Bioinformatics

CS300

Domains according to UniProt and String

Fall 2019

Oliver BONHAM-CARTER

Proteins Fold Into Specific Structures for Functionality

Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins

Donald B. Wetlaufer

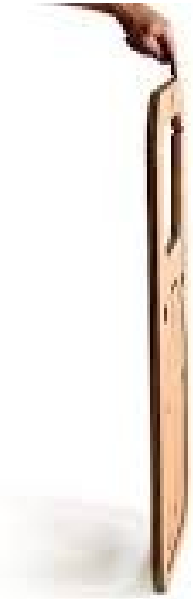
Wetlaufer, Donald B. "Nucleation, rapid folding, and globular intrachain regions in proteins." *Proceedings of the National Academy of Sciences* 70.3 (1973): 697-701.

Abstract

Distinct structural regions have been found in several globular proteins composed of single polypeptide chains. The existence of such regions and the continuity of peptide chain within them, coupled with kinetic arguments, suggests that the early stages of three-dimensional structure formation (nucleation) occur independently in separate parts of these molecules. A nucleus can grow rapidly by adding peptide chain segments that are close to the nucleus in aminoacid sequence. Such a process would generate three-dimensional (native) protein structures that contain separate regions of continuous peptide chain. Possible means of testing this hypothesis are discussed.

Different regions in same protein (*domains*) performing specific tasks.

Structures For Functions



Structures for Functions

Windows to allow driver to see out while driving

Ventilation for cooling

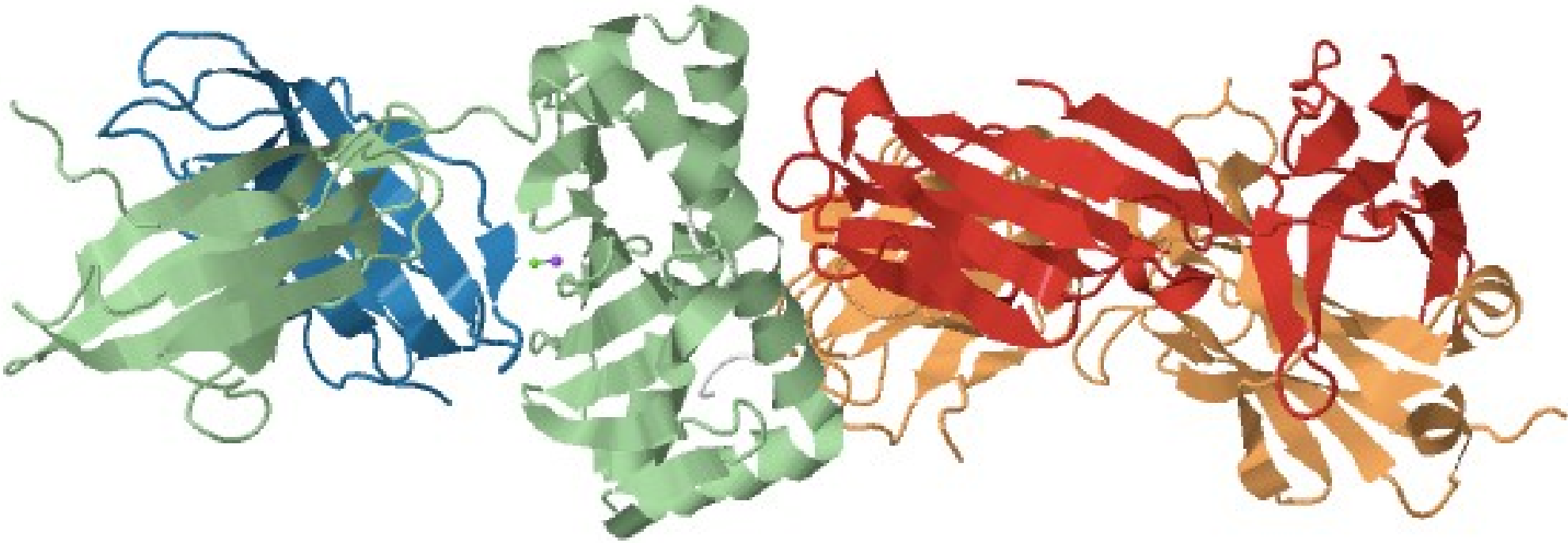
Headlights for driving at night

License plate: for Identification

Door for letting driver into the car

Wheels, necessary for mobility

Proteins Also Have Specific Functional Regions, Too!



Protein Data Bank:

5WLG

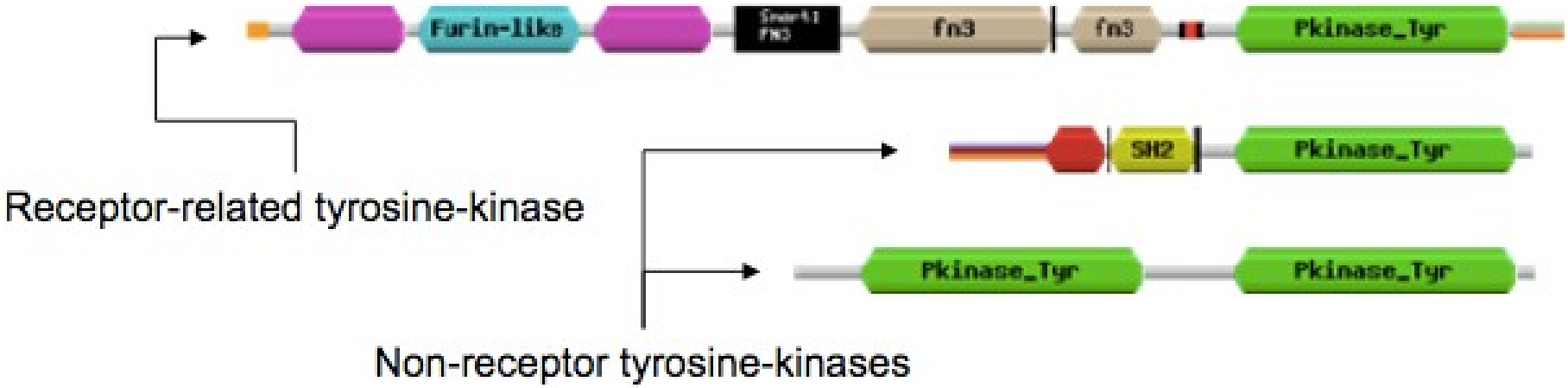
Domains

- A protein **domain** is a conserved part of a given protein sequence and (tertiary) structure.
- Can evolve, function, and exist independently of the rest of the protein chain
- Each domain forms a compact three-dimensional structure
- Often can be independently stable and folded.

