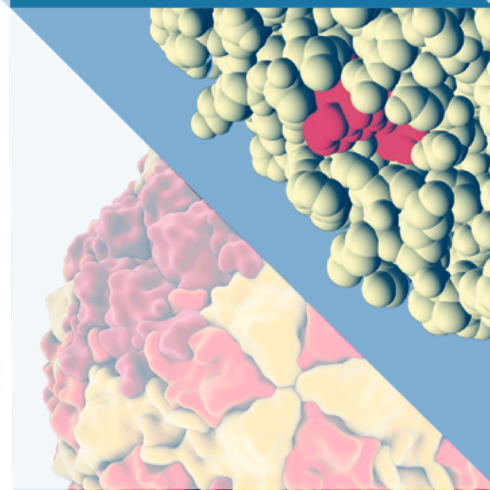
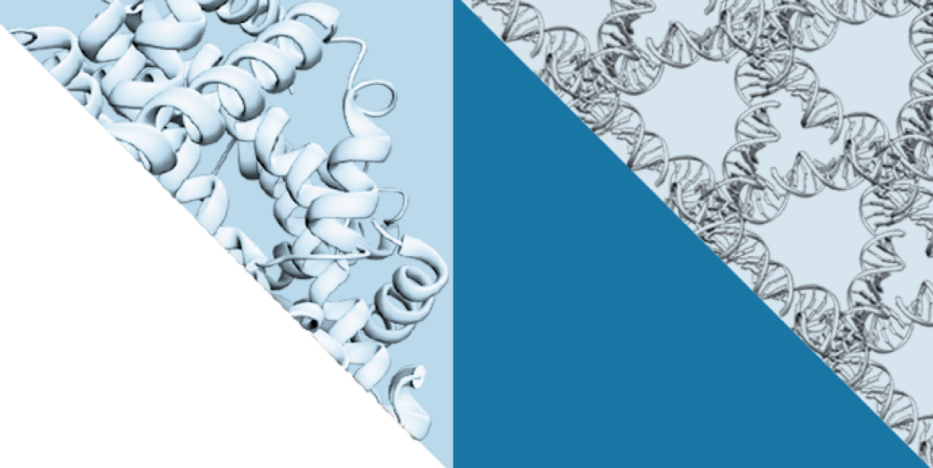


# Disorder in Protein Structures from the Protein Data Bank

Stephen K. Burley, M.D., D.Phil.  
University Professor and Henry Rutgers Chair  
Founding Director, Institute for Quantitative Biomedicine  
Director, RCSB Protein Data Bank  
Member, Rutgers Cancer Institute of New Jersey  
Rutgers, The State University of New Jersey

February 25<sup>th</sup> 2019

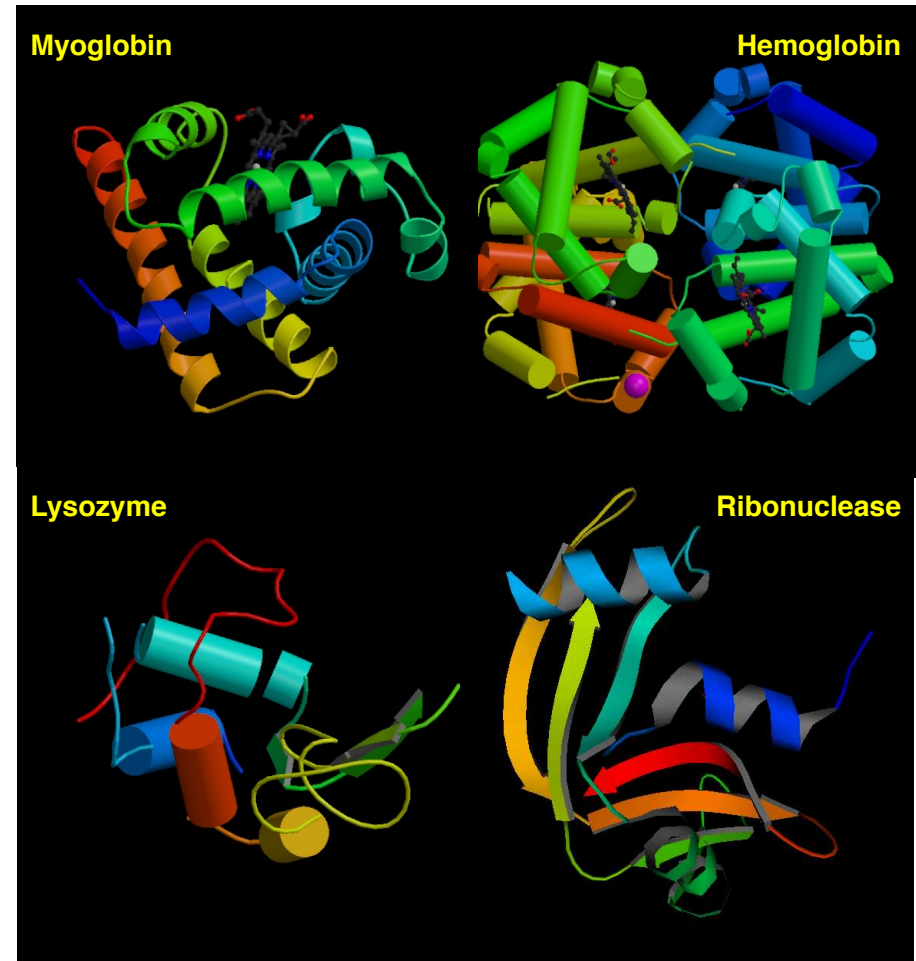


# Outline

- Protein Data Bank History
- RCSB PDB Sustaining a Living Data Resource
- IDPs and Regulation of Translation in Eukaryotes
- Accessing PDB Data *via* RCSB.org
- Learning about IDPs from PDB Data (MX)
- Acknowledgements

# Protein Data Bank History

- PDB 1<sup>st</sup> Open Access digital data resource in all of biology
- Founded 1971 with 7 X-ray structures of proteins
- Single global **archive** for protein and DNA/RNA experimental structures
- Today, Open Access to ~150,000 structures
- wwPDB collaboration US (RCSB PDB), EU (PDBe), Japan (PDBJ), and BMRB



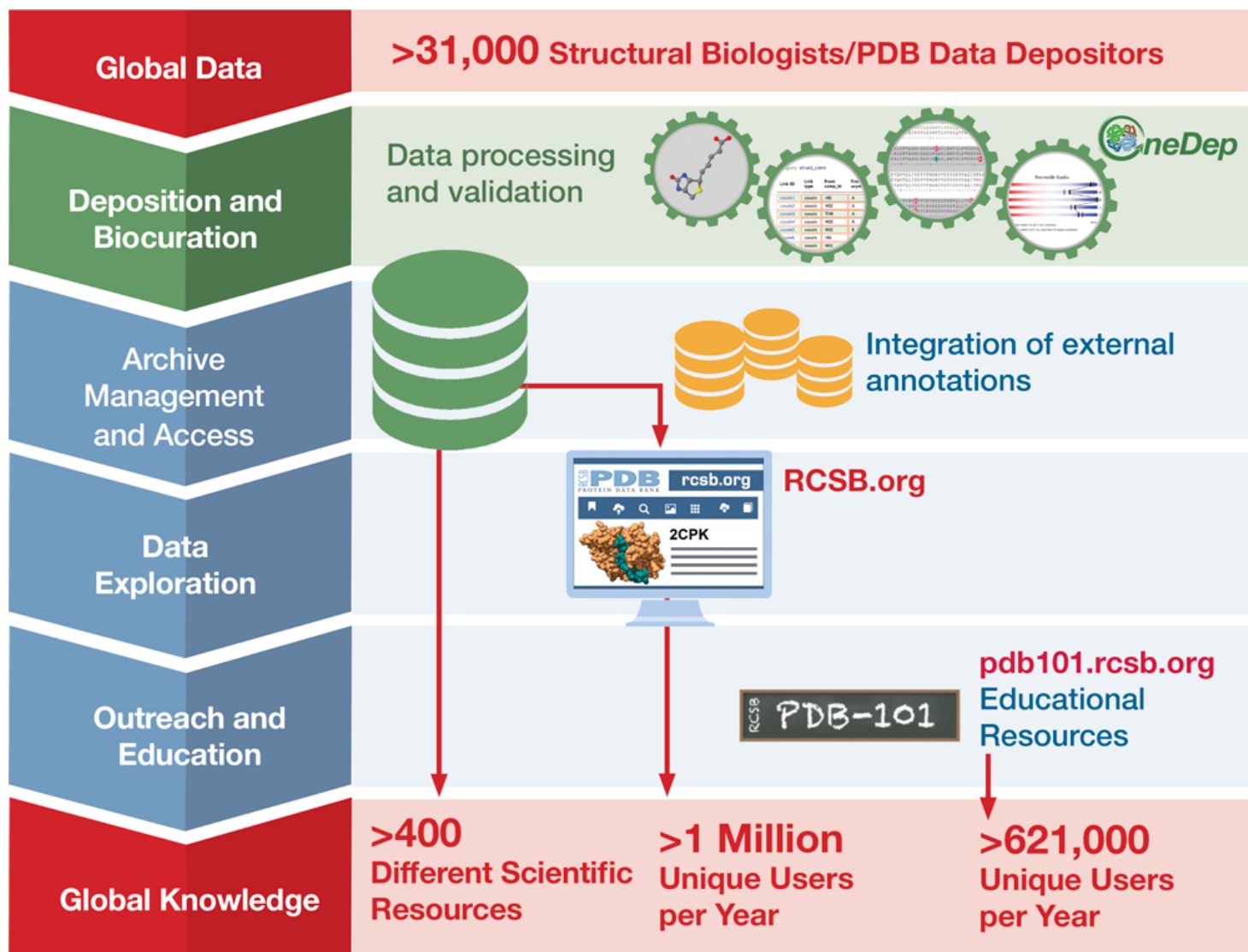
Some of the earliest structures in the PDB

