

Proteins: Structure and Domains

Bioinformatics for Biologists

What is a Protein?

Watch before continuing



WHAT IS A PROTEIN?

RCSB PDB-101



0:00 / 6:57

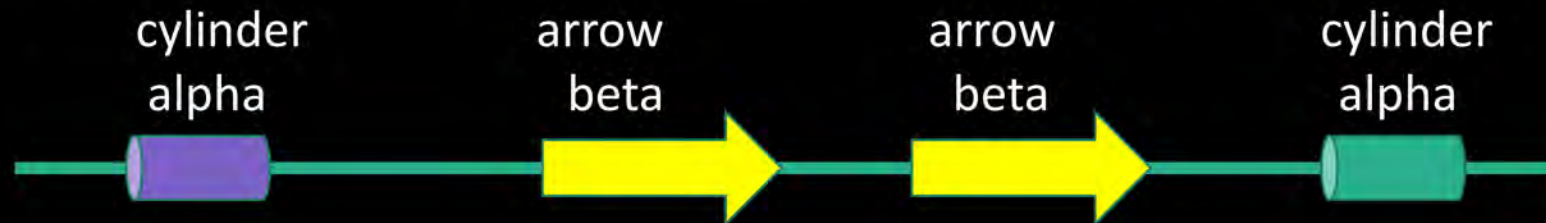


<https://www.youtube.com/watch?v=wvTv8TqWC48>

Protein structure

- Primary: Sequence

-
- Secondary: **alpha helix** and **beta strands**



- Tertiary: 3D



- Quaternary: Oligomeric state



Protein: PH4H_HUMAN (P00439)

1 architecture

1 sequence

0 interactions

1 species

17 structures

- Summary
- Sequence
- Structures
- TreeFam
- Jump to...
- enter ID/acc

Summary

This is the summary of UniProt entry [PH4H_HUMAN](#) (P00439).

Description:	Phenylalanine-4-hydroxylase EC=1.14.16.1
Source organism:	Homo sapiens (Human) (NCBI taxonomy ID 9606) View Pfam proteome data.
Length:	452 amino acids
Reference Proteome:	✓

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms the basis of the overview that you see here. It is important to note that, although some UniProt entries may be removed *after* a Pfam release, these entries will not be removed from Pfam until the *next* Pfam data release.

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains. [More...](#)

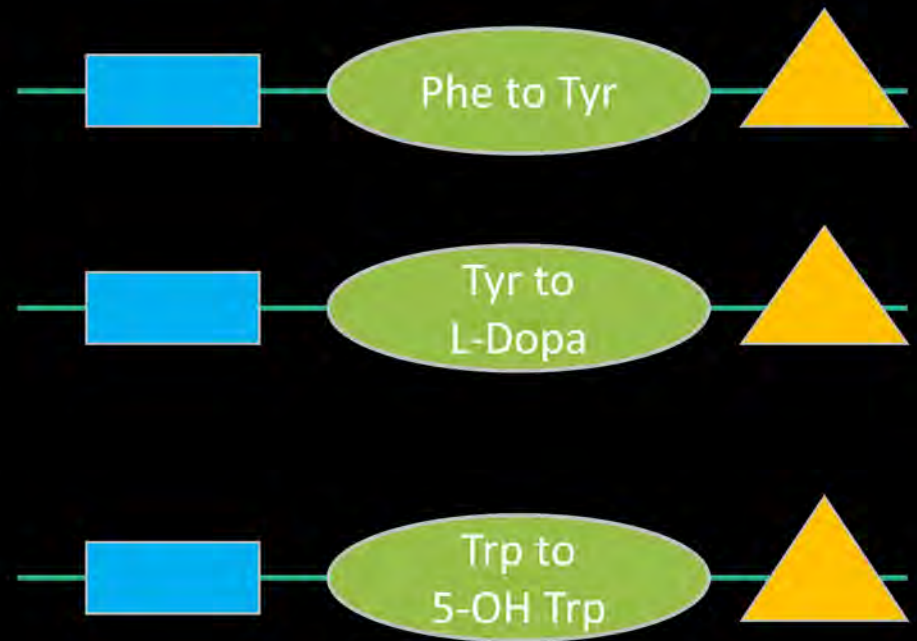
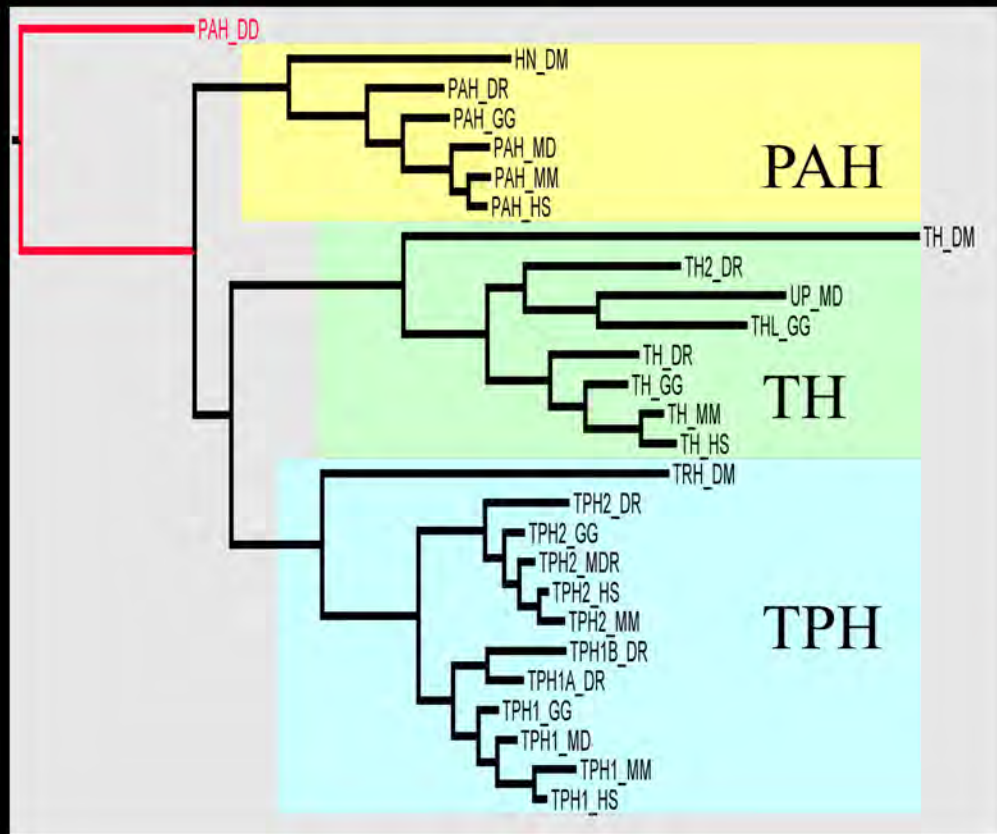


[Download](#) the data used to generate the domain graphic in JSON format.

