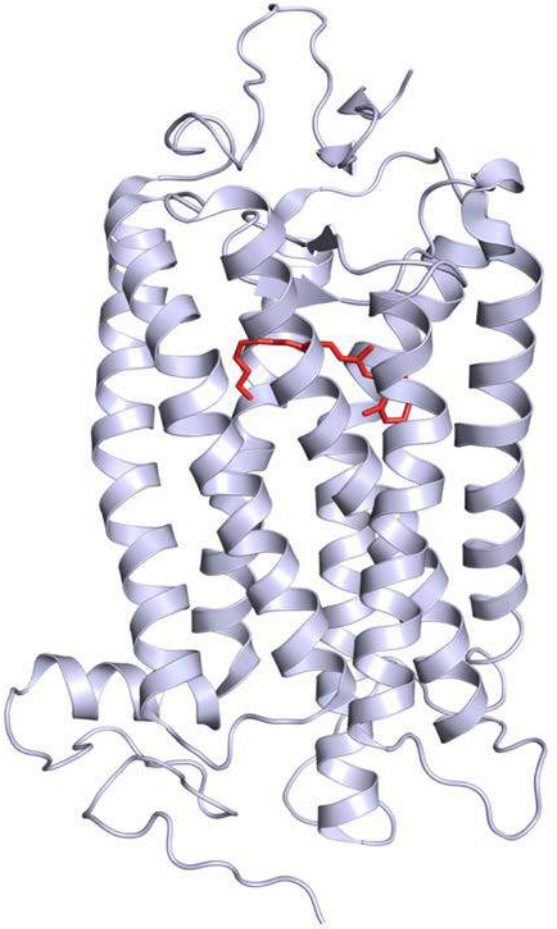


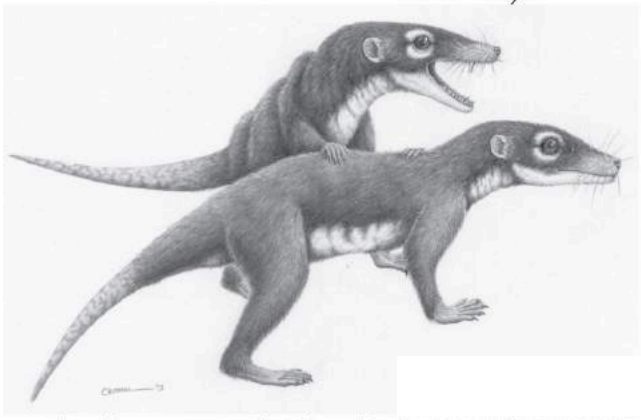


Gaucher et al. 2008 Nature

Rhodopsin evolution: Nocturnality of early mammals?







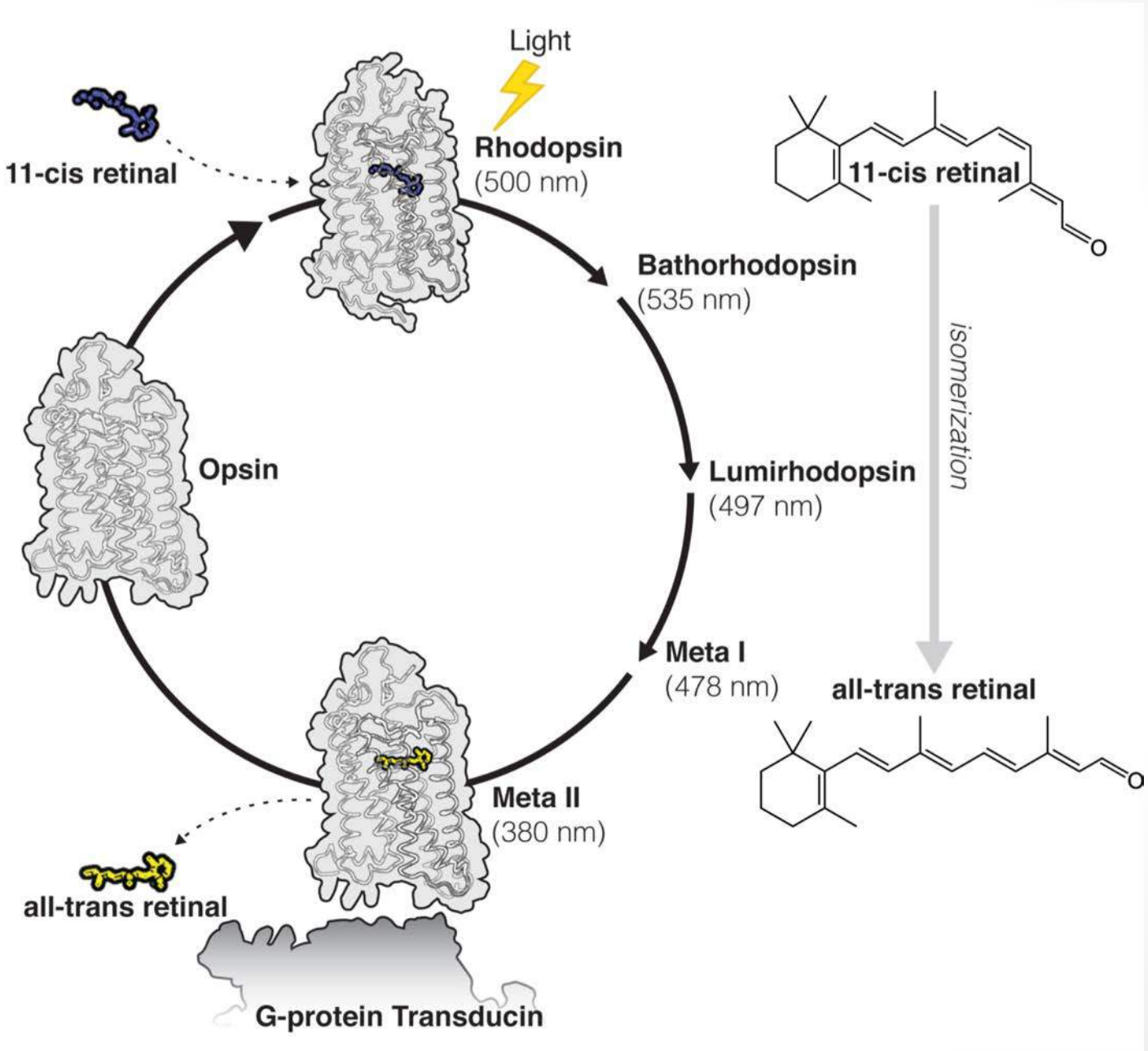
http://www.seirim.net/Cont/Draw/Vert/Morganucodon.jpg