SMART domain 'bubblegram' for human fibroblast growth factor (FGF) receptor 1 (type P11362 into web site: smart.embl.de)

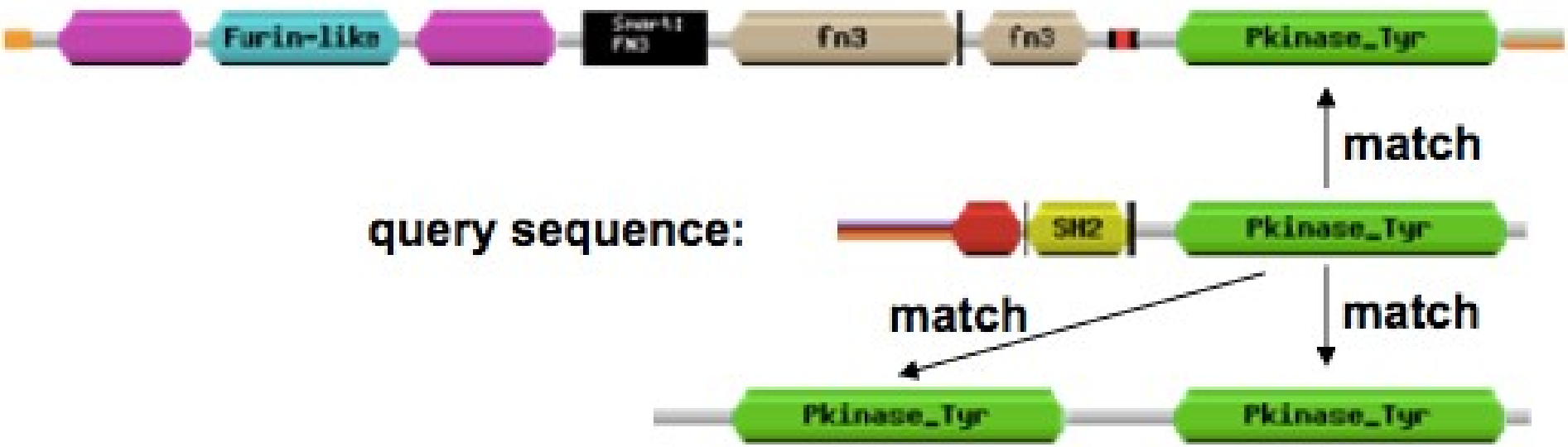
Protein Modularity

- Discrete functional units
- Found domains do not occur in the same order across proteins.
- Domains are considered separately in protein function predictions



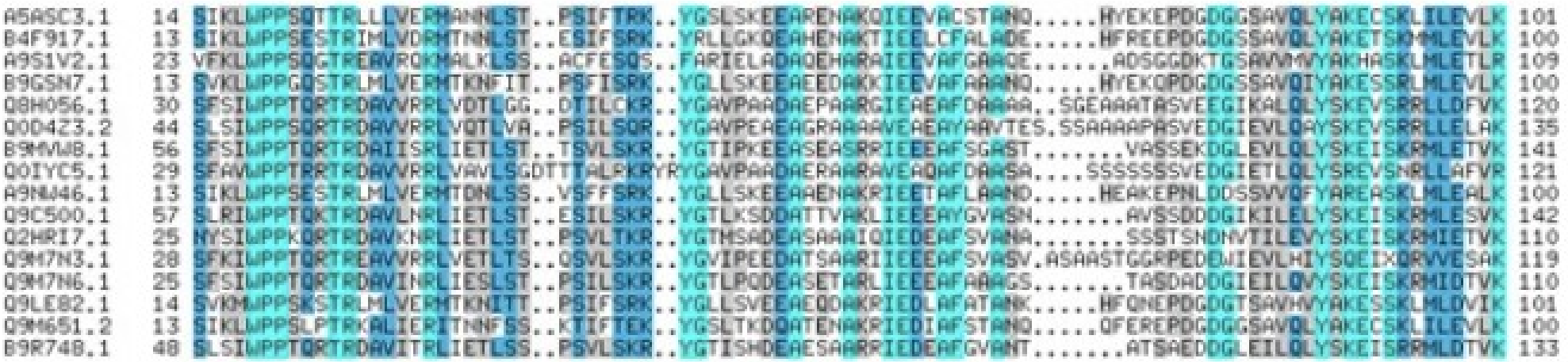
Finding a Domain?

- Alignment across proteins may show domains
- Use databases to match similar parts of proteins
  - Pfam, Smart, Interpro and others