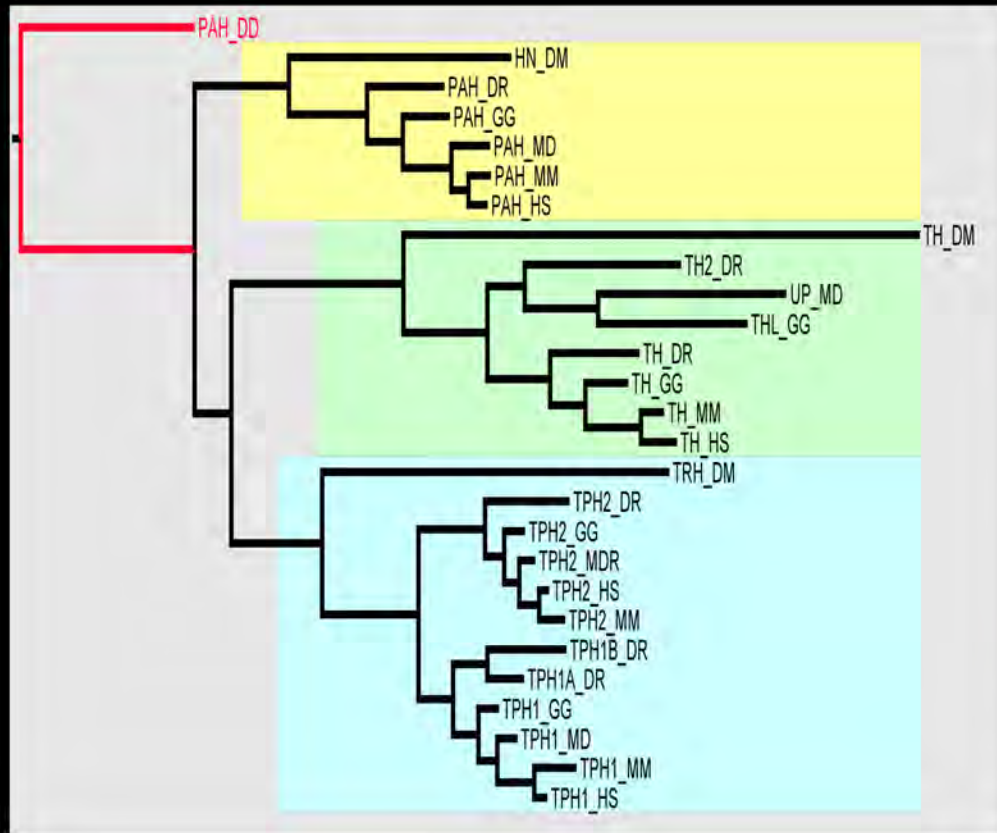
Source	Domain	Start	End
Pfam	ACT	35	100
disorder	n/a	66	67
Pfam	Bioperin_H	119	449
low_complexity	n/a	359	370

[Show](#) or [hide](#) domain scores.

A familiar sight?



Proteins have functional domains



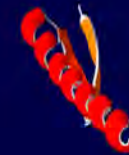
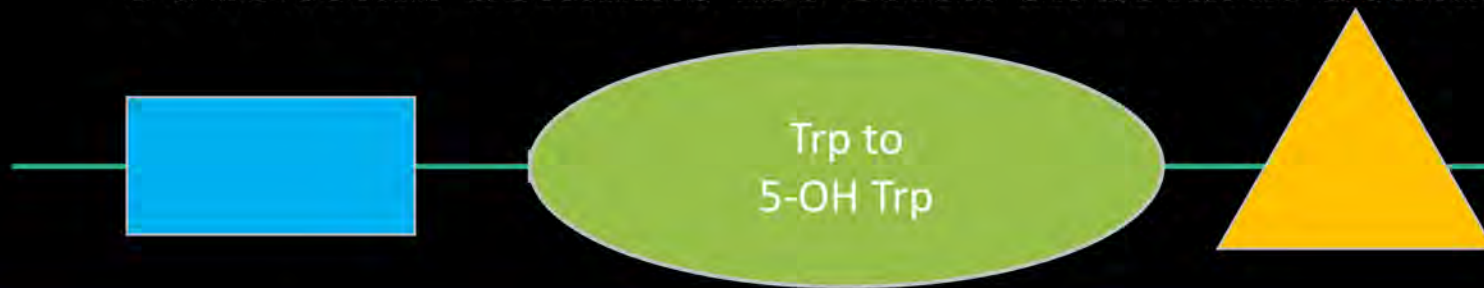
PFAM