Previous Hypothesis about early mammals:

1. Living in **Nocturnal Niche**
(Crompton, Taylor and Jagger 1978 Nature)

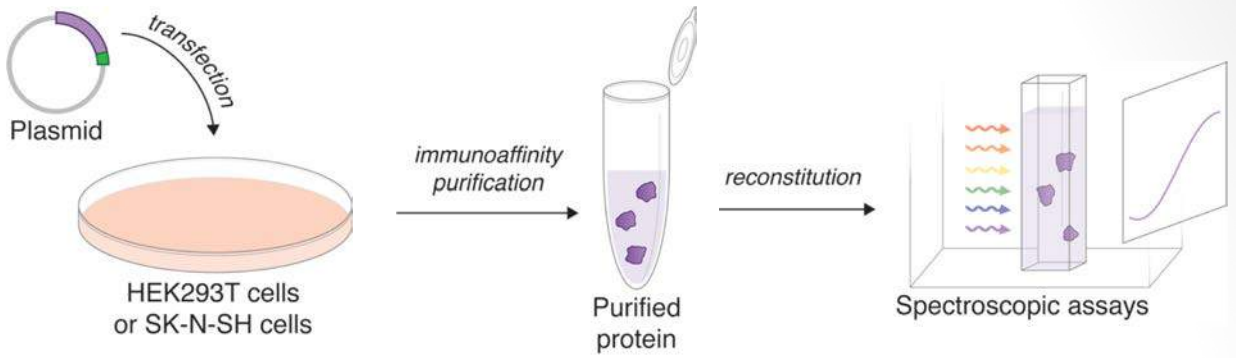
2. Adaptive Changes In Rod Photoreceptors to improve dim-light vision
(Walls 1942 ; Ahnelt and Kolb 2000 Prog. Retinal and Eye Res.)