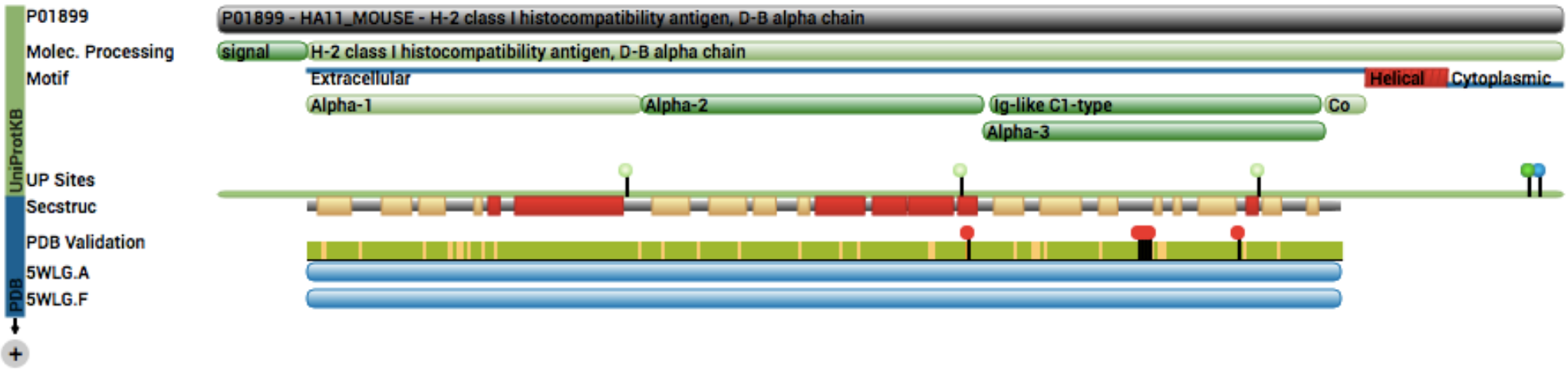
Alignment?!

- Provide more info about a protein’s family, relatedness and other details.
- Domain landmarks include: low-complexity or disorder to suggest that these regions may have a specific syntax or pronounced grammar.



Domains By PDB

- Domains give the protein special qualities:
  - Domain Names: *Alpha1, Alpha2, Alpha3, Ig-like C1-type*



Protein Data Bank ID: 5WLG

  - This protein:
    - https://www.rcsb.org/pdb/explore/explore.do?structureId=5WLG
  - Help with features
    - https://www.rcsb.org/pages/help/featureView

Domains By Uniprot

- Domains give the protein special qualities:
  - Domain Names: *Alpha1, Alpha2, Alpha3, Ig-like C1-type*

Family & Domains[i]

Domains and Repeats

Feature key	Position(s)	Description	Actions	Graphical view	Length
Domain[i]	209 – 297	Ig-like C1-type	Add BLAST		89

Region

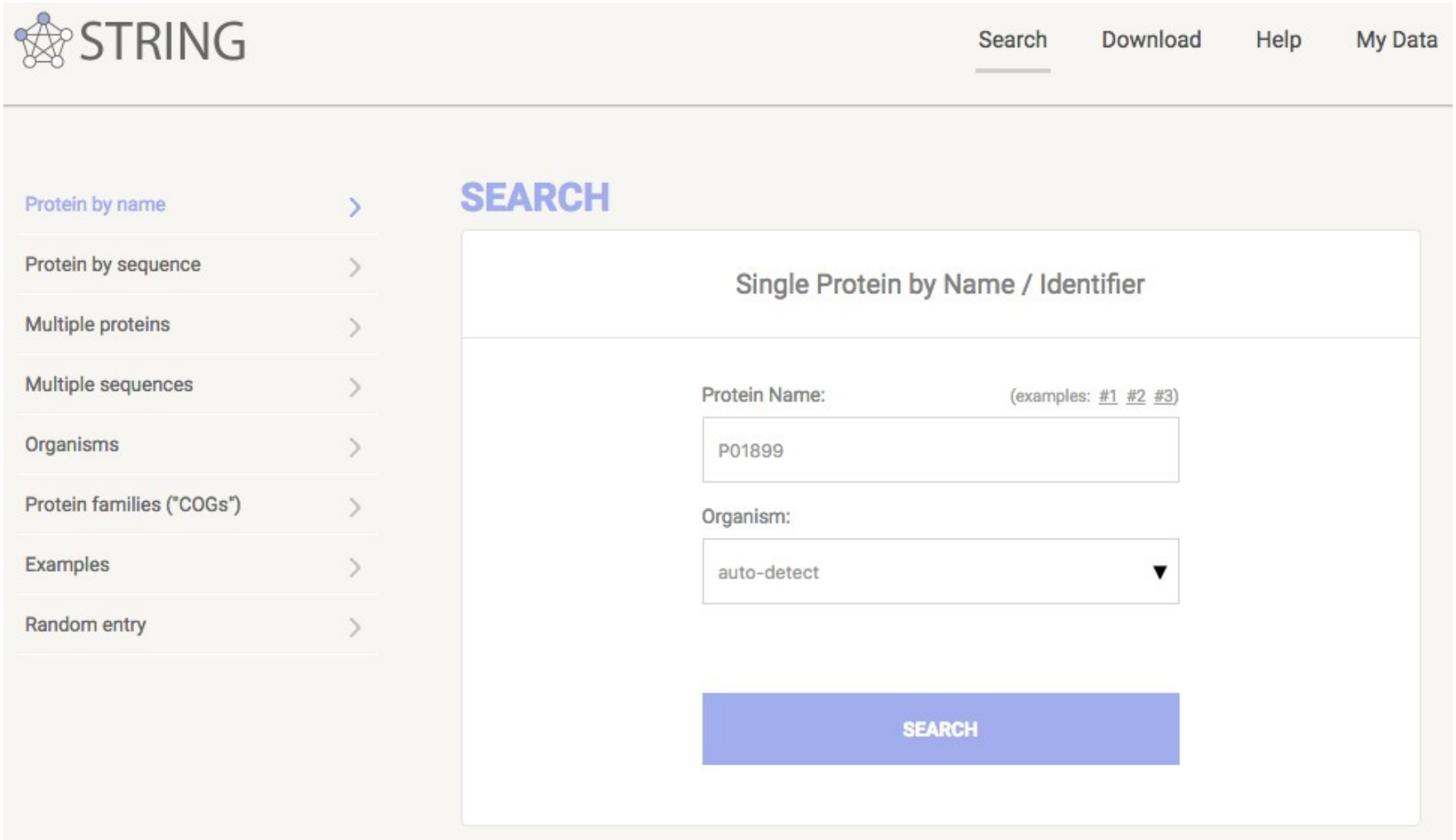
Feature key	Position(s)	Description	Actions	Graphical view	Length
Region[i]	25 – 114	Alpha-1	Add BLAST		90
Region[i]	115 – 206	Alpha-2	Add BLAST		92
Region[i]	207 – 298	Alpha-3	Add BLAST		92
Region[i]	299 – 309	Connecting peptide	Add BLAST		11