Regulatory Catalytic Regulatory

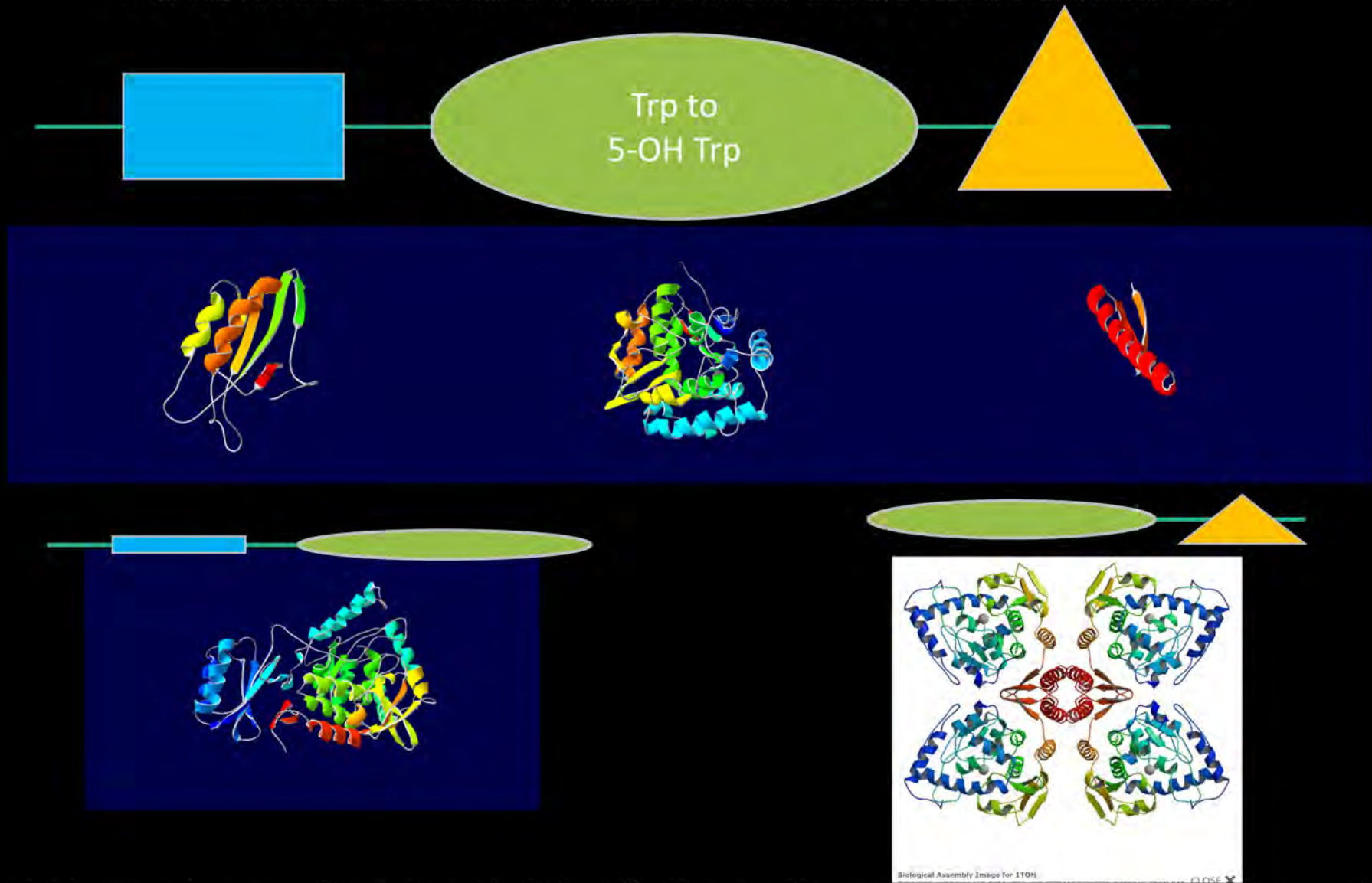


Functional domains are often structural domains.

Functional domains are often structural domains



Functional domains are often structural domains



Multidomains: often individual domains or pair combinations, but rarely full-length. The linkers between the domains are flexible and difficult to structurally determine.

Regulatory

Catalytic

Regulatory

TH: Tyrosine Hydroxylase



TPH: Tryptophan Hydroxylase



PAH: Phenylalanine Hydroxylase



Regulatory Catalytic Regulatory

TH: Tyrosine Hydroxylase



TPH: Tryptophan Hydroxylase



PAH: Phenylalanine Hydroxylase

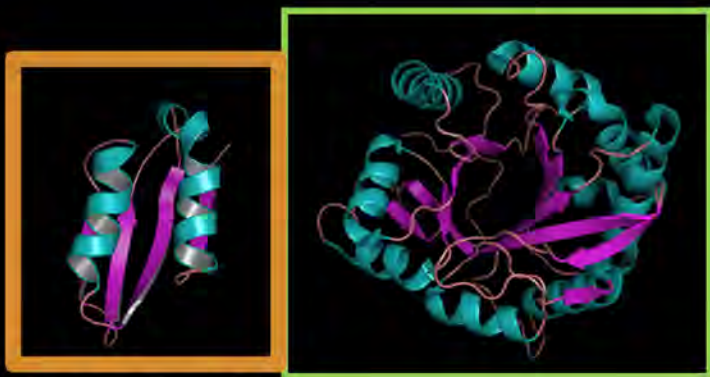


The main functional change (substrate specificity) is in the catalytic domain.

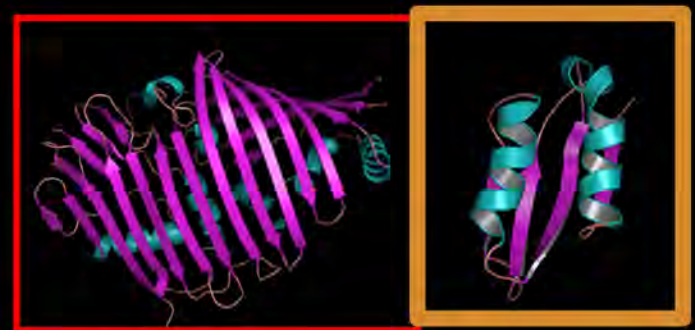
The structures of the domains have remained conserved, but there may be changes in regulation, domain packing, and protein dynamics.

The Domain – an evolutionary unit

- The smallest functional unit, basically structurally independent.
- Domains are units that reappear in different proteins.
- Many domains share a common ancestor, they are close or remote homologs.
- Homologs frequently have related functions.
- New domain combinations many yield new functions.