Visual cycle: conformational changes in rhodopsin

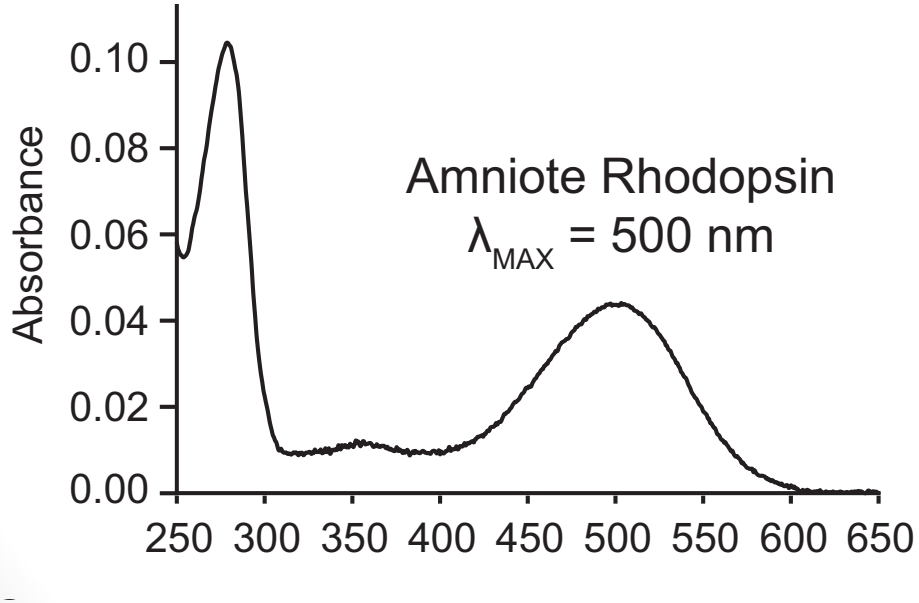


Spectroscopic assays of rhodopsin function

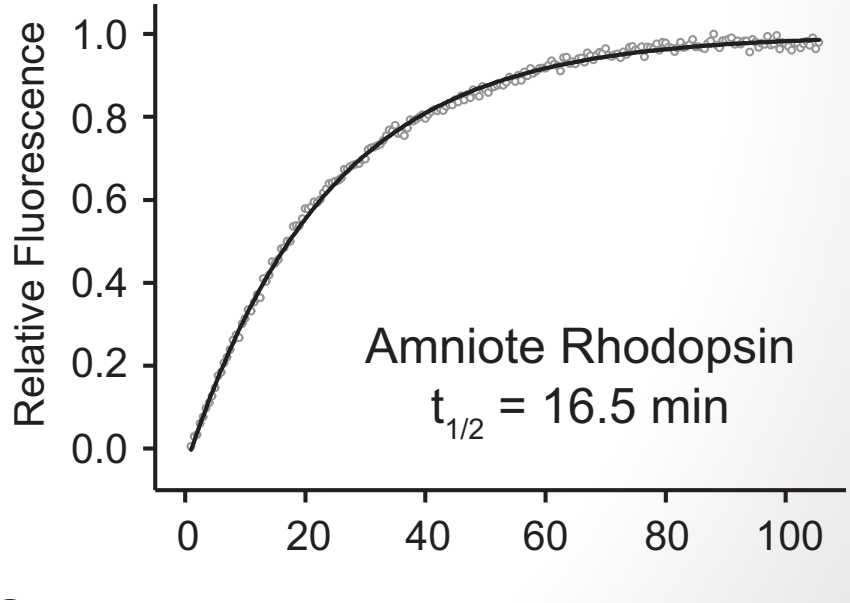
In vitro expression
& purification



Rhodopsin spectral tuning

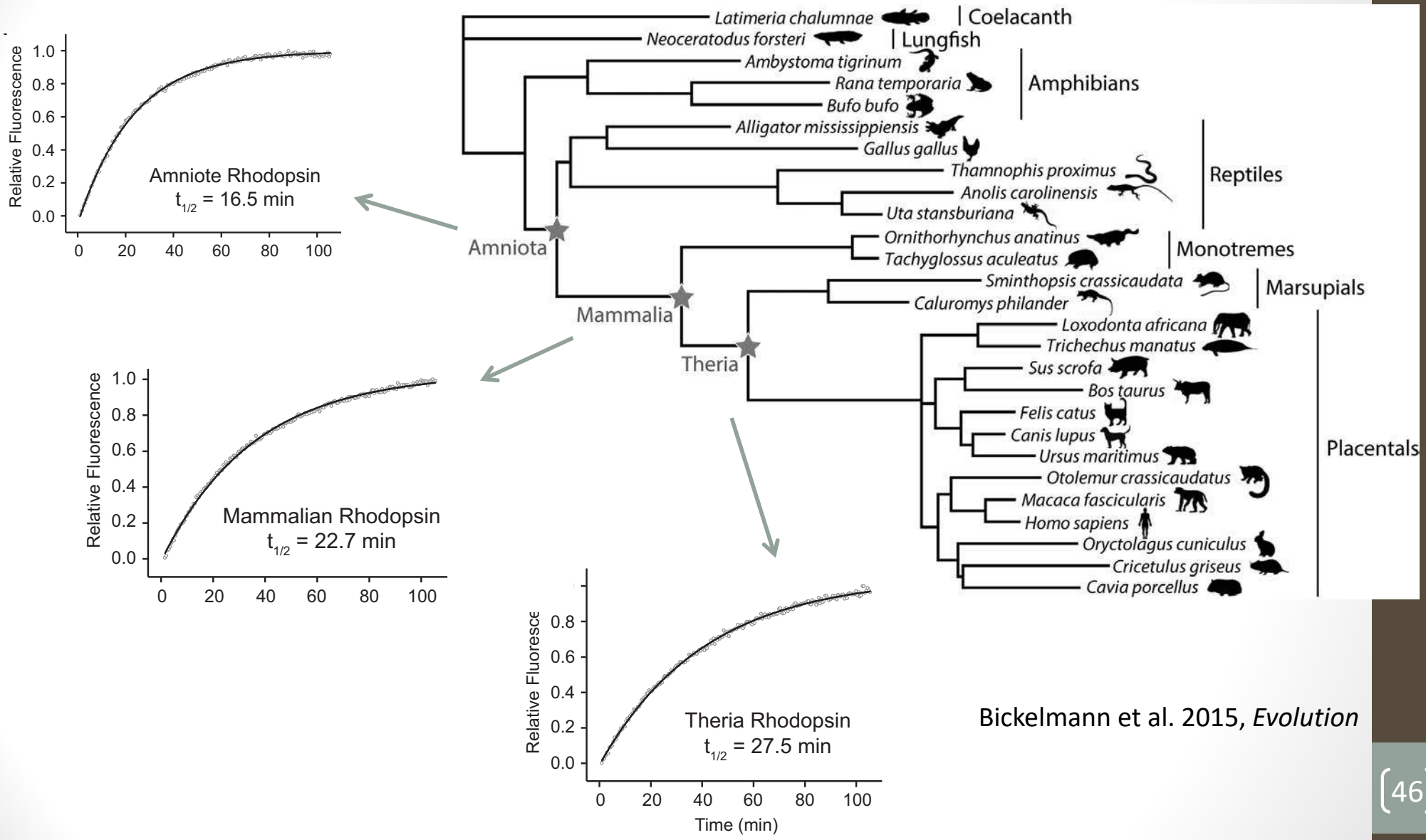


Lifetime of activated state



Bickelmann et al. 2015, *Evolution*

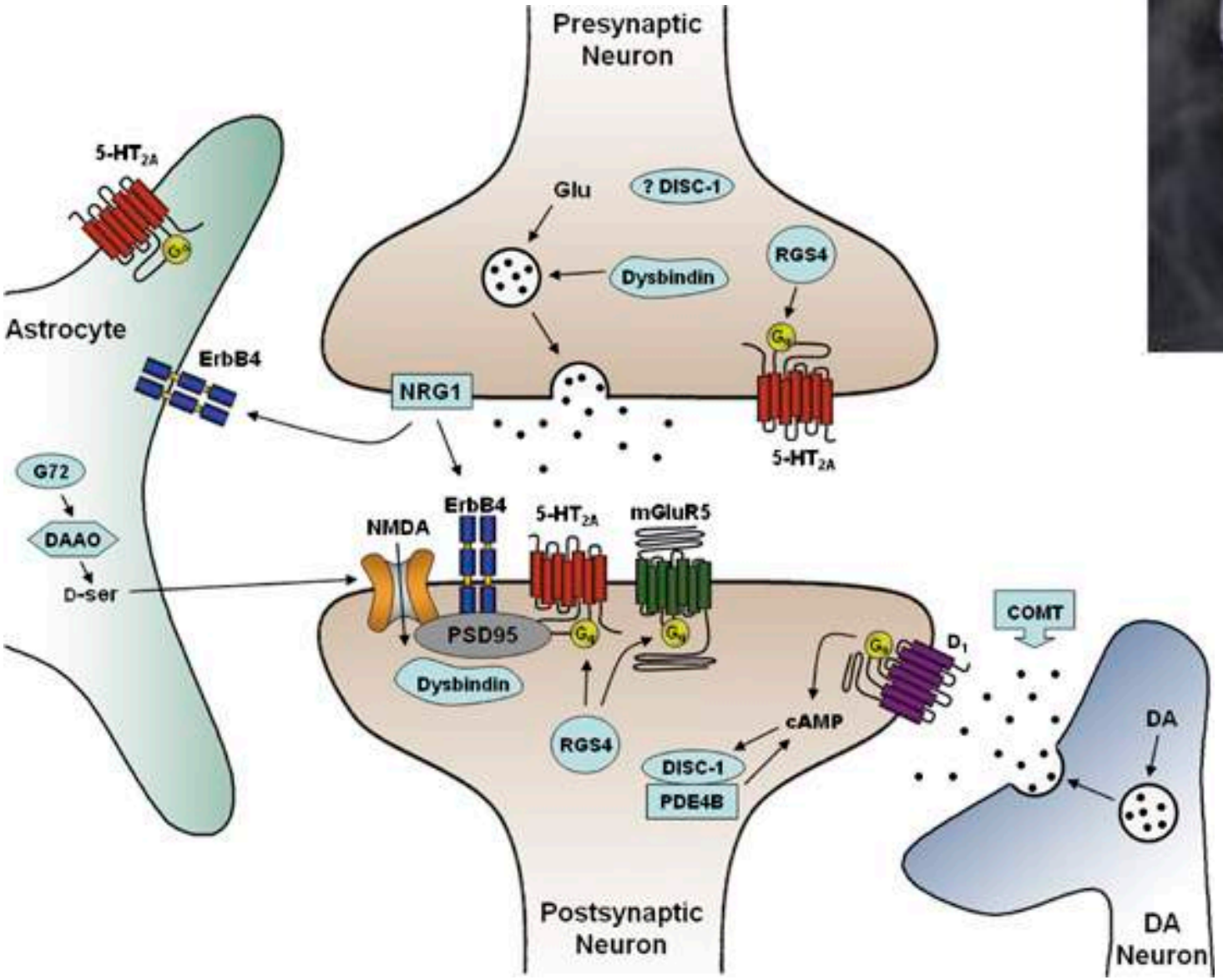
Kinetic rates of light-activated rhodopsin lifetimes



