



Bioinformatics in Protein Science

Jesús Alvarado Valverde

Toby Gibson's group

Predoc course 2019
November 11th / 13th

EMBL



Gibson Team



Toby Gibson

Jesús



Lena



Hugo

Malvika – BioIT

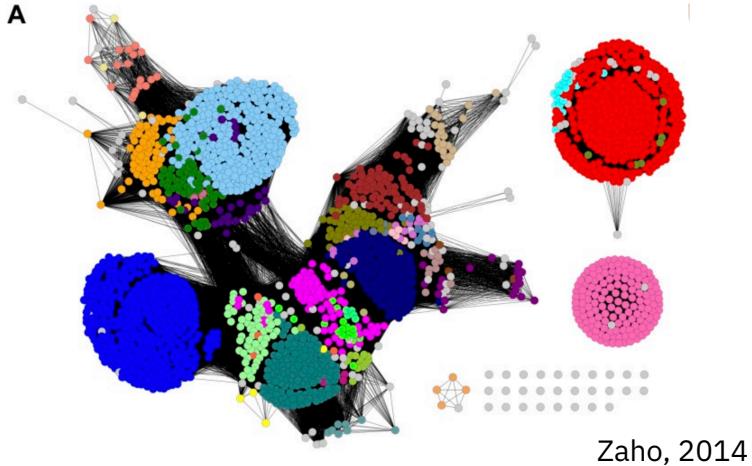


Manjeet – ELM database

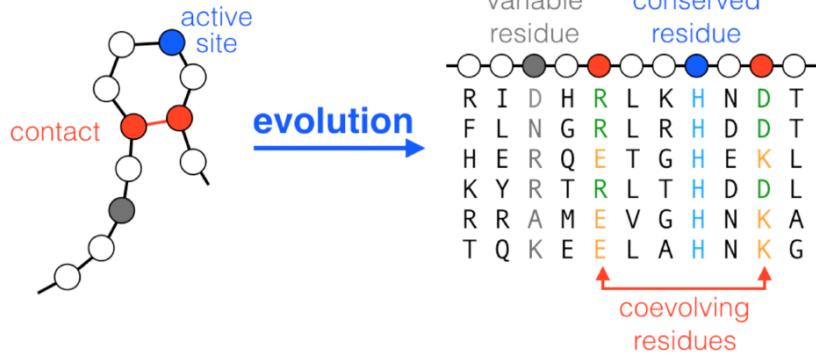


Elizabeth (visitor)

Previous experience

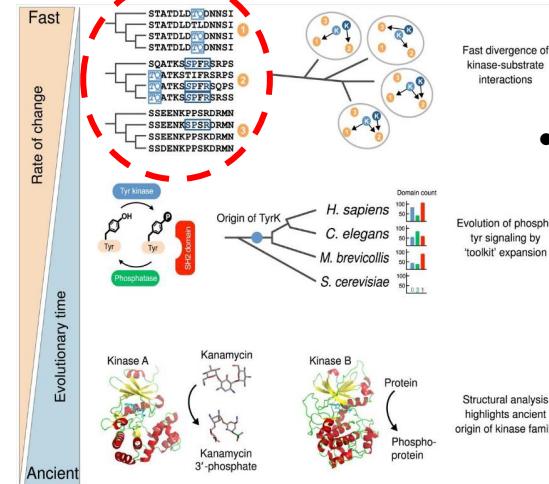


- Sequence Similarity Networks

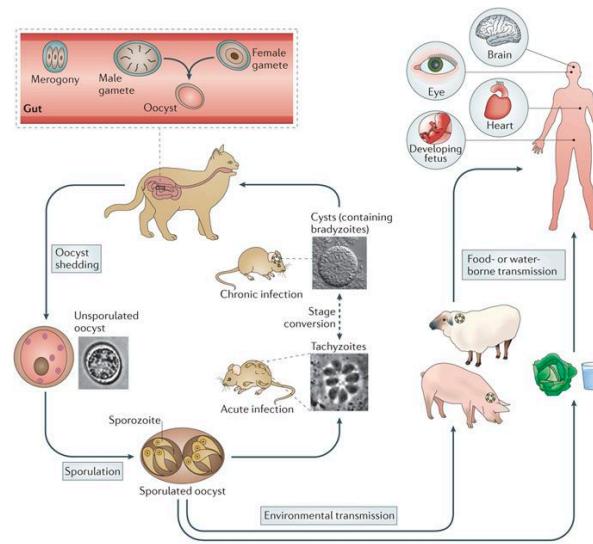


- Hsp90 – DCA & Molecular dynamics

Cocco, et al. 2017



(Beltrao et al., 2013)