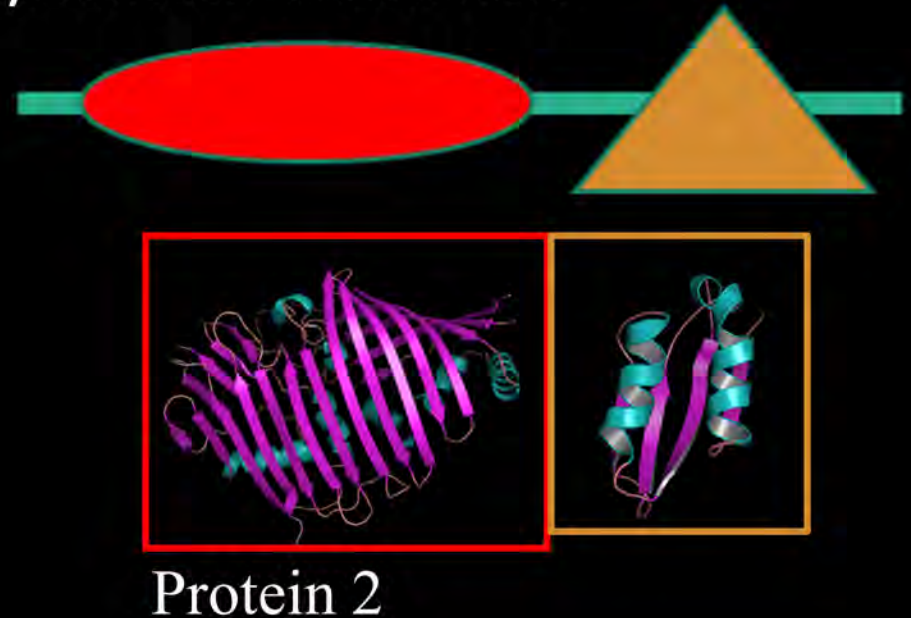
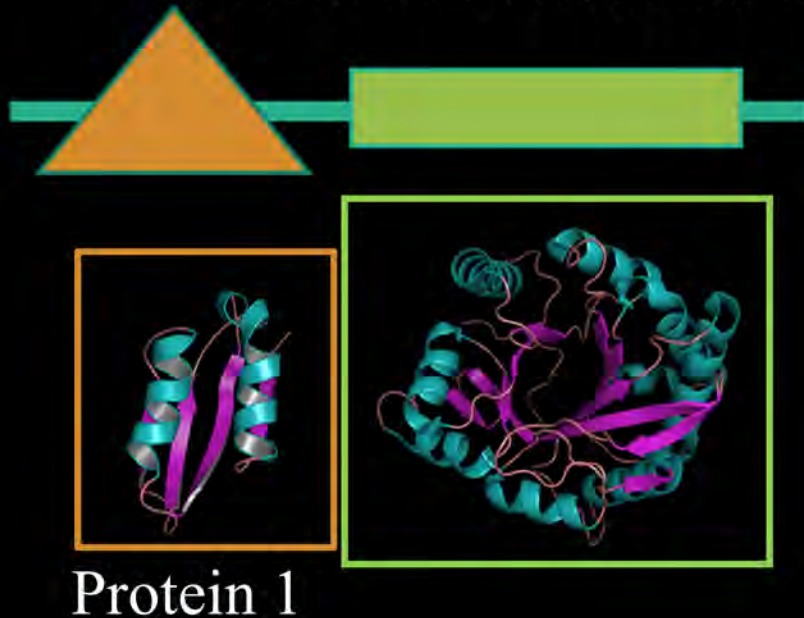
Protein 1



Protein 2

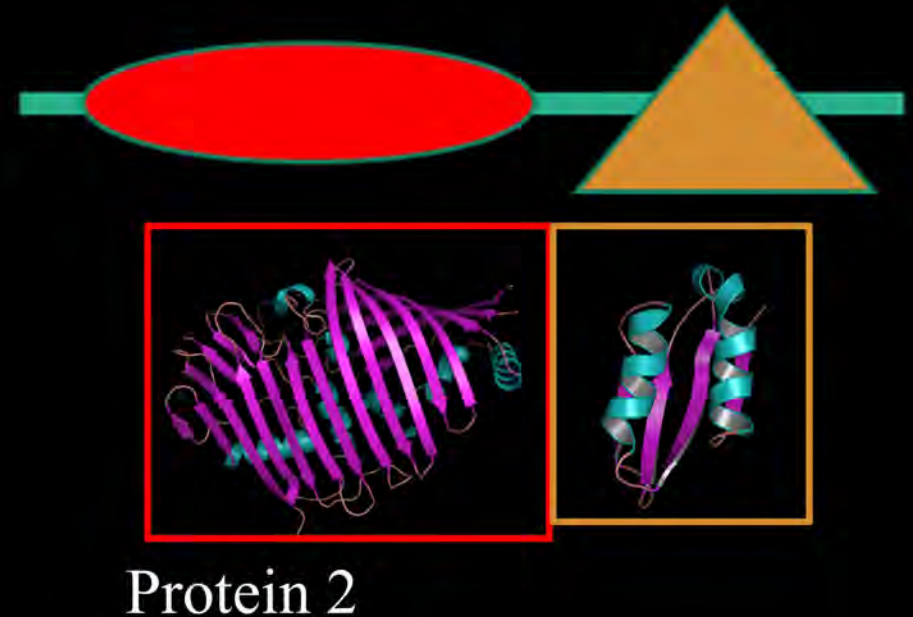
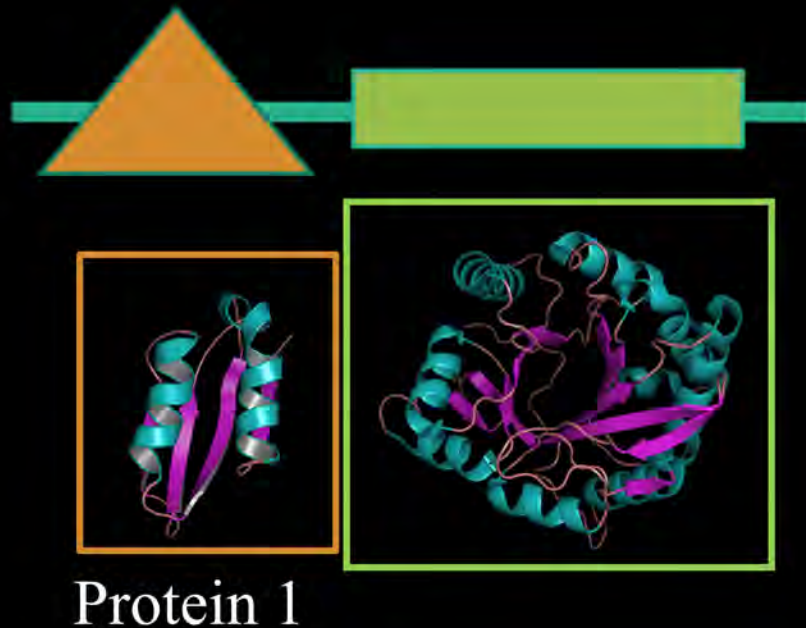
The Domain – an evolutionary unit

- The smallest functional unit, basically structurally independent.
- Domains are units that reappear in different proteins.
- Many domains share a common ancestor, they are close or remote homologs.
- Homologs frequently have related functions.
- New domain combinations many yield new functions.