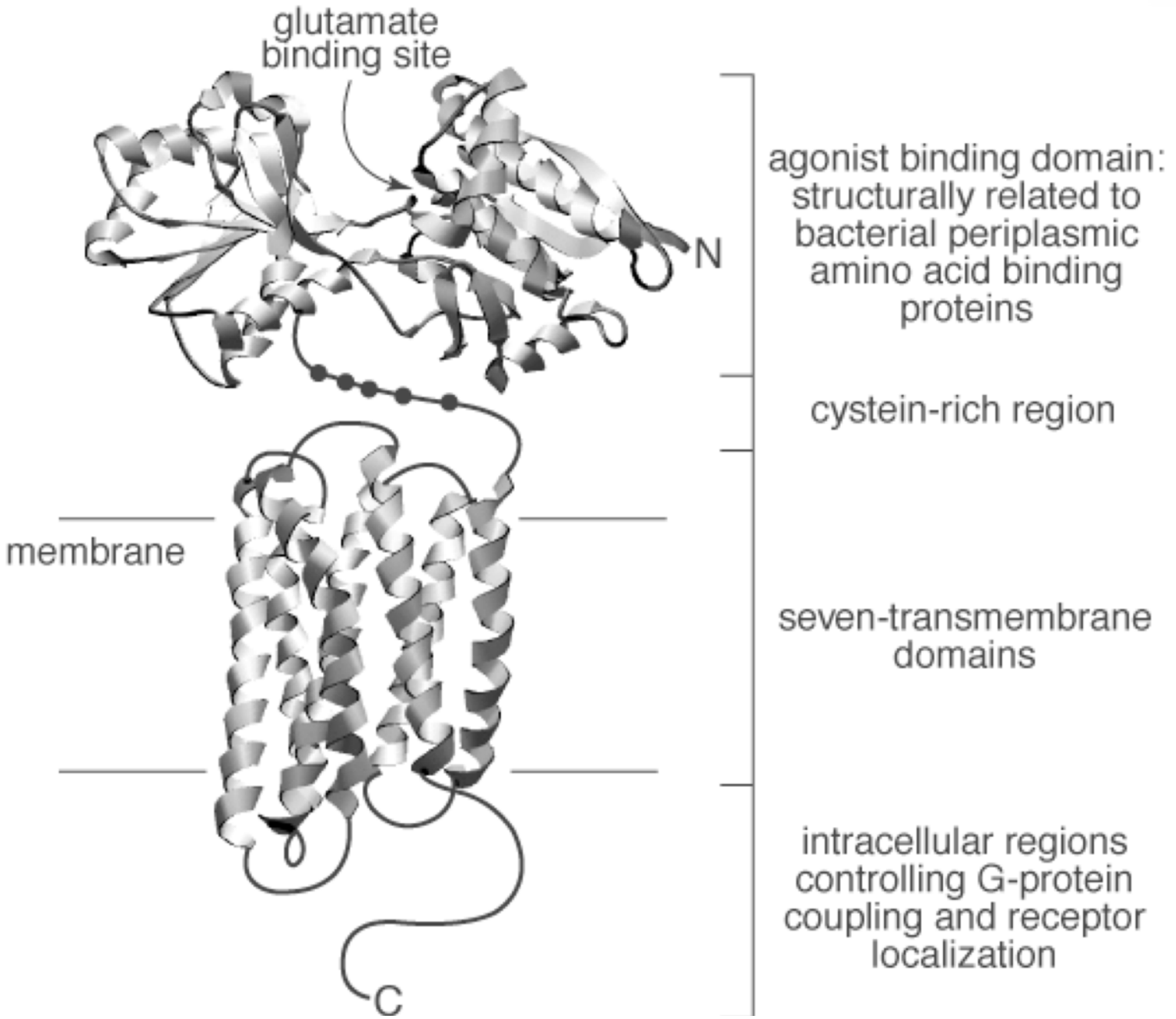
Bickelmann et al. 2015, *Evolution*

-> Increased lifetime of activated state of rhodopsin in mammalian and therian ancestors