# RCSB PDB Supporting US Research/Education in Fundamental Biology, Biomedicine, and Energy

- Managed since 1999 by [RUTGERS](#) / [UC San Diego](#)
- Collaborates with Worldwide PDB in global support of Data Depositors
- Supports US Research ([RCSB.org](#))
- Outreach and Education ([PDB101.RCSB.org](#))
- Funded jointly by NSF, NCI, NIGMS, and DOE

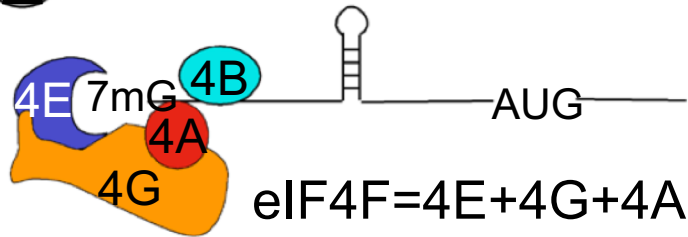


# RCSB PDB Services Sustain a Living Data Resource



# Assembling the Translation Machinery

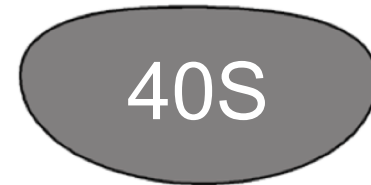
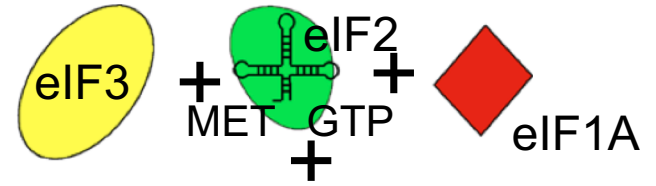
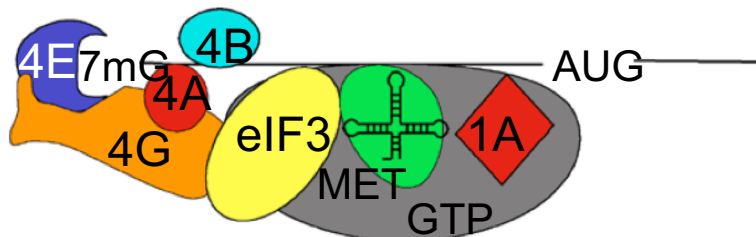
②