UniProt ID: P01899

A Protein Knowledge Base

http://www.uniprot.org/uniprot/P01899#family_and_domains

STRING: Functional Protein Association Networks

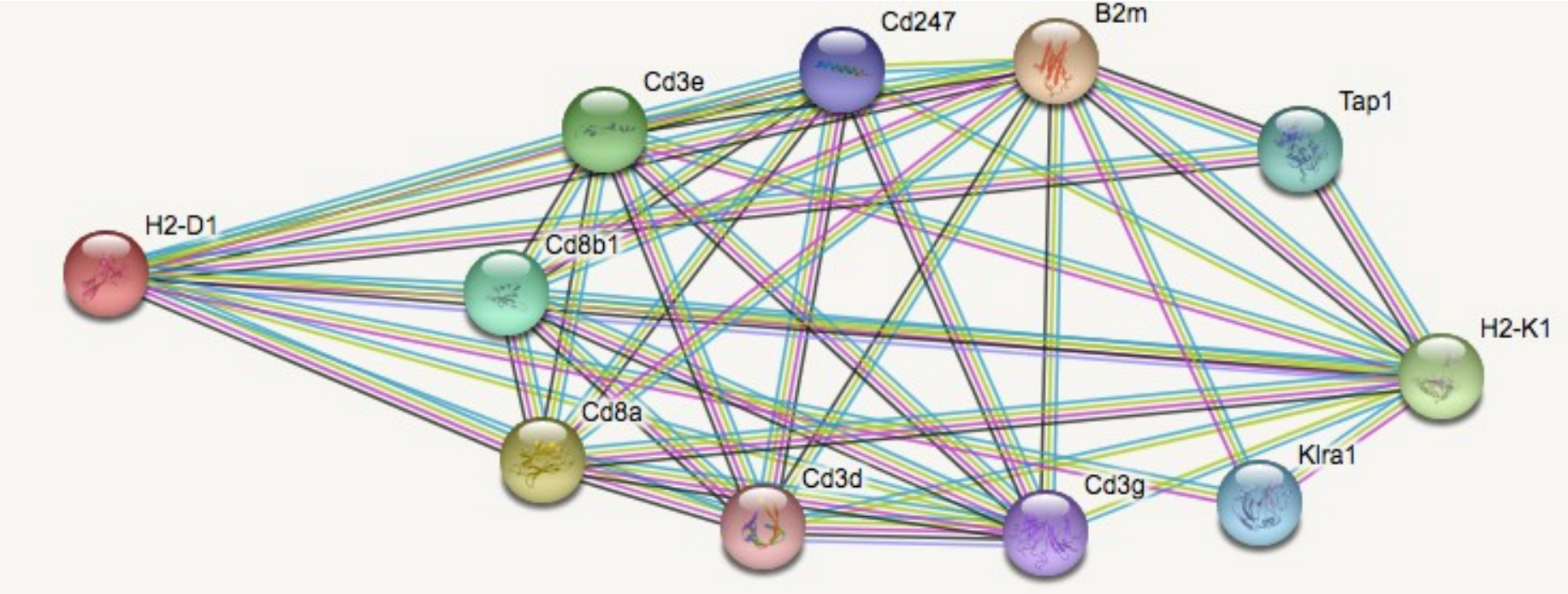


String DB ID P01899

http://string-db.org/

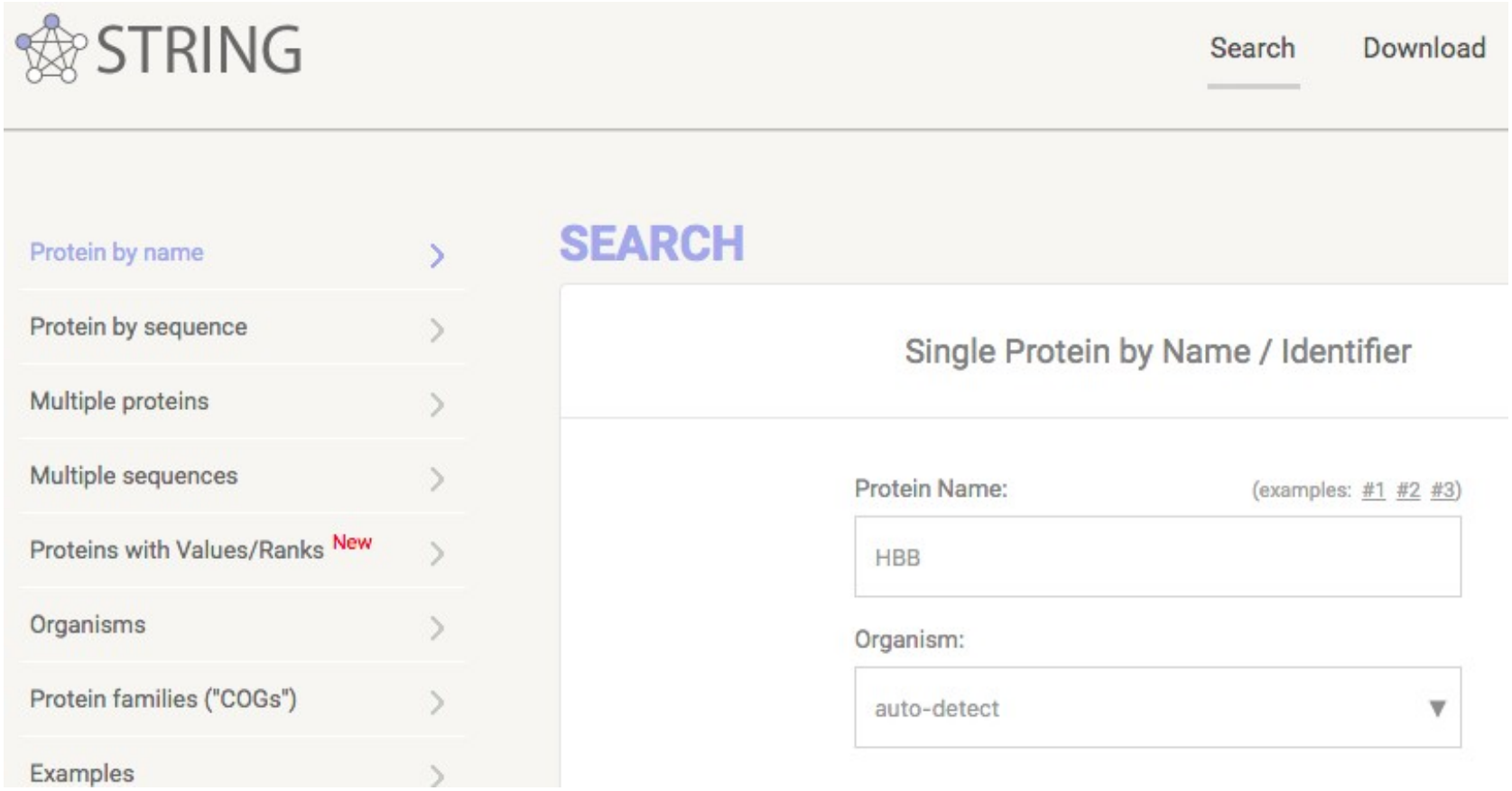
STRING: Functional Protein Association Networks