New domain combinations yield new functions

- Despite being in different domain context and perhaps even having a different function, distantly related domains often have very similar arrangements and connectivity of protein secondary structure within the domain.



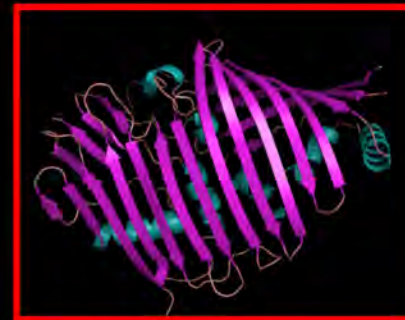
New domain combinations yield new functions

Proteins or domains with very similar arrangements and connectivity of protein secondary structure have the same

FOLD.



Protein 1

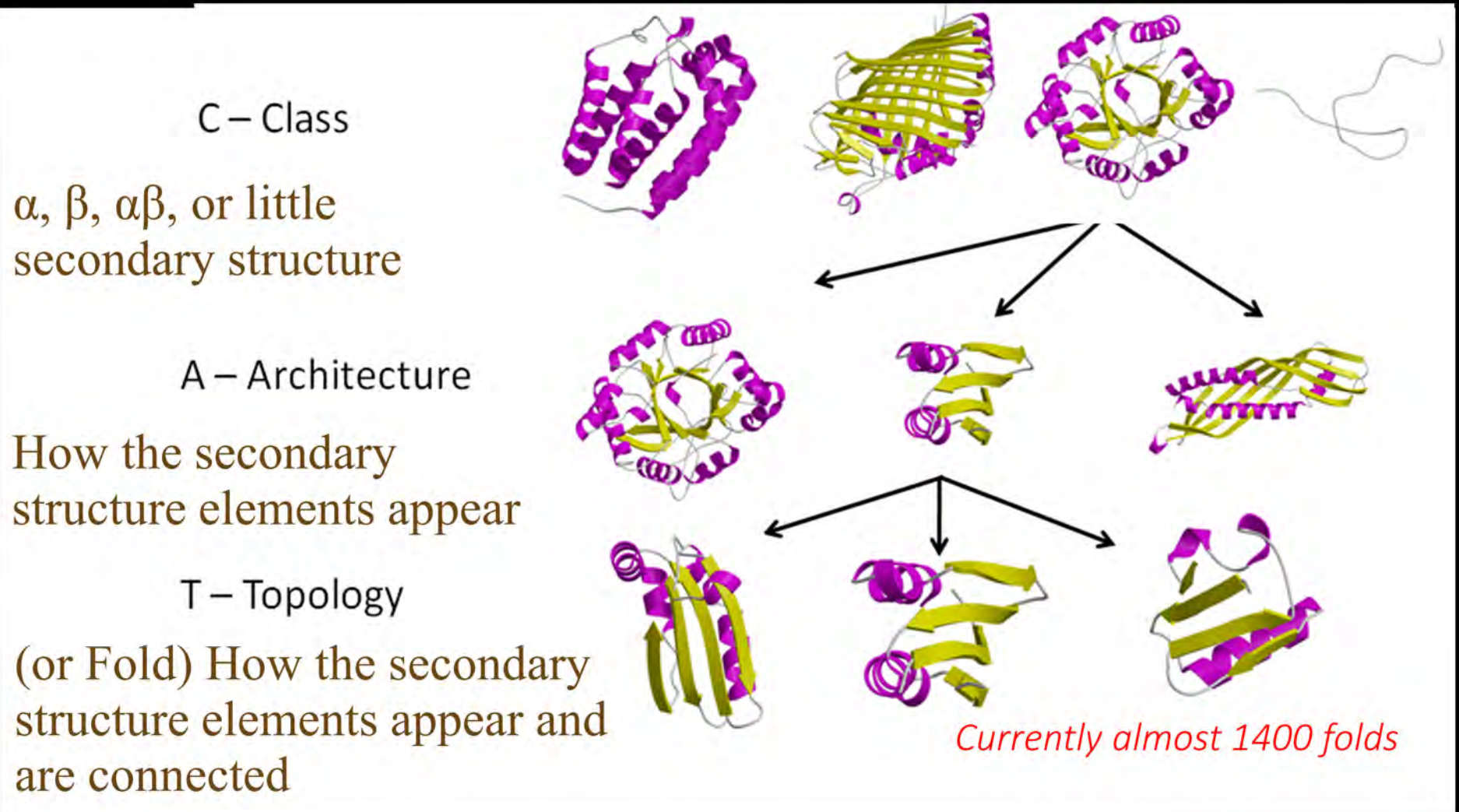


Protein 2

Protein Space

CATH:

Hierarchical classification of Protein Structure



Protein Space

CATH:

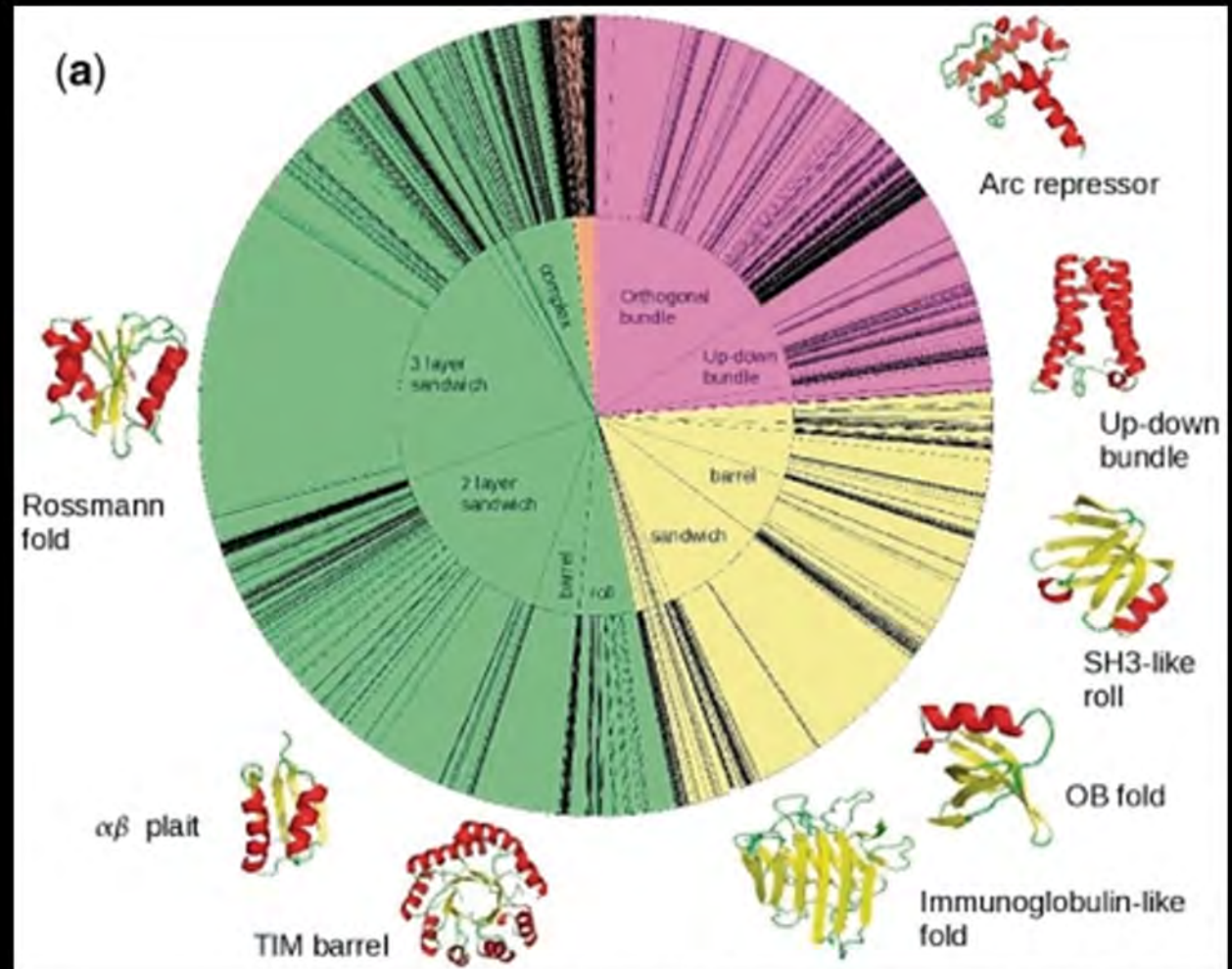
Hierarchical classification of Protein Structure

- C** Class: α , β , $\alpha\beta$, or little secondary structure
- A** Architecture: How the secondary structure elements
- T** Topology (or Fold) How the secondary structure elements appear and are connected
- H** Homologous superfamily: sometimes very remote (distantly related/very divergent sequences => sometimes too divergent to be identified as related)

Protein Space

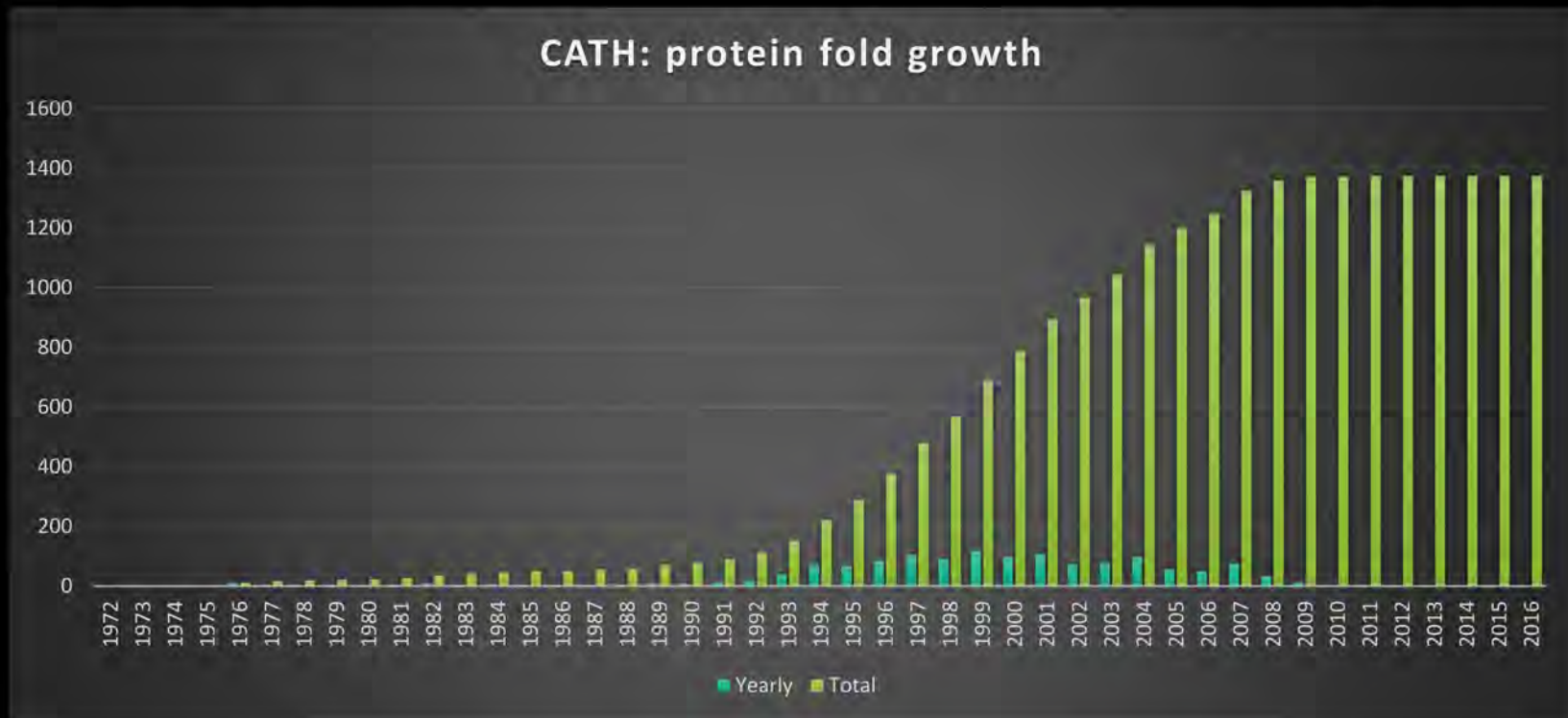
– Fold distribution

100 folds = 60% of
the 100,000 pdb
chains classified by
CATH



Cuff et al., Nucleic Acids Res. 2011; 39: D420–D426.