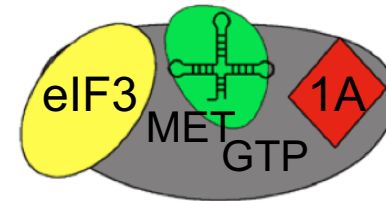
+ATP



③

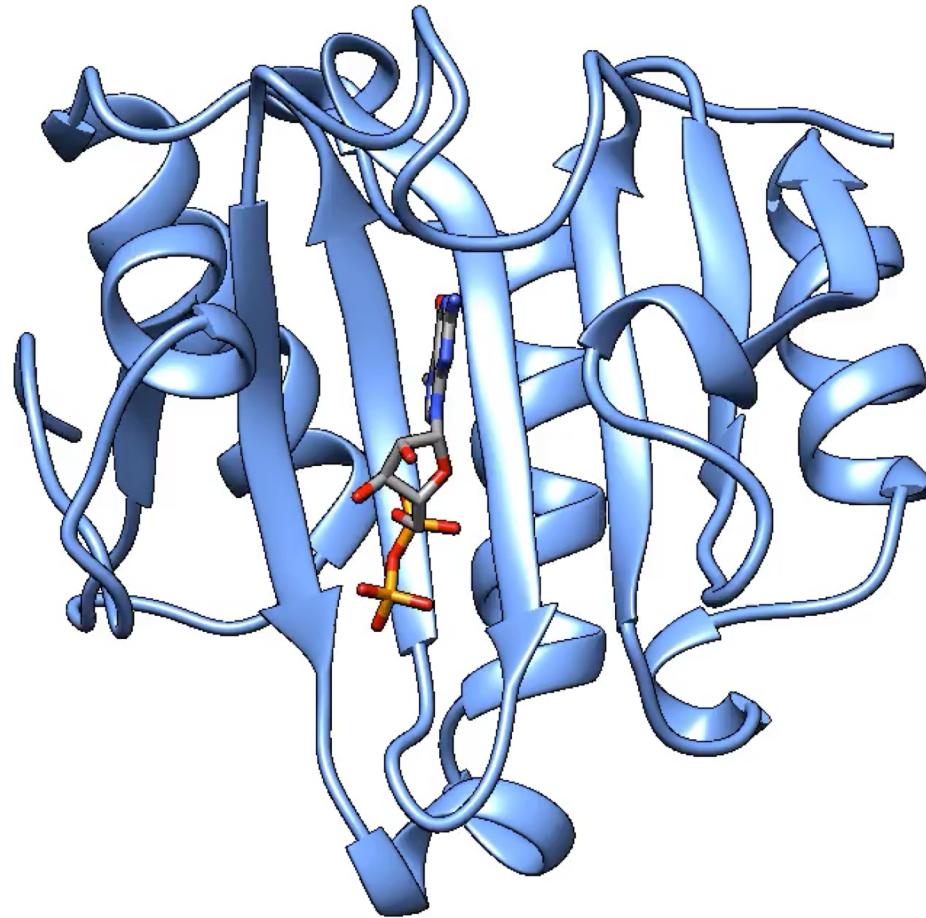


①



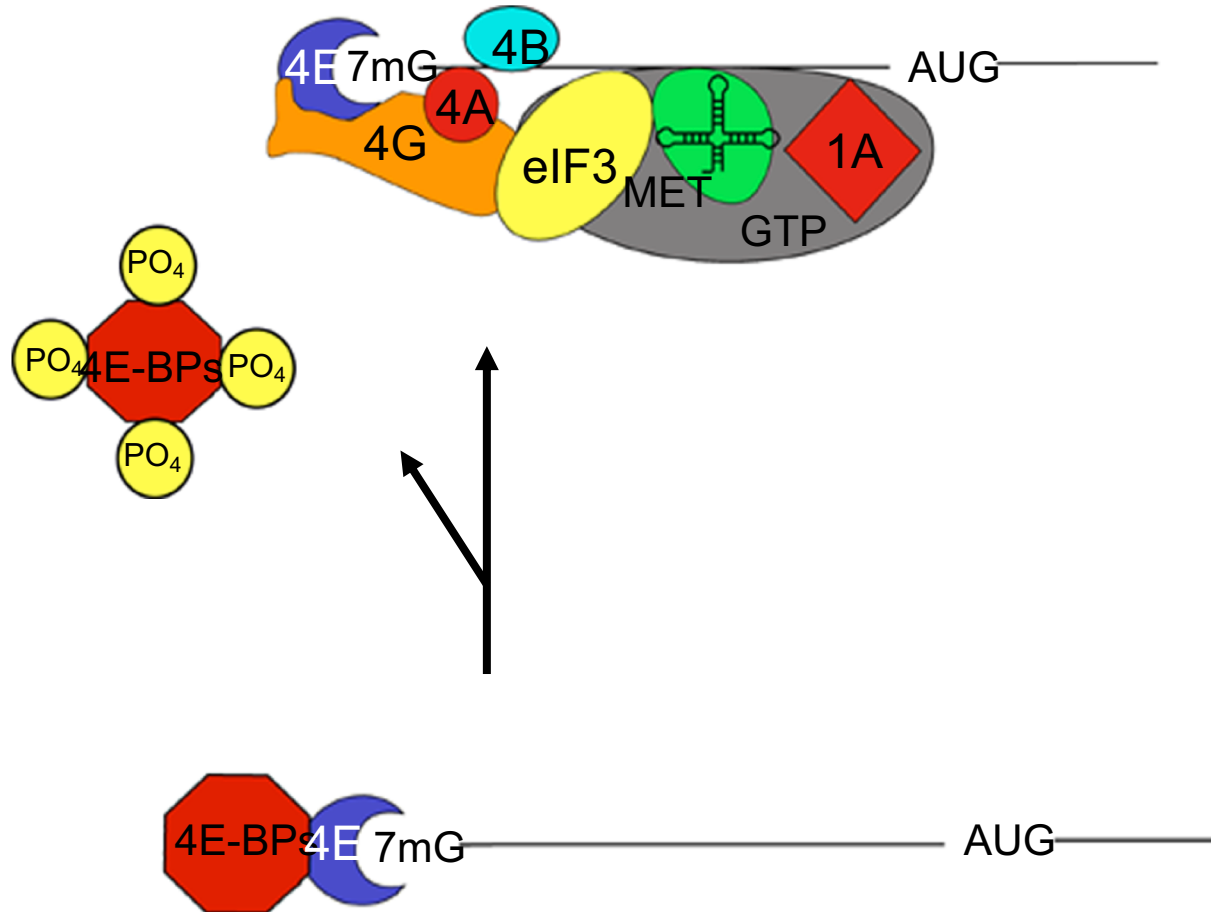
43S Ribosomal  
preinitiation complex

# eIF4E Recognizing the mRNA Cap (PDB 1ej1)



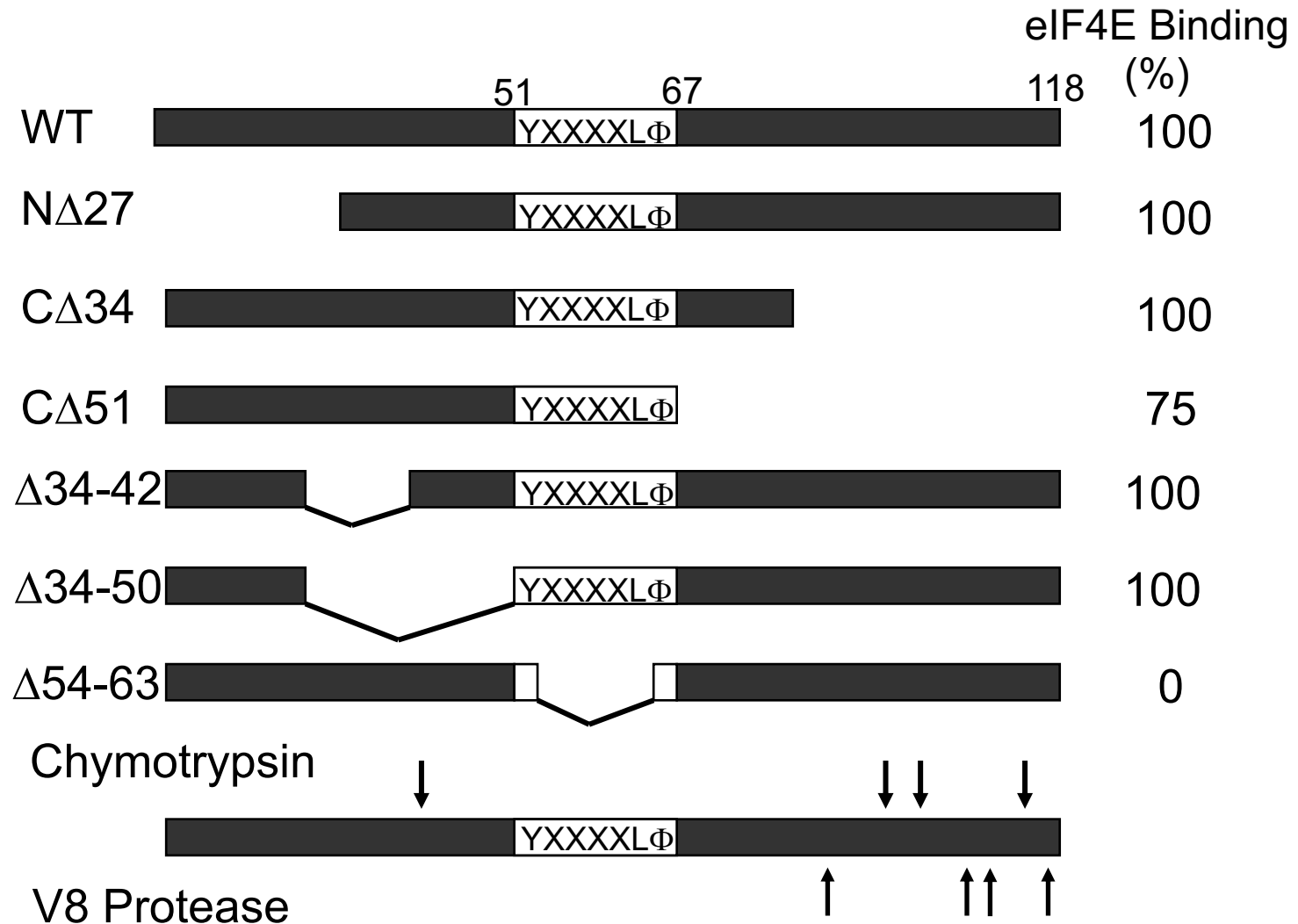
Marcotrigiano *et al.* (1997) *Cell* 89, 951-961.