- Known and predicted protein-protein interactions
- How does a protein interact with others?
- What types of interactions are these (across all known genomes, of any organism)?



STRING: Functional Protein Association Networks

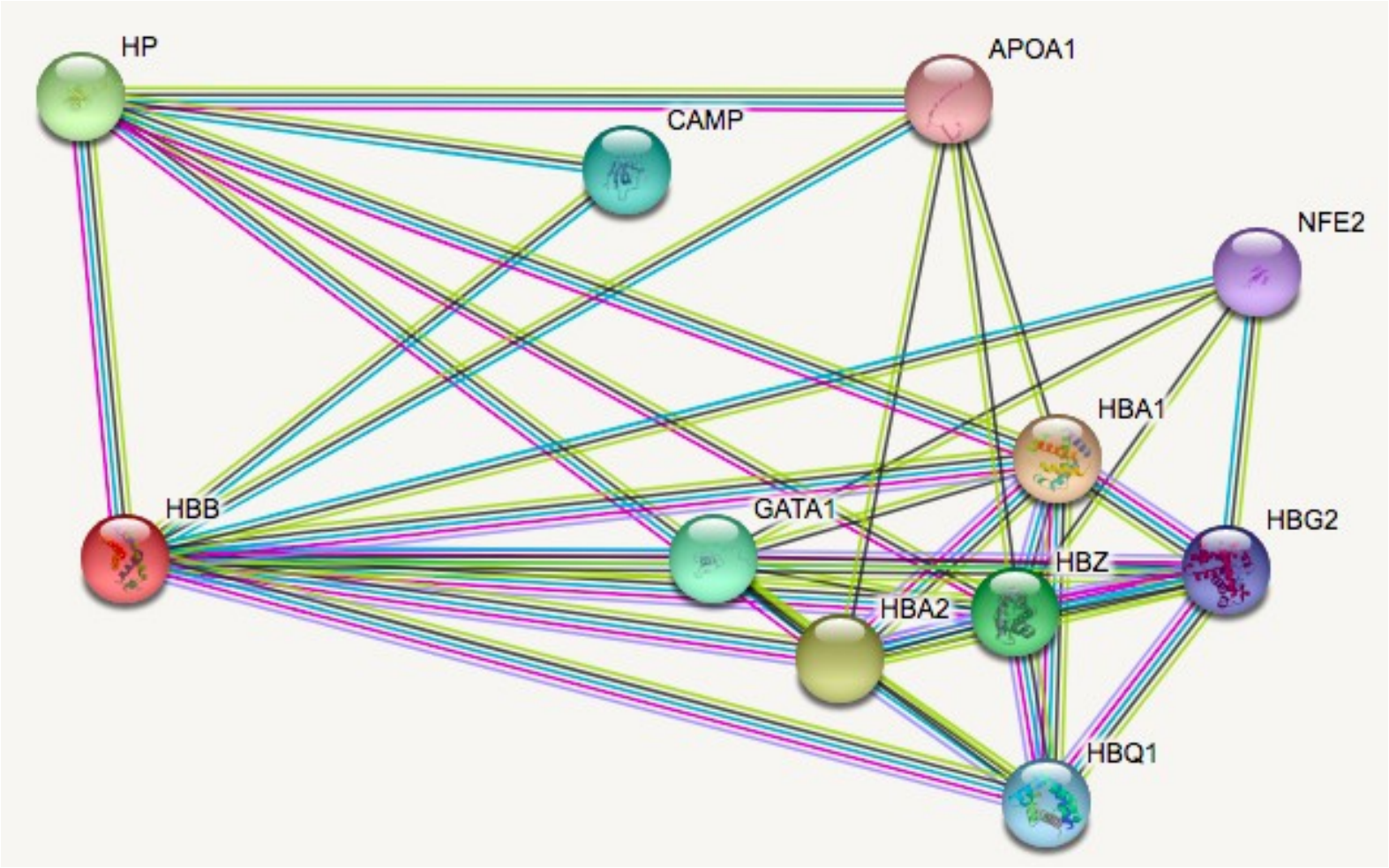
- Question: What proteins (from genes) interact with **HBB** protein (from the gene)?