The number of structures in PDB have grown rapidly, but very few new folds have been discovered.



<http://www.rcsb.org/pdb/>

Structure is more conserved than sequence?

Protein Structure is assumed to be conserved among evolutionarily related proteins

- Many folds (usually rare folds) can not withstand many mutations.
- Some folds are highly designable: sequences can diverge far beyond recognition (below 20% sequence identity) but still fold into the same structure.



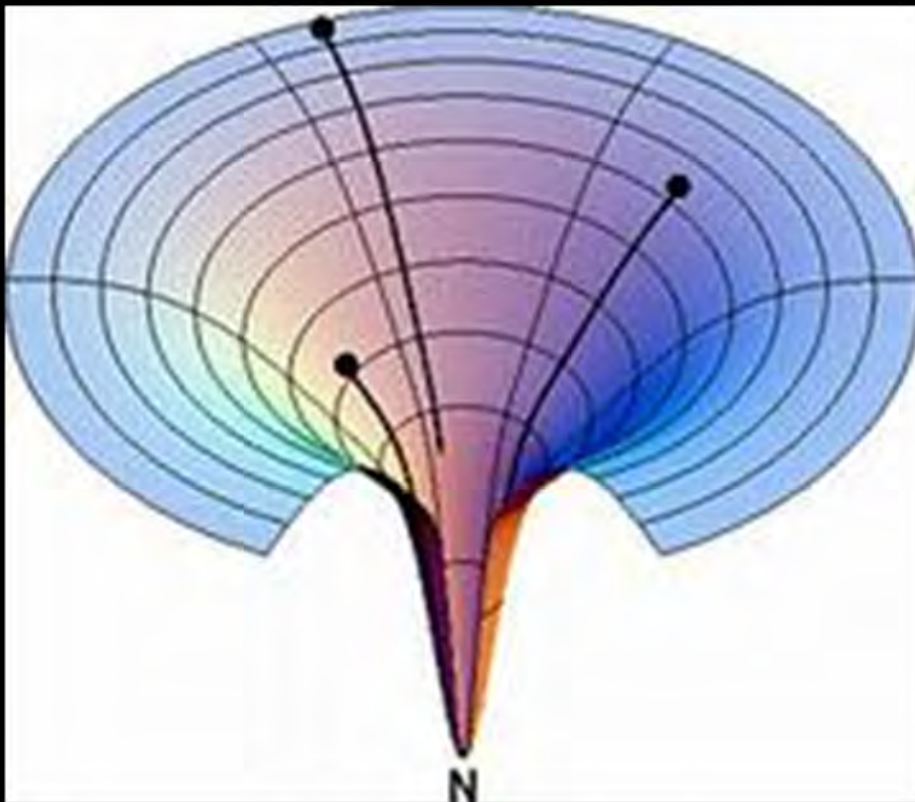
Myoglobin in gray

Beta subunit of hemoglobin in red

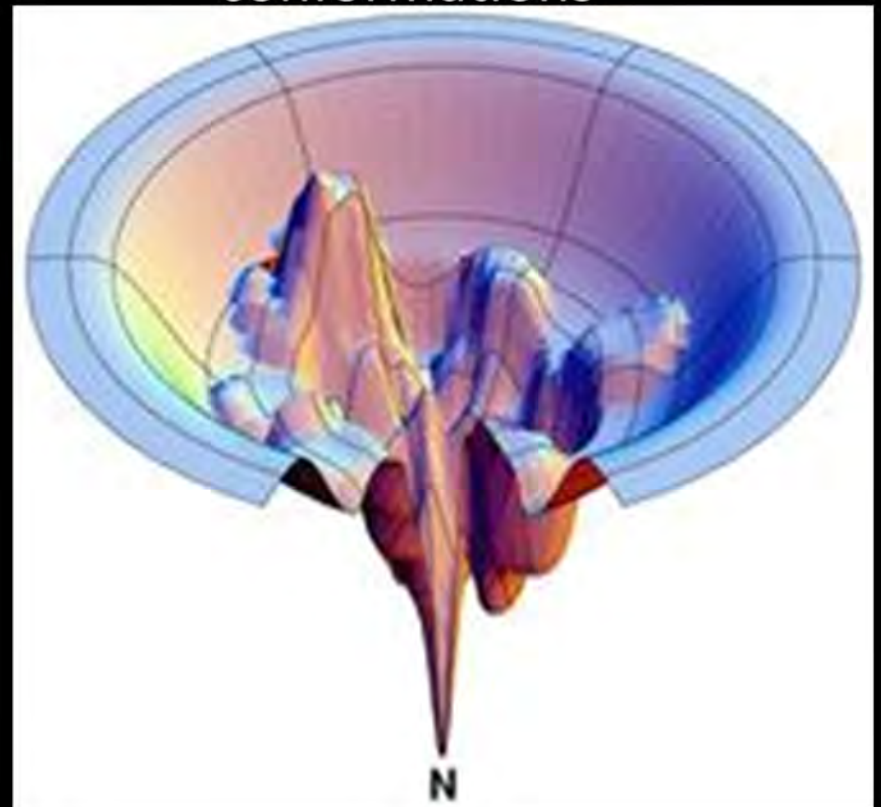
Sequence identity: 25%

Energy landscapes of protein folding

Smooth
only one minima



Rugged
local minima, alternative non-native
conformations

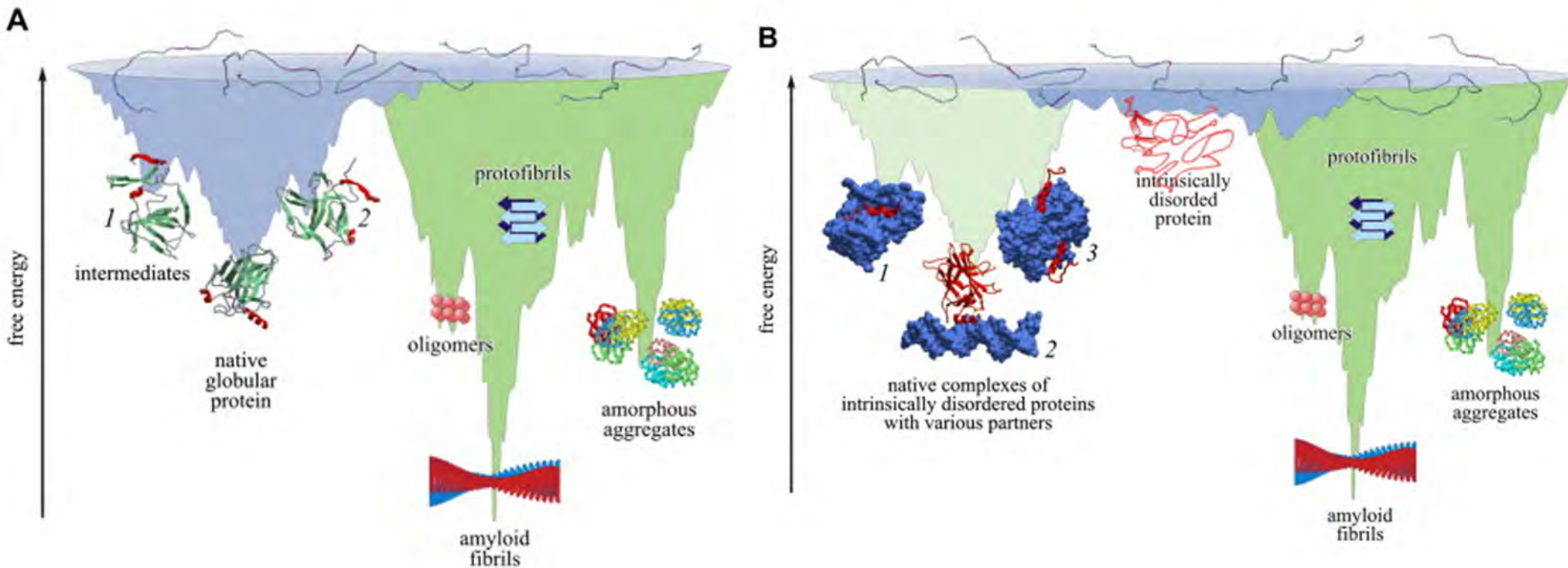