# Regulating Translation Initiation with IDPs

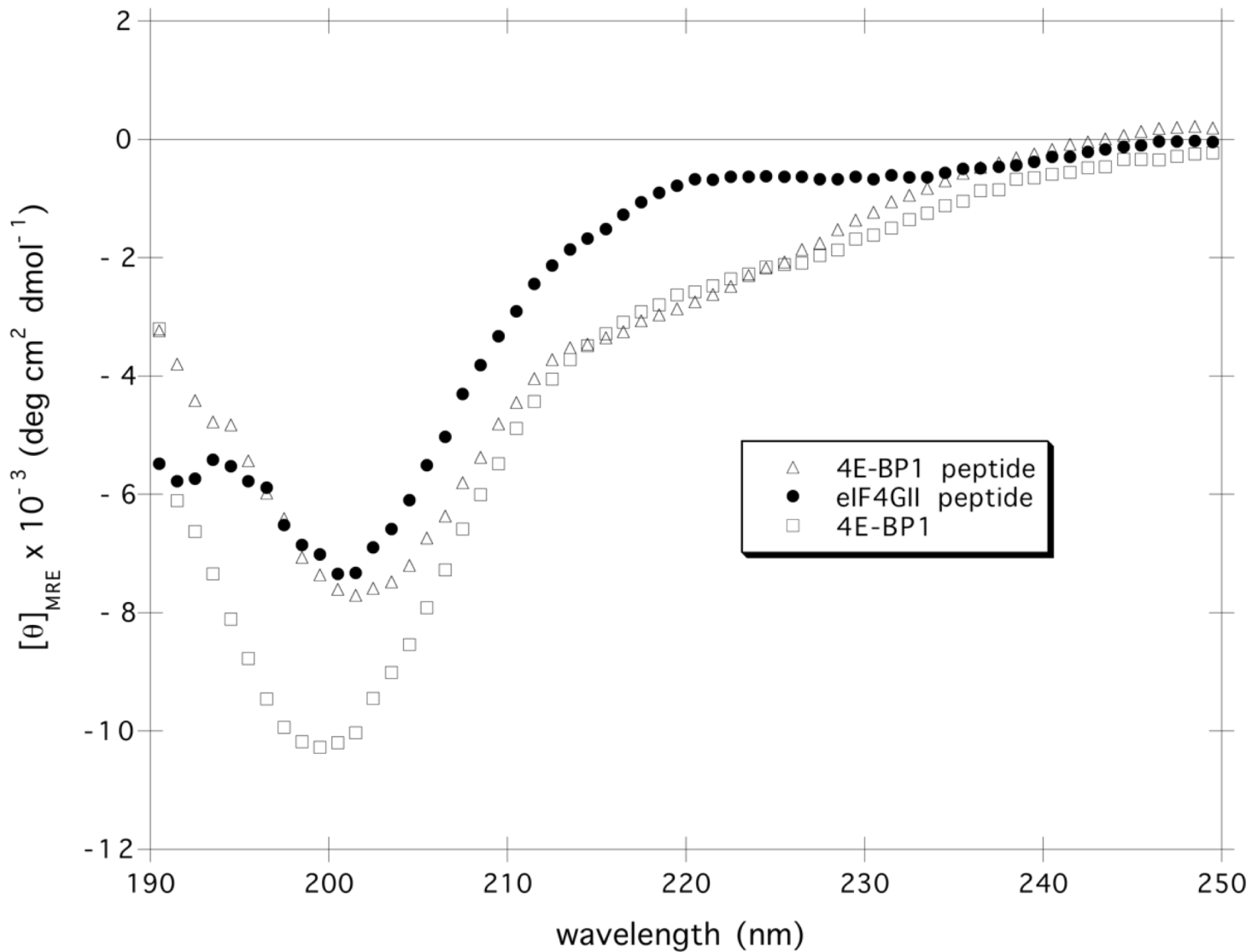




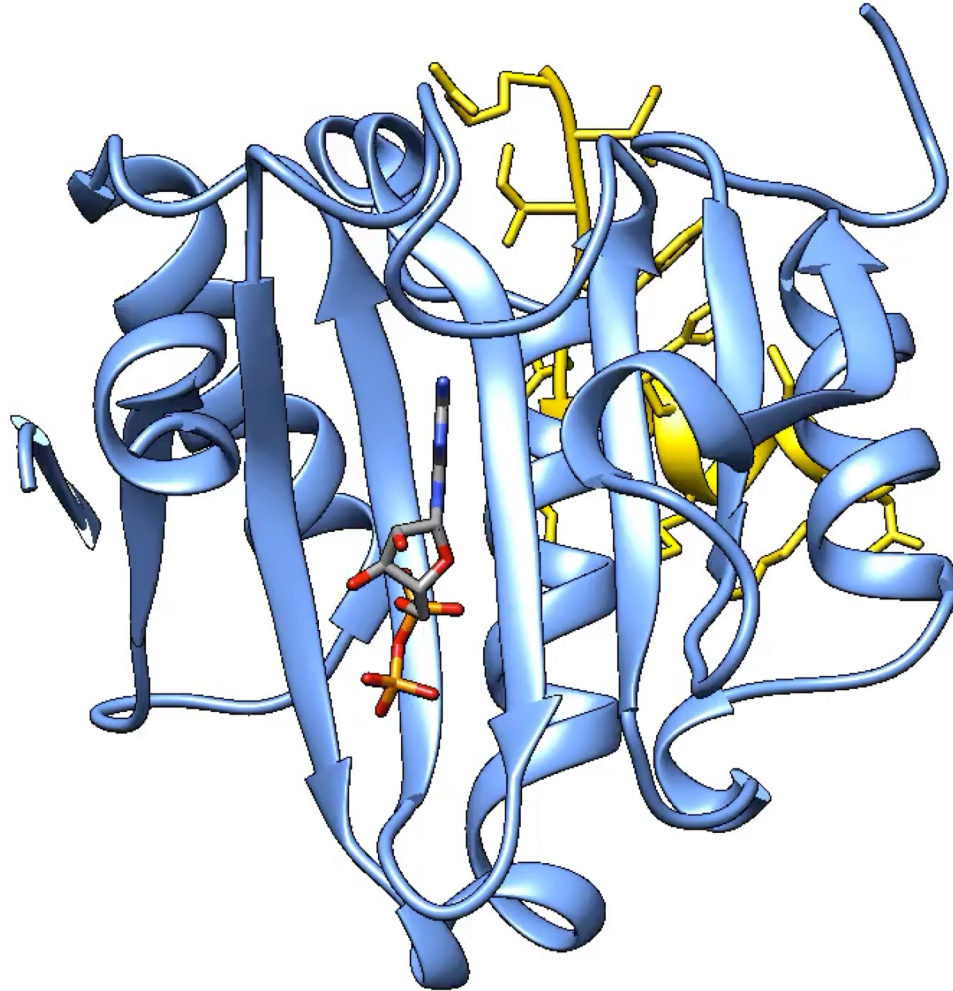
# Functional Characterization of 4E-BP1



# 4E-BP1 is an IDP! As is eIF4G Peptide!



# 4E-BP1 Recognizing eIF4E-Cap (PDB 1ej4)



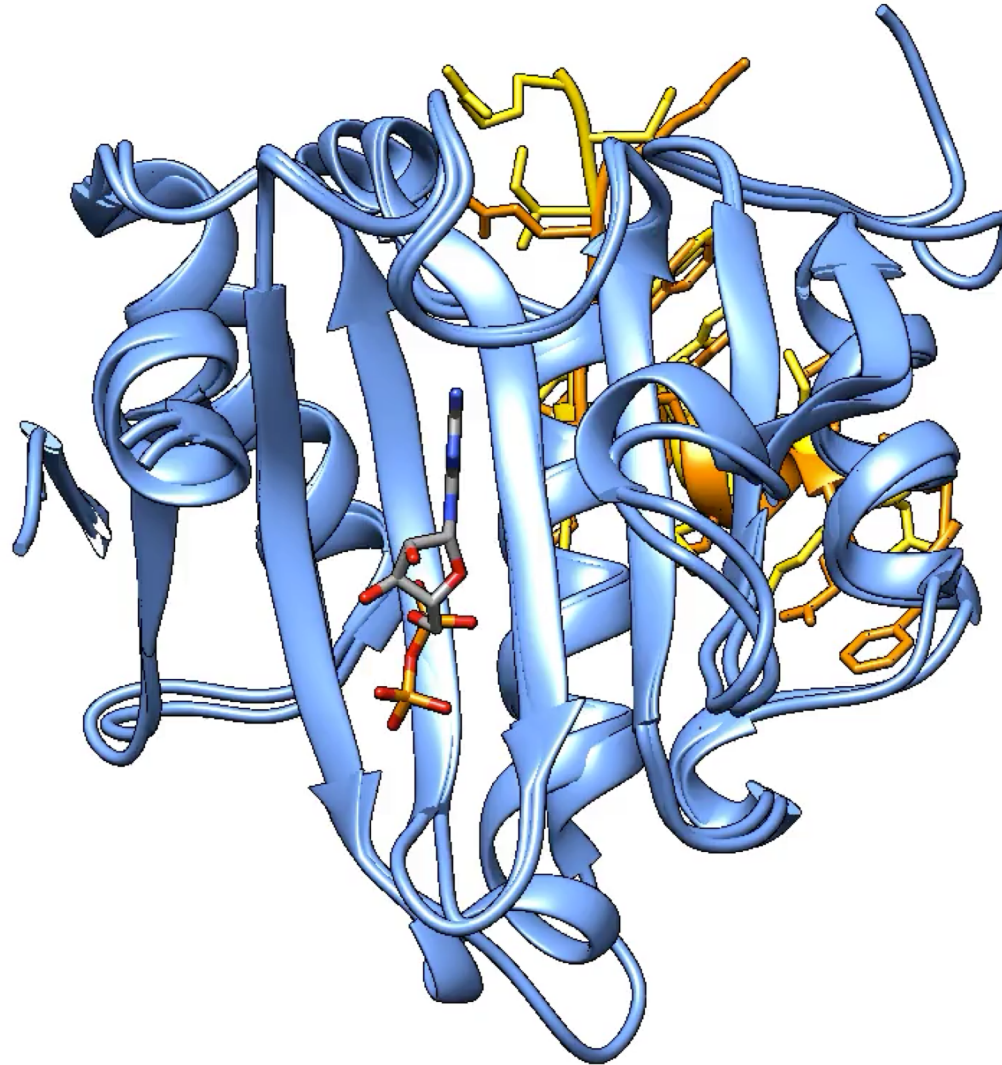
Marcotrigiano *et al.* (1999) *Molecular Cell* 3, 707-716.