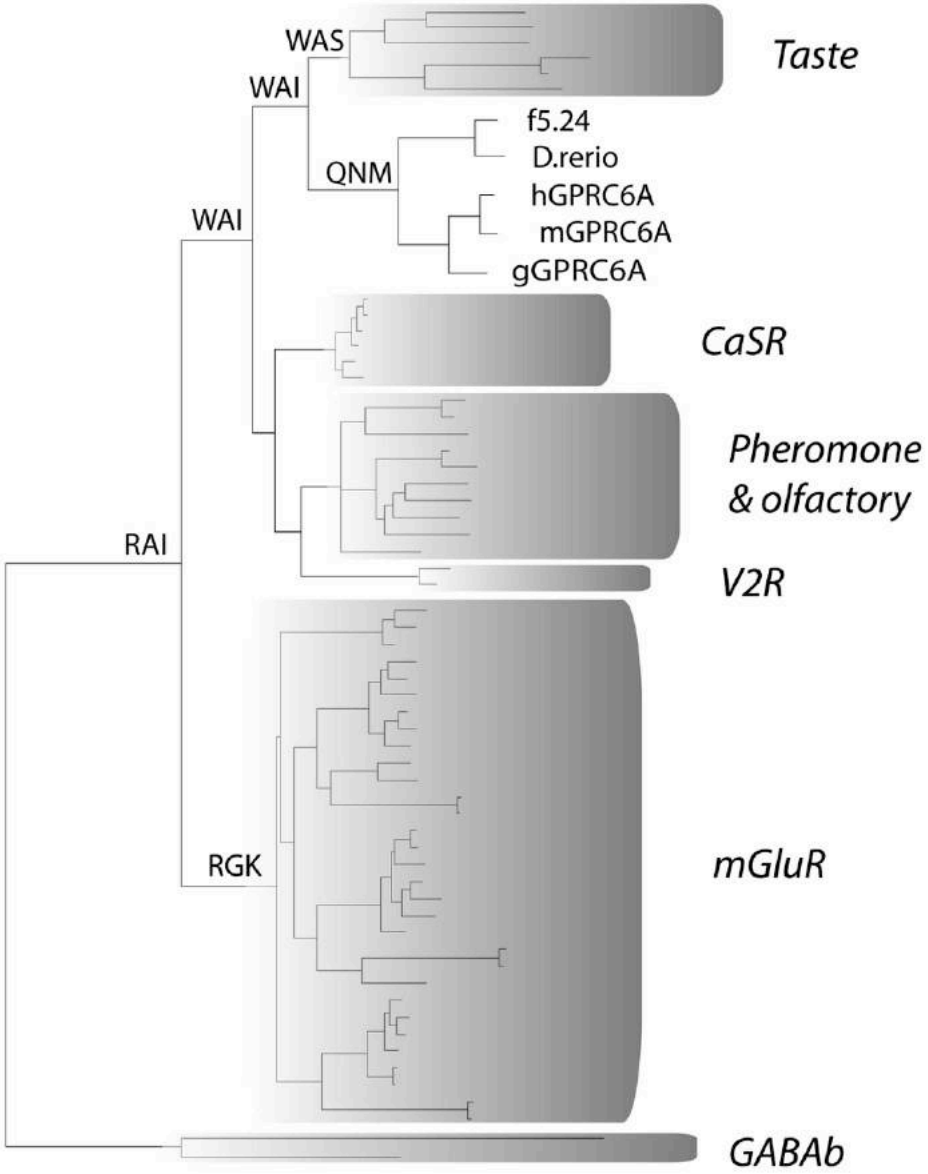
Synaptic neurotransmission



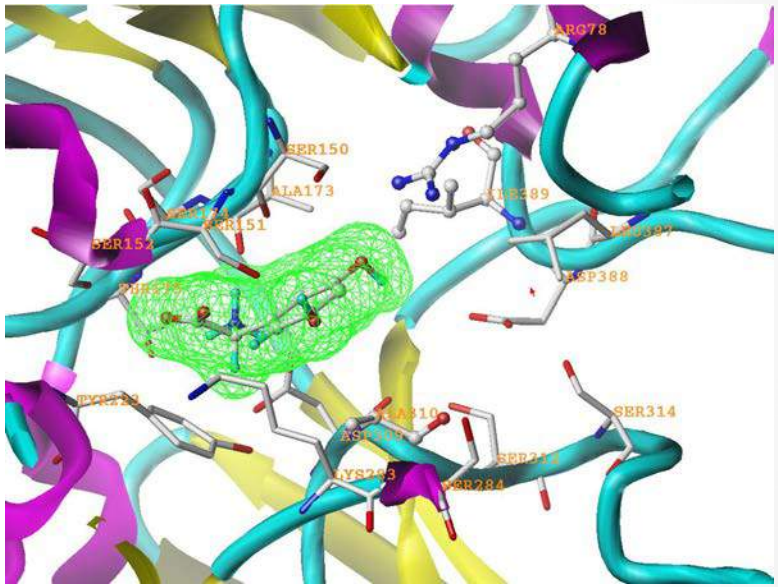
Glutamate receptors: Excitatory synaptic neurotransmission