https://string-db.org/

STRING: Functional Protein Association Networks

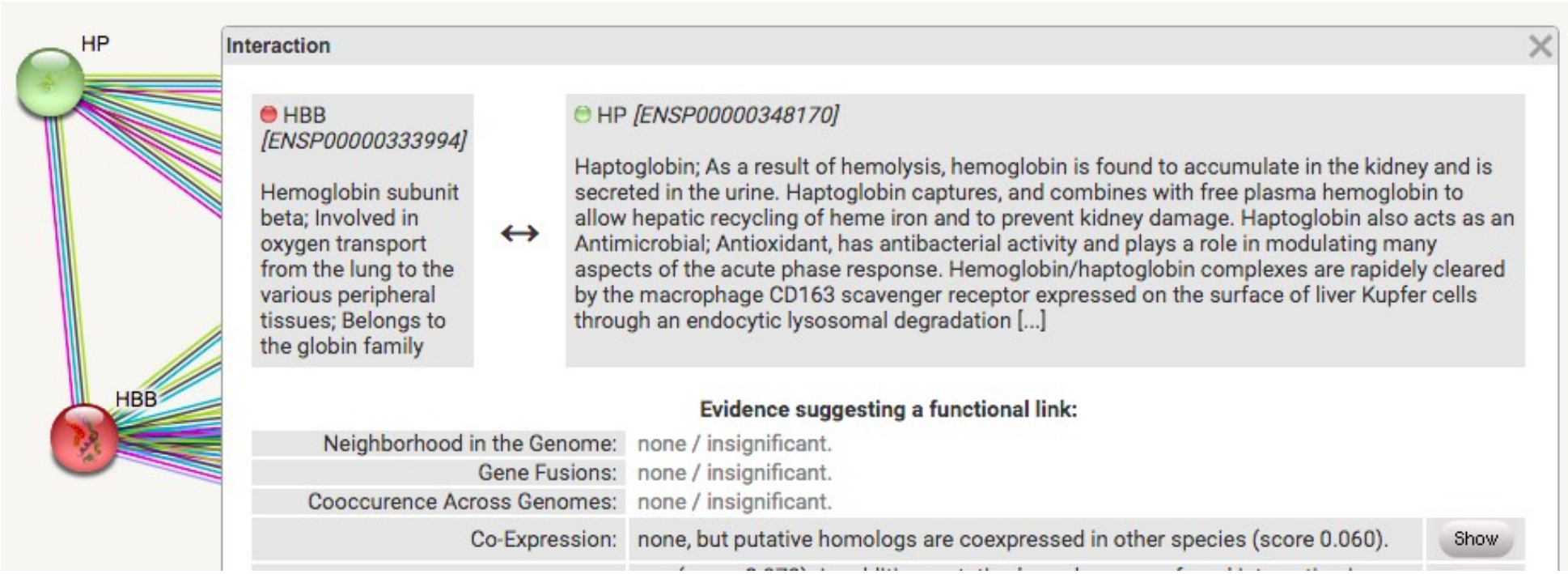
- Answer: Lots!



https://string-db.org/

STRING: Functional Protein Association Networks

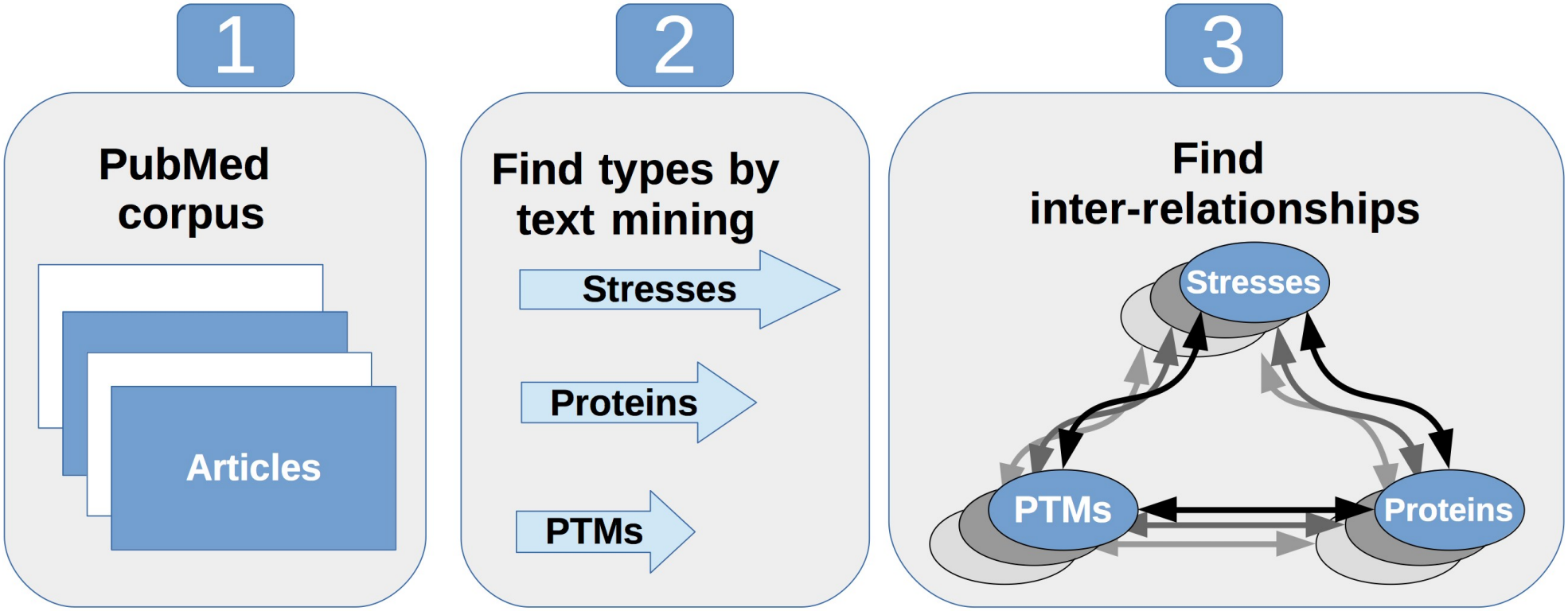
- What kinds of interactions?