# eIF4G Recognizing eIF4E-Cap (PDB 1ejh)



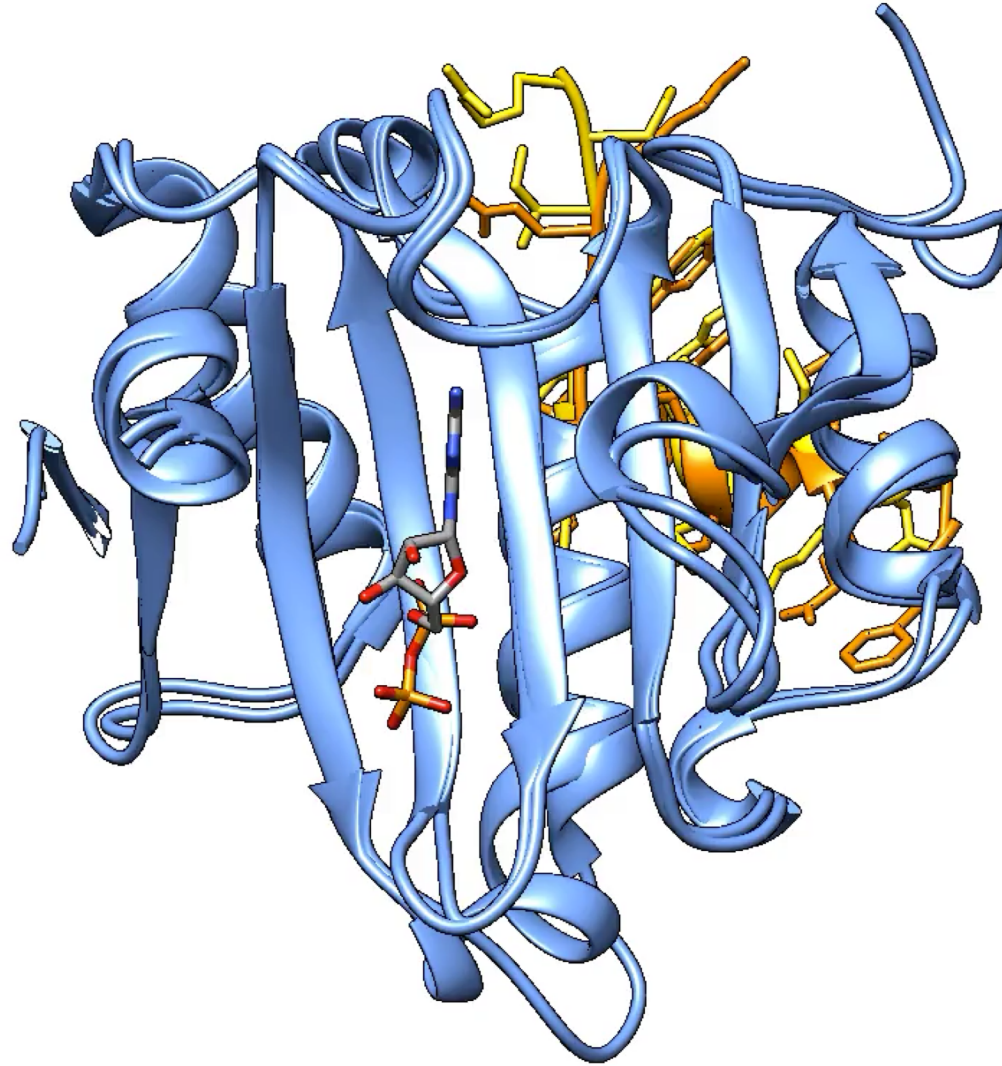
Marcotrigiano *et al.* (1999) *Molecular Cell* 3, 707-716.

# 4E-BP1 is a Molecular Mimic of eIF4G



Marcotrigiano *et al.* (1999) *Molecular Cell* 3, 707-716.

# 4E-BP1 and eIF4G Exhibit Disorder→Order



Marcotrigiano *et al.* (1999) *Molecular Cell* 3, 707-716.

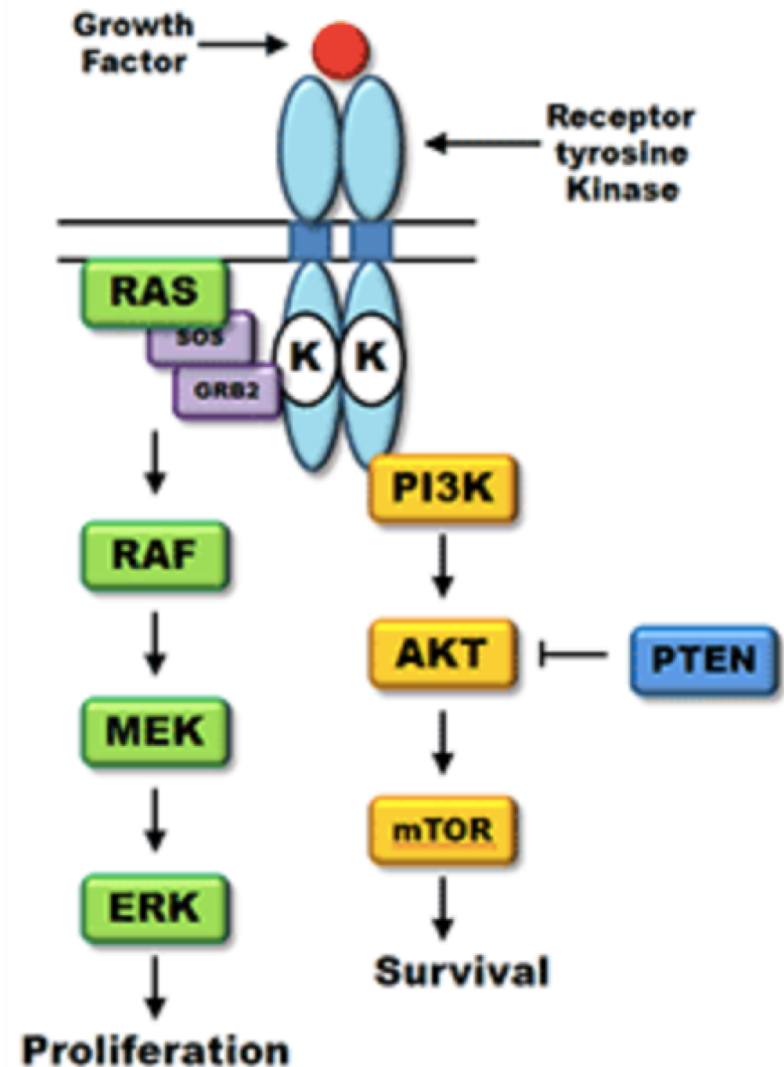
# Accessing PDB Data *via* RCSB.org

- Deposited Structure Data (X-ray, NMR, and 3DEM)
  - Atomic Coordinates
  - Experimental Data
  - Ligand Information: SMILES string, common name
  - Metadata: Organism, Sequence, Sample, etc.
- Data Integrated with ~40 External Data Resources
  - Functional Annotations, etc. (updated Weekly)
- Sequence/Structure Similarity and Visualization