Reconstructed ancestral AA-binding GPCR

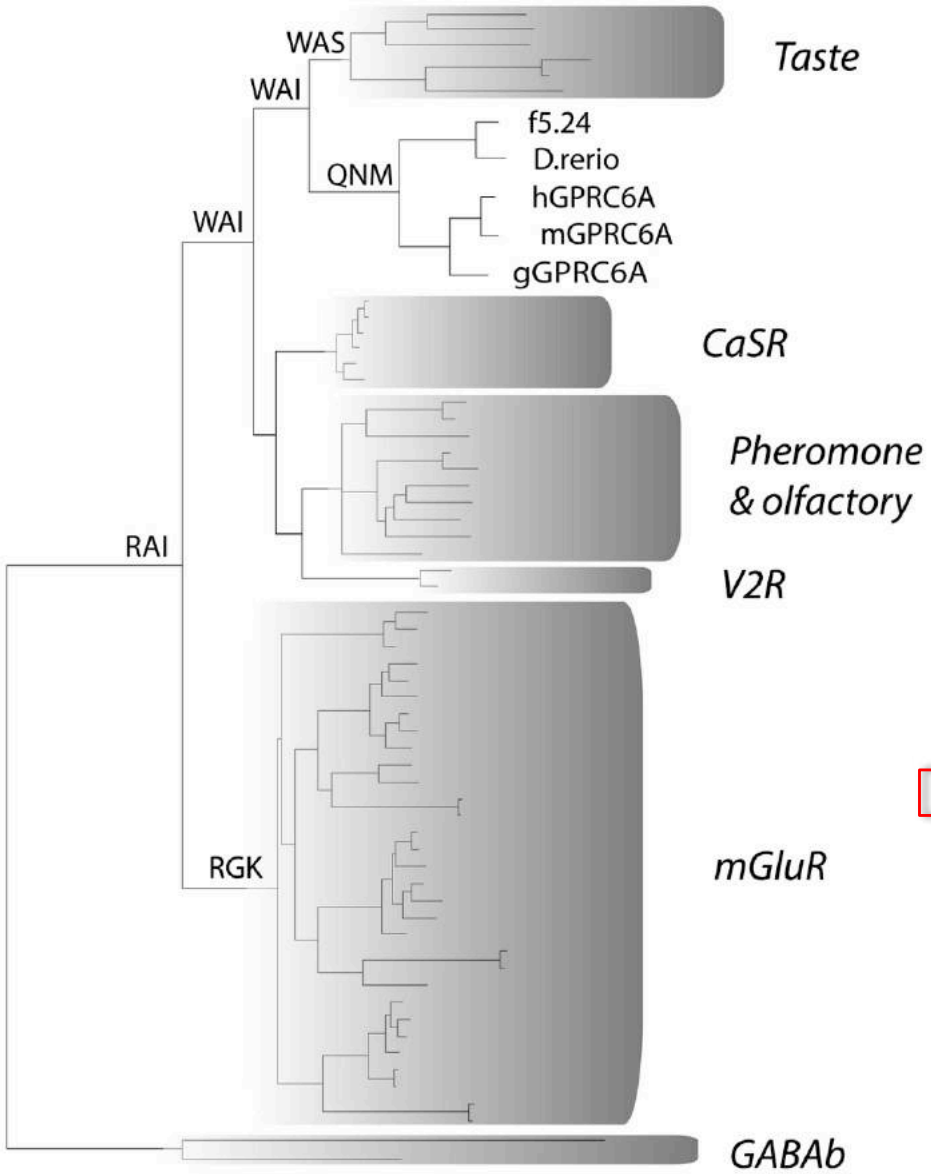


Ancestral ligand binding site
of AA-binding GPCR



Kuang et al., 2006
PNAS (103):14050

Reconstructed ancestral AA-binding GPCR

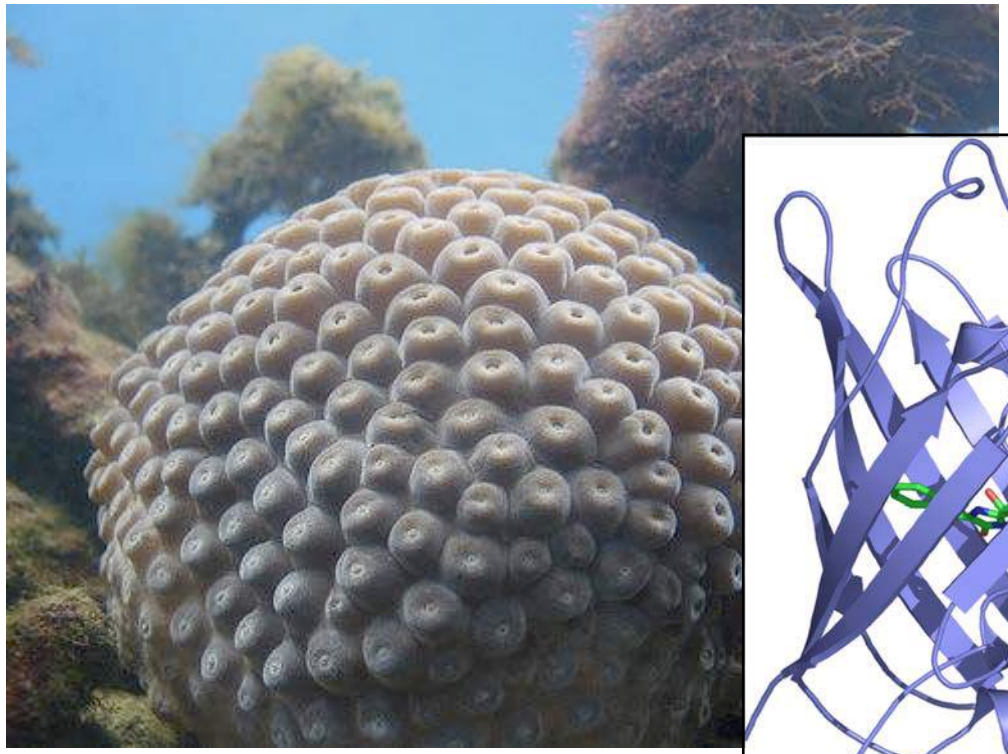


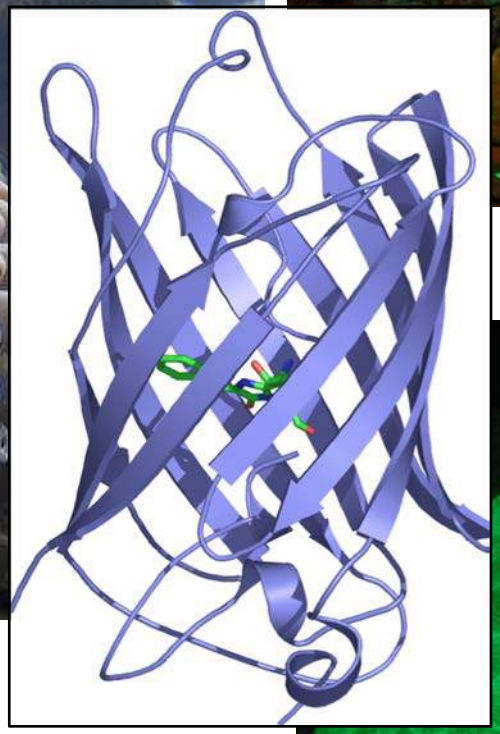
Comparison of potency (EC50)
Glutamate R agonists