https://string-db.org/

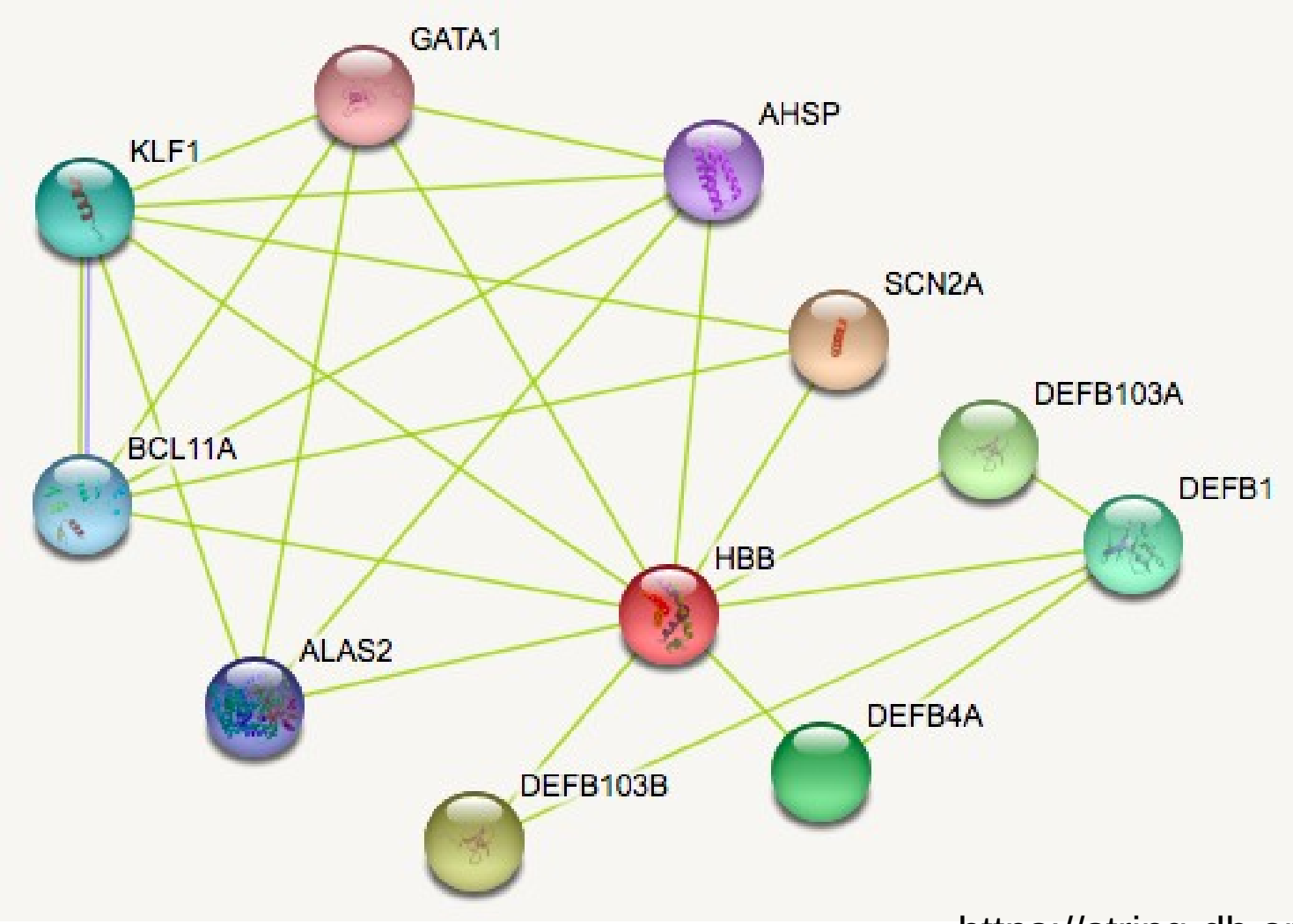
Criteria to Determine Relations

- There are many ways to measure the distance between two different proteins
  - Text Mining



String: by Text Mining

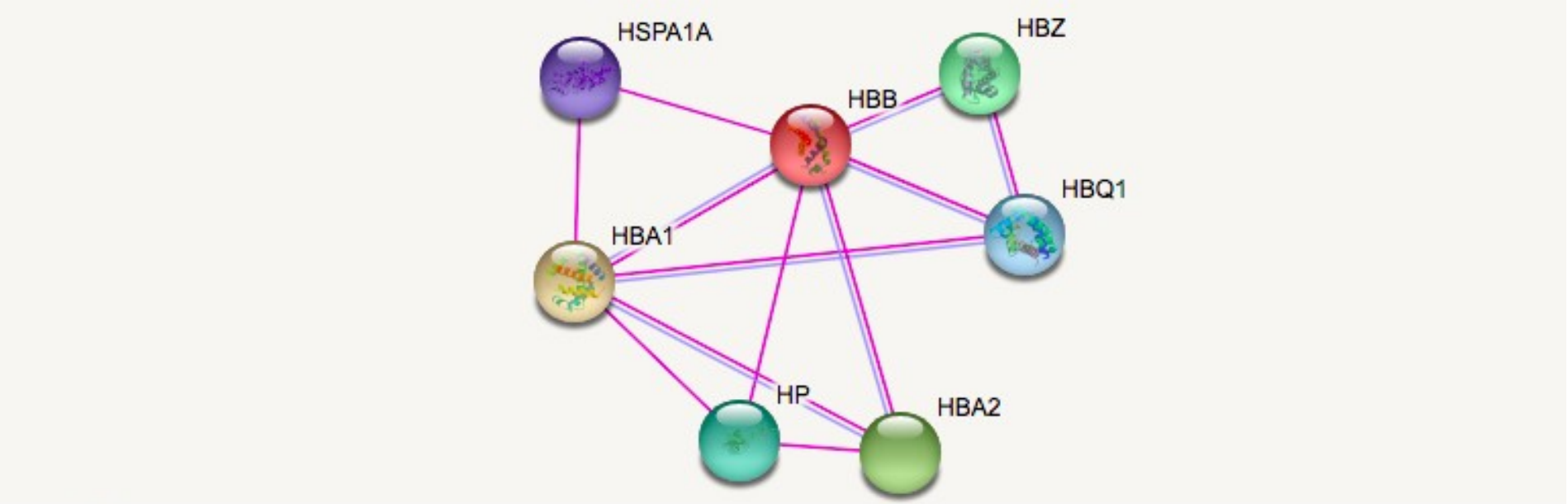
- HBB's interactions according to the literature



https://string-db.org/

String: Linked Experimentally

- Experiments performed to show that protein are related



Interaction

HBB *[ENSP00000333994]*

Hemoglobin subunit beta; Involved in oxygen transport from the lung to the various peripheral tissues; Belongs to the globin family