**Living Structure Data Eco-system:  
Going Well Beyond Original Publications**

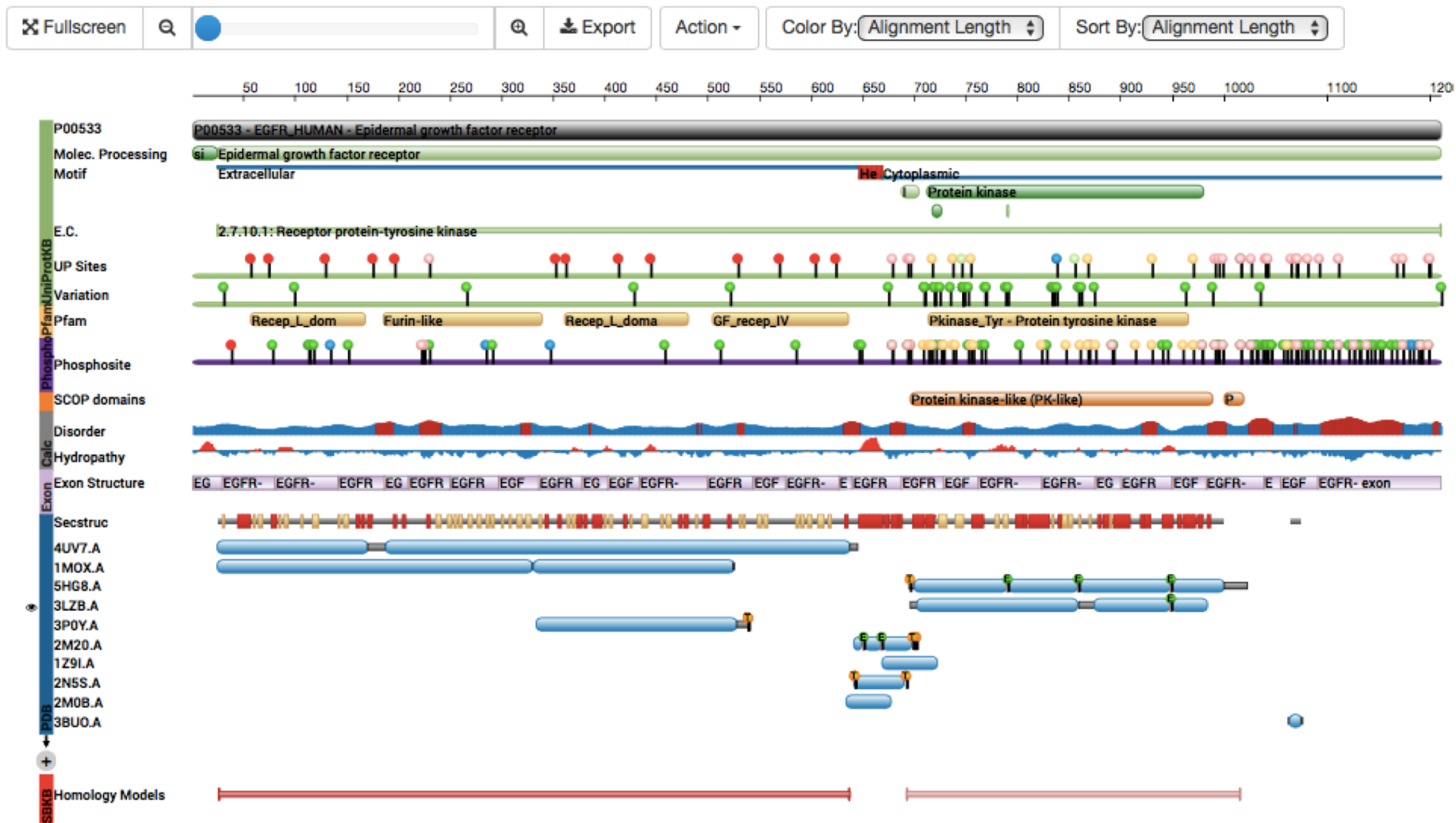
# EGFR Receptor Tyrosine Kinase

- Epidermal Growth Factor Receptor (EGFR) is a receptor tyrosine kinase closely related to *v-erb-B*
- Activating somatic mutations and gene amplifications correlated with NSCLC
- EGFR exome sequencing routine at CINJ