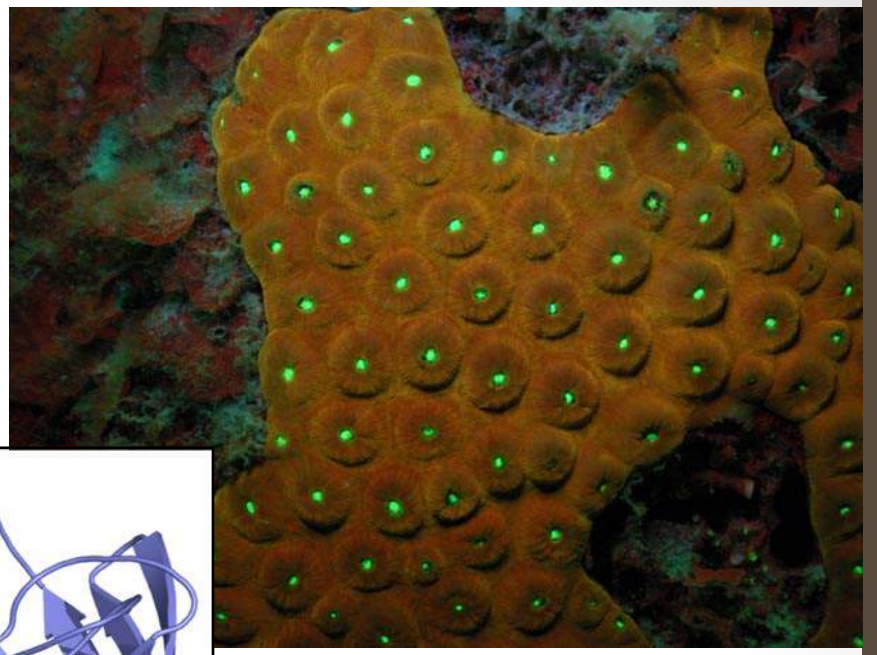
Agonists	5.24 receptor	Ancestral receptor	Q78R/ N310G	Q78R	mGluR1a
L-CCG-1	+	4.9 ± 1.2	3.5 ± 0.8	1.3 ± 0.1	24.7 ± 1.9
ibotenate	183.7 ± 22.5	0.8 ± 0.3	3.2 ± 0.4	6.4 ± 0.3	4.4 ± 0.9
quisqualate	230.0 ± 58.0	0.8 ± 0.2	3.5 ± 0.8	8.3 ± 0.4	0.4 ± 0.2
(*S*)-DHPG	–	2.4 ± 1.3	3.8 ± 1.1	+	3.3 ± 0.7
DCG-IV	–	–	–	–	–
L-SOP	–	–	–	–	–

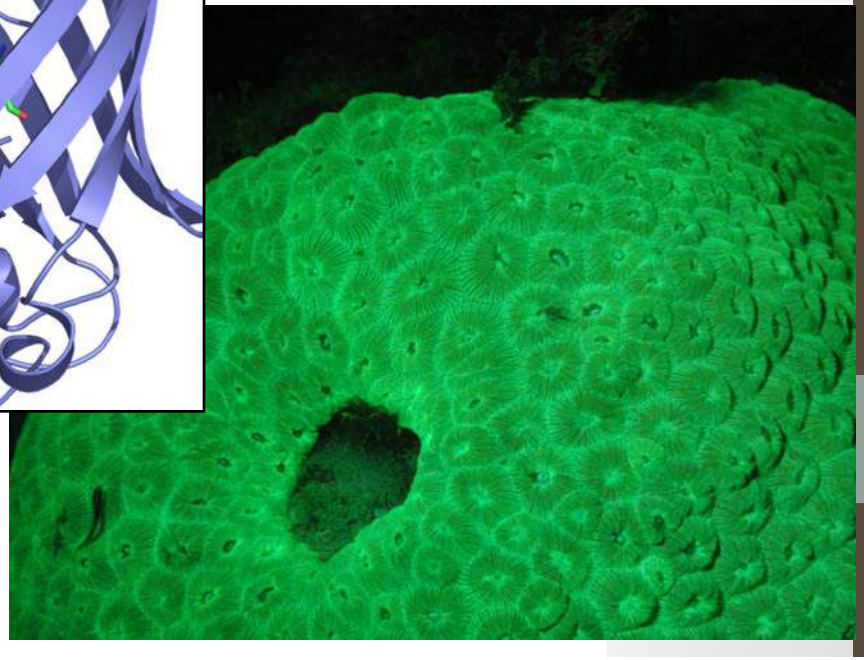
Kuang et al., 2006
PNAS (103):14050

Coral pigments



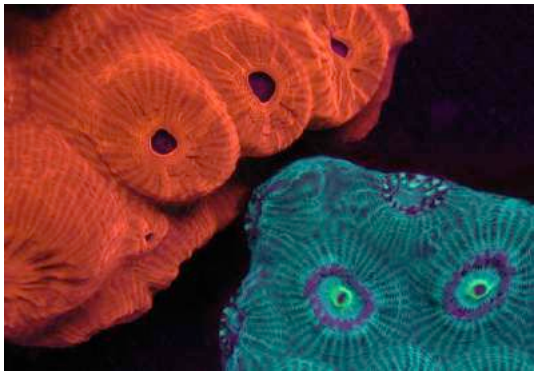


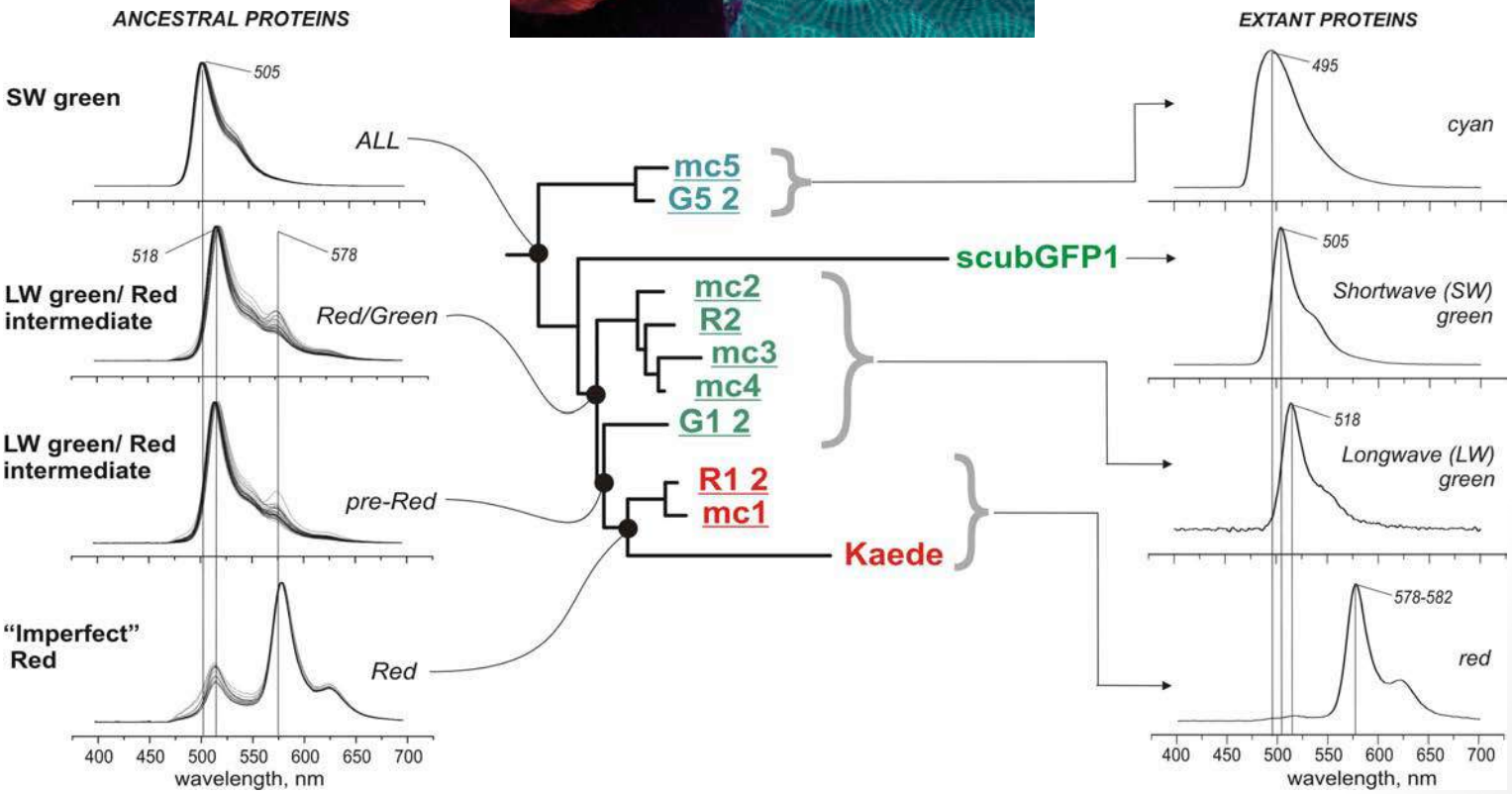




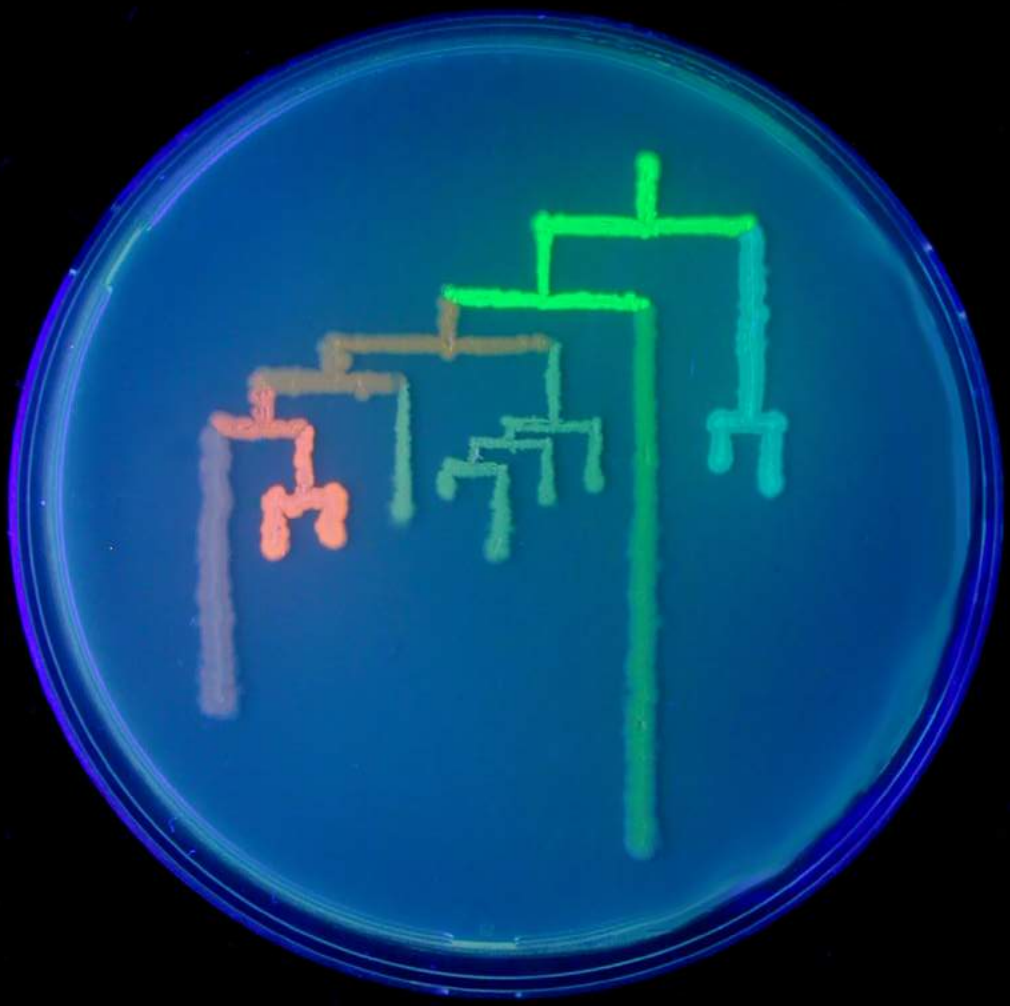
Reconstructed GFP-like proteins from coral

great star coral *(Montastraea cavernosa)*





Ugalde et al., 2004
Science (305):1433



Adaptive protein evolution

Conclusions: Ancestral reconstruction

Ancestral reconstruction approaches can offer a window into the past in studying ancient adaptive shifts in protein function

Computational analyses can be used to generate specific evolutionary hypotheses that can then be tested experimentally

Experimental approaches should not be viewed as applications of computational methods, instead serve to extend the hypothesis testing framework to study the evolution of protein function

Need for more interaction between computational and experimental methods in order to provide better insight into both approaches in the study of molecular evolution