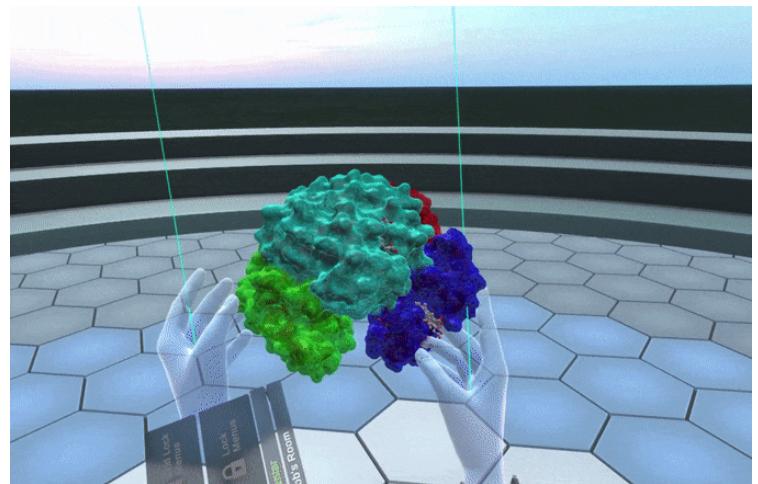


- Toxoplasma Motif prediction

(Hunter & Sibley, 2012)

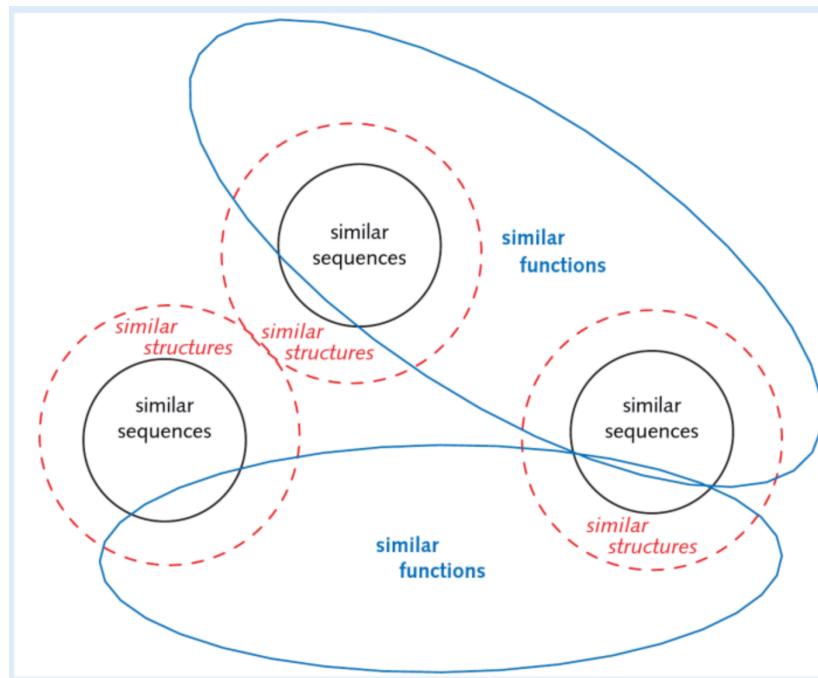
Nature Reviews | Microbiology

Protein Science: studying the invisible



Protein sequence determines structure

The primary amino acid sequence of polypeptides contains all of the necessary information to direct their folding into functional native proteins (Anfinsen, 1973).

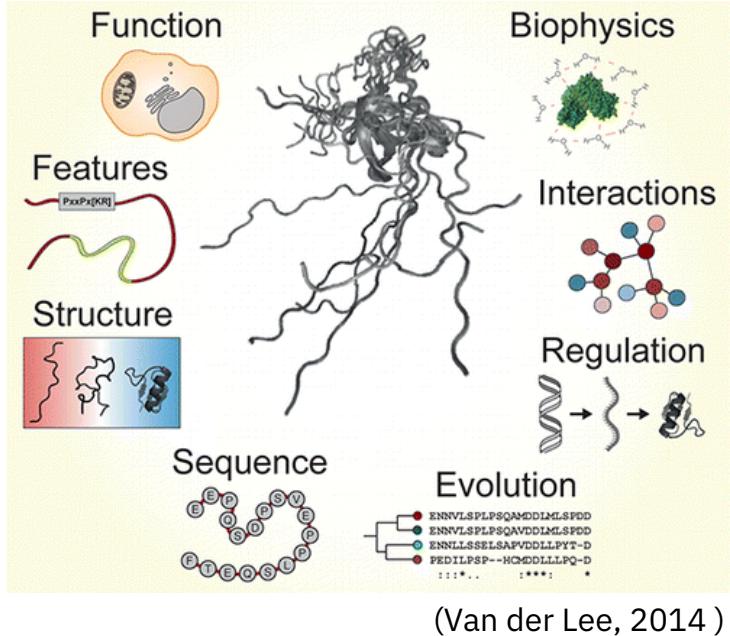