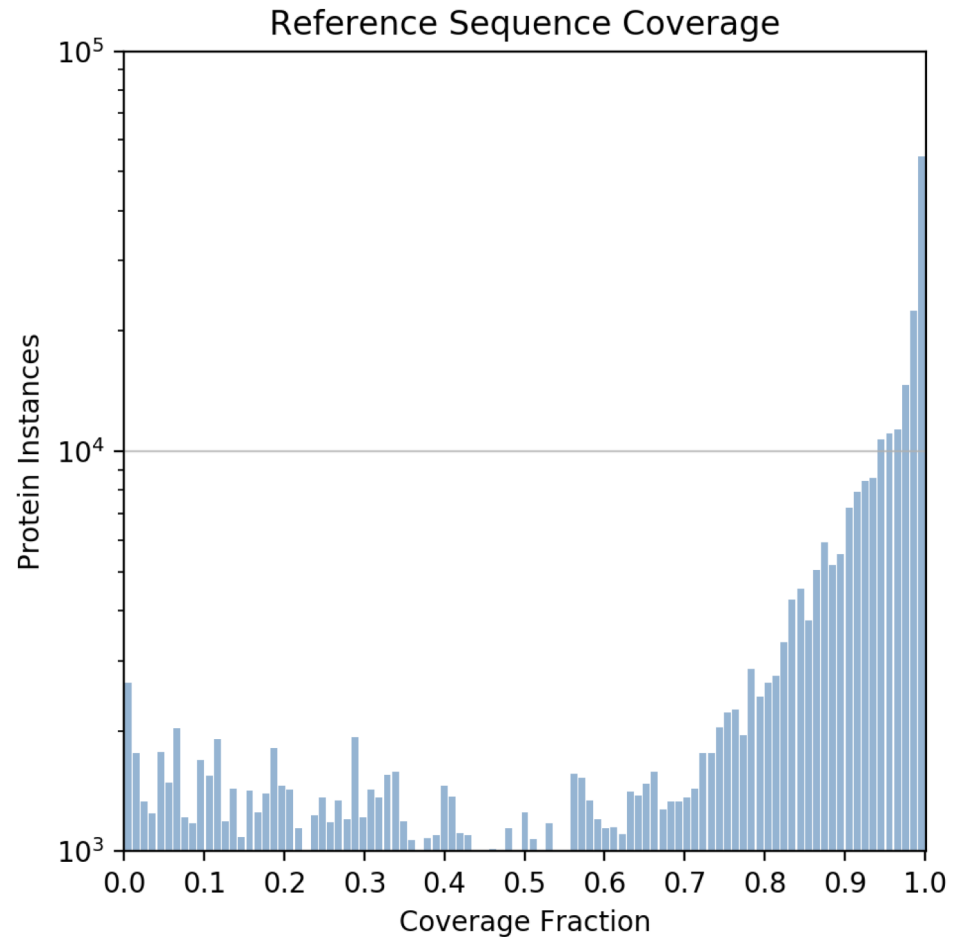


# RCSB.org Protein Feature View for EGFR



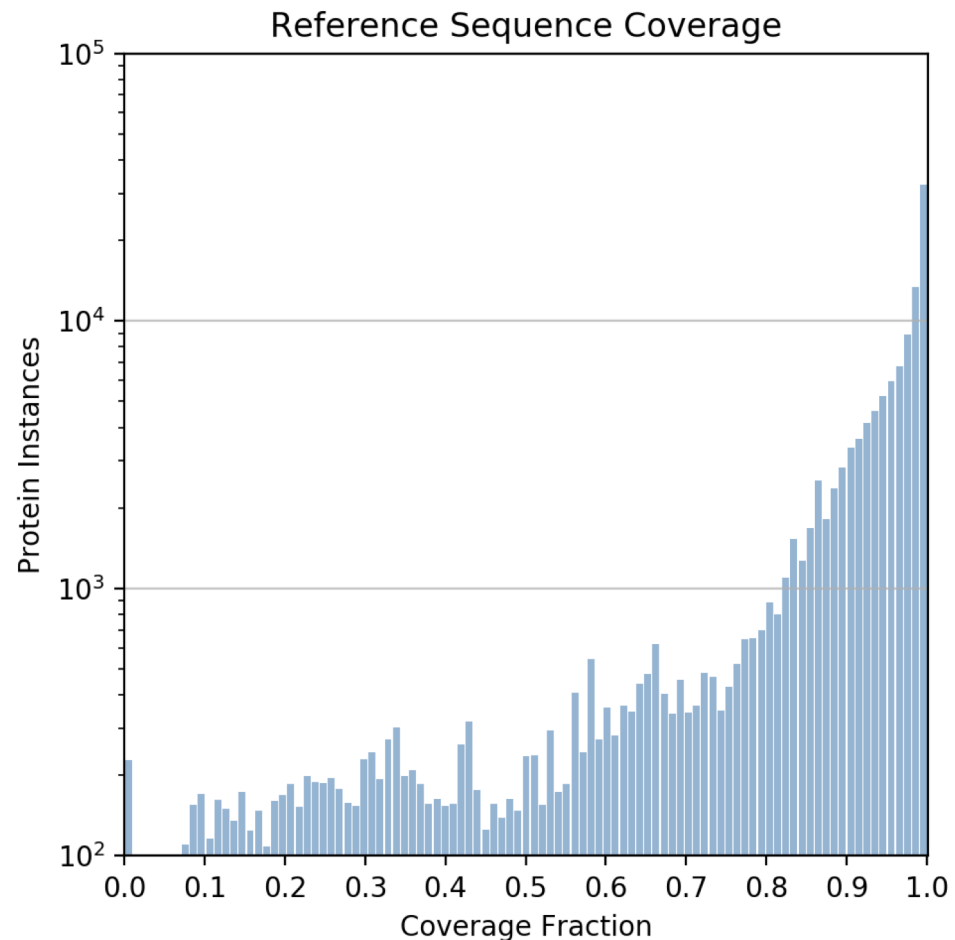
# Many Protein MX Structures in PDB: Cover Only Part of the Full-Length Sequence

- UniProt sequence coverage for 311,405 all-organism protein sequences in 126,830 MX structures (<3.5Å)



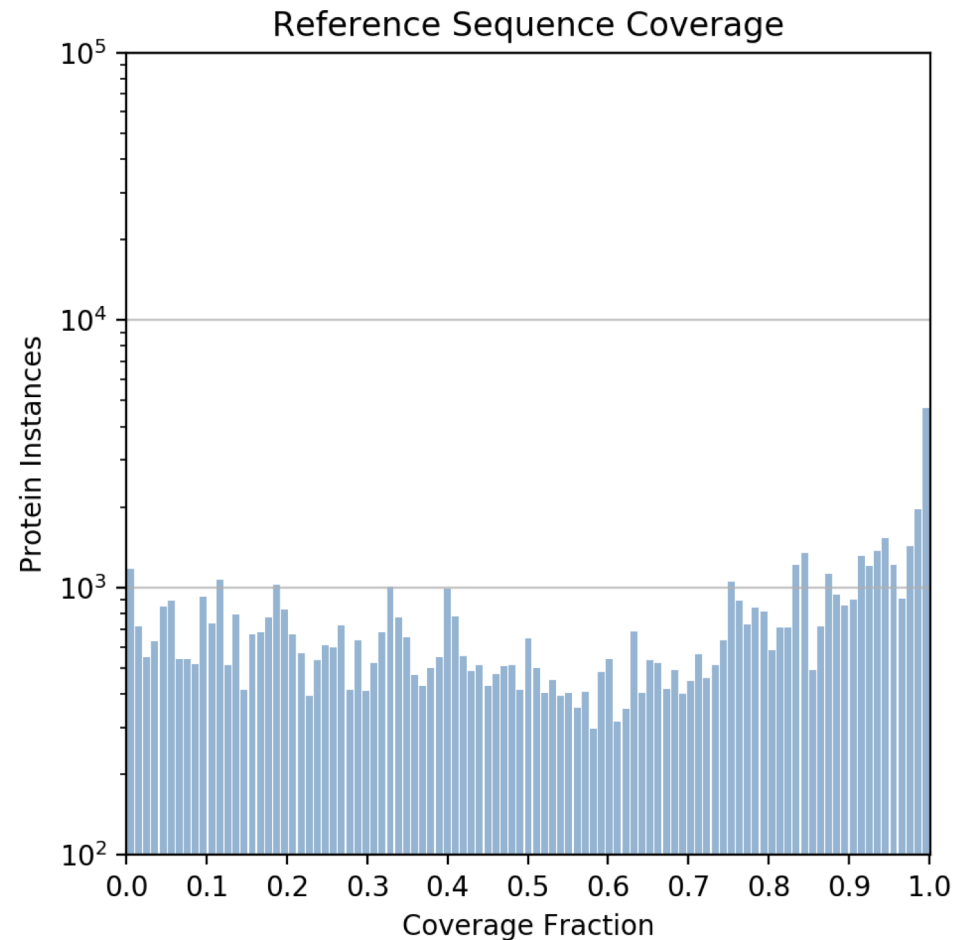
# Bacterial Protein MX Structures in PDB: More Likely to Cover Full-Length Sequence

- UniProt sequence coverage for 125,077 **bacterial** protein sequences in 41,975 MX structures (<3.5Å)



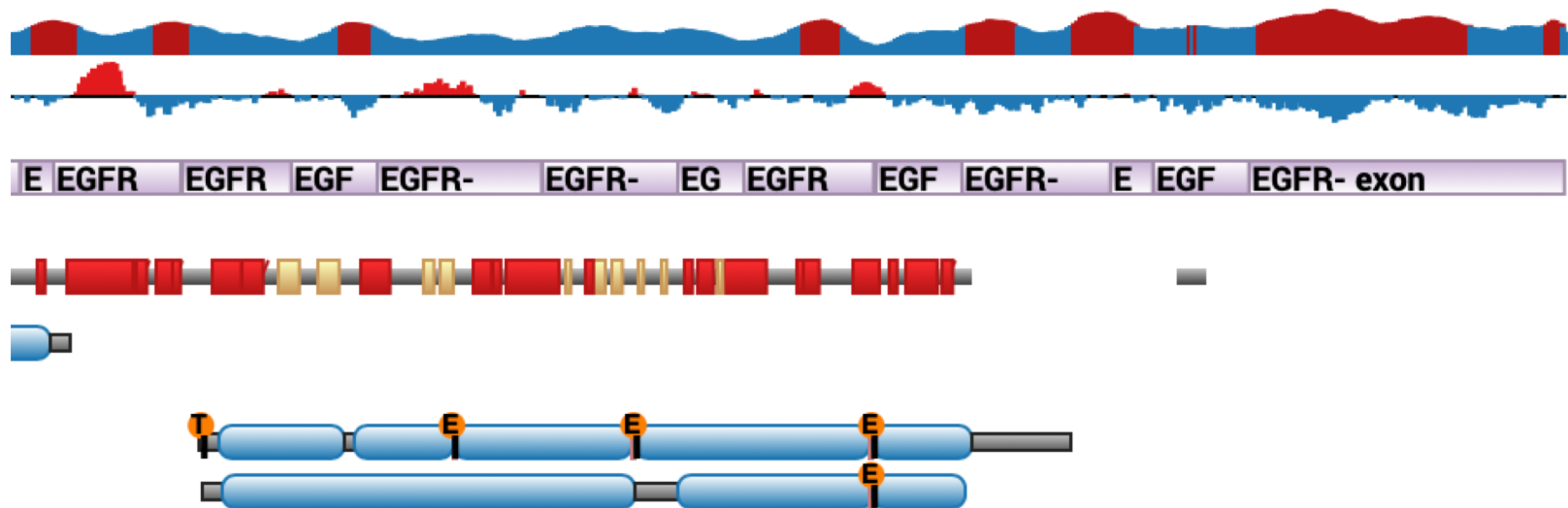
# Human Protein Structures in PDB: Less Likely to Cover Full-Length Sequence