↔

HBZ *[ENSP00000252951]*

Hemoglobin subunit zeta; The zeta chain is an alpha-type chain of mammalian embryonic hemoglobin

String: Linked Experimentally

- Learn about the experiments

LAB EXPERIMENTS

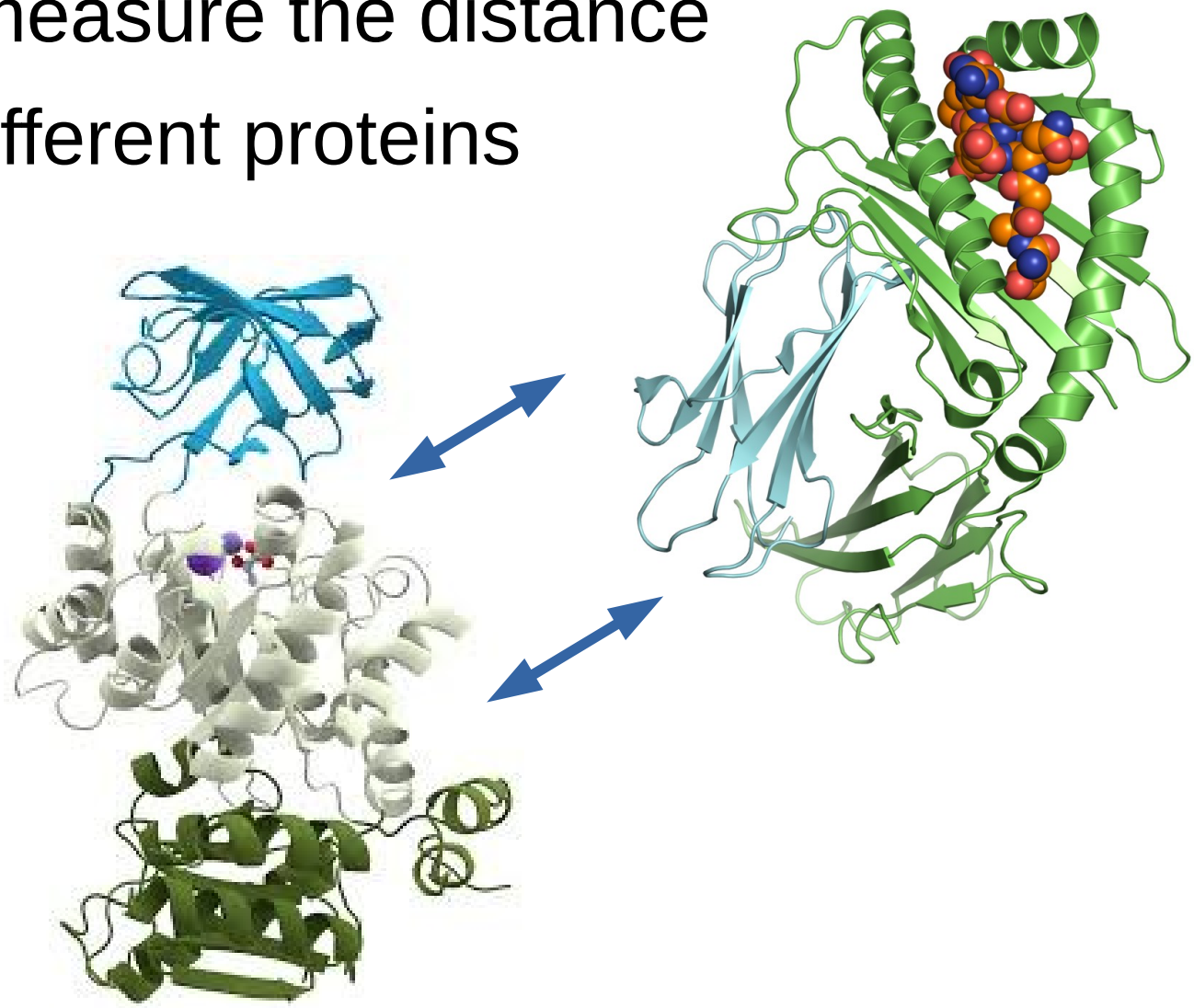
Relevant datasets in Mus musculus:

protein-protein interaction (intact) *Detected by psi-mi:"MI:0027"(cosedimentation) assay*	H2-D1 B2m [... and 1527 other proteins]
protein-protein interaction (mint) *Detected by psi-mi:"MI:0027"(cosedimentation) assay*	H2-D1 B2m [... and 1527 other proteins]
protein-protein interaction (dip) *Detected by x-ray crystallography assay*	H2-D1 B2m
protein-protein interaction (intact) *Detected by psi-mi:"MI:0114"(x-ray crystallography) assay*	H2-D1 B2m

Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling.
Kislinger T, Cox B, Kannan A, Chung C, Hu P, Ignatchenko A, Scott MS, Gramolini AO, Morris Q, Hallett MT, Rossant J, Hughes TR, Frey B, Emili A
Cell. 125(1):173-86 (2006).

Criteria to Determine Relations

- Other ways to measure the distance between two different proteins
  - Neighborhood
  - Experiments
  - Databases
  - Co-Expression
  - And others...



More Information?

- Unify the representation of gene and gene product attributes across all species information
  - **AmiGO 2: Gene ontology**
    - http://amigo.geneontology.org/amigo/landing

- Information of effects of genetic variation on human health
  - **Genetics Home Reference**
    - https://ghr.nlm.nih.gov/

Header

- Pick your favorite protein and head-over to:
  - http://www.uniprot.org/

    example: P01899, gene name: H2-D1
- Then check out the networks at:
- https://string-db.org/