N = Native state (the folded conformation with the lowest energy).

<http://dillgroup.stonybrook.edu/#!/landscapes>

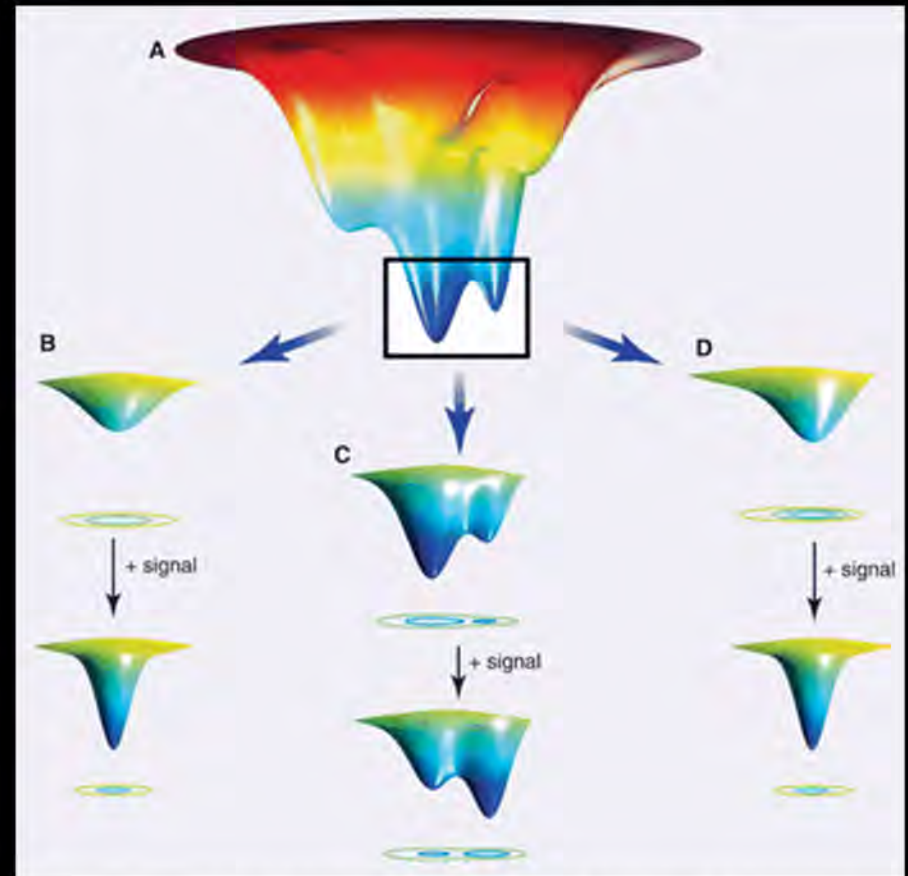
Protein space

Folding – energy landscapes



Proteins are dynamic and flexible

- The energy landscape can change in response to various signals (pH, ligands, allosteric effectors, interacting biomolecules, post-translational modifications, temperature... light)



Smock et al., Science, 324 (5924): 198-203

Protein structure space

- The fold and function determines the rate of sequence evolution.
- Different domains evolve at different rate.
- Protein structure evolves slower than protein sequence.
- Homologs often have similar structures.

How do we know this? We know this from the PDB data.

The PDB data will be covered in class, not in another video!