- UniProt sequence coverage for 72,801 **human** protein sequences in 33,457 MX structures (<3.5Å)



# RCSB.org Protein Feature View for EGFR

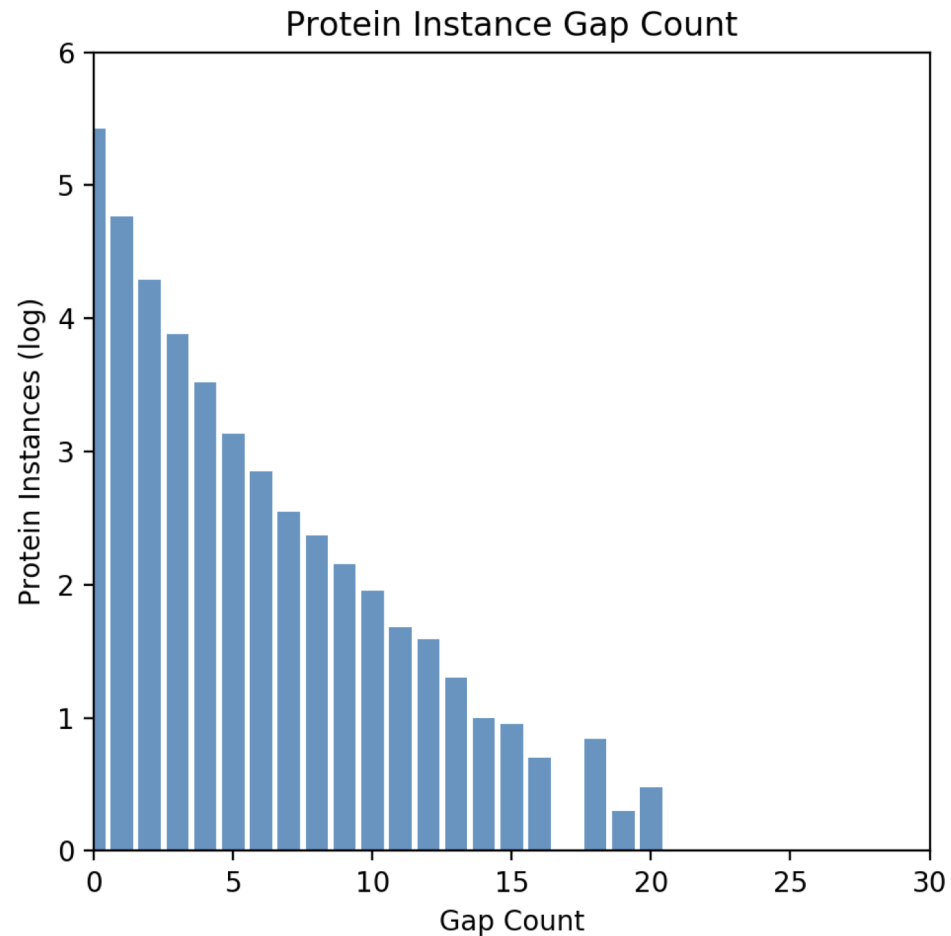
- Predicted regions of disorder (red) correlated with
  - Gaps in PDB structures (grey)
  - Absence of structural data (white space on right)



Disorder predicted using RONN – Yang *et al.* (2005) *Bioinformatics* 21, 3369-3376.

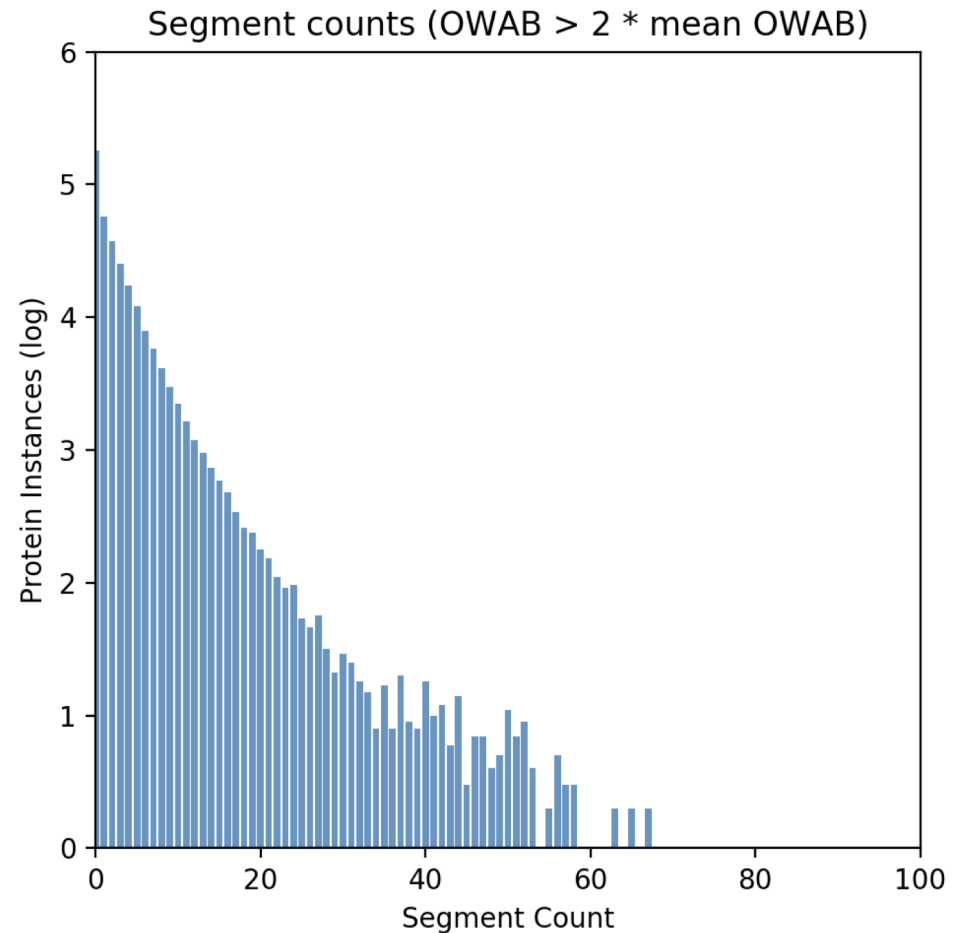
# Many Protein MX Structures in PDB: Gaps in Polypeptide Chain Electron Density

- Gap counts for 356,696 all-organism protein chains in 126,830 MX structures (<3.5Å)
- Electron Density Gap lengths=1→140+



# Many Protein MX Structures in PDB: Have Poorly Ordered Electron Density Segments

- $B\text{-factor} > 2 * \langle B\text{-factor} \rangle$   
for 356,696  
all-organism  
protein chains in  
126,830 MX structures  
( $< 3.5\text{\AA}$ )
- High B-factor segment  
lengths =  $1 \rightarrow 100$



# Learning about IDPs from PDB Data (MX)

- Structures of IDPs post Disorder→Order Transition
  - 4E-BP1 } Marcotrigiano *et al.* (1997) *Cell* 89, 951-961.
  - eIF4G } Marcotrigiano *et al.* (1999) *Molecular Cell* 3, 707-716.
- Many proteins really do look like “beads on a string”
  - Eukaryotic even more so than Bacterial
- Many protein structures have disordered segments that cannot be visualized with crystallography
- Many protein structures have poorly-ordered segments that are hard to see with crystallography



# Acknowledgements



**RCSB.ORG**

[info@rcsb.org](mailto:info@rcsb.org)

## Funding

RCSB PDB is funded by a grant (DBI-1338415) from the National Science Foundation, the National Cancer Institute, the National Institute of General Medical Sciences and the US Department of Energy.



## Management

The RCSB PDB is managed by:

**RUTGERS**

**UC San Diego**

**SDSC** SAN DIEGO SUPERCOMPUTER CENTER

Follow us



The RCSB PDB is a member of the Worldwide Protein Data Bank partnership (wwPDB; [wwpdb.org](http://wwpdb.org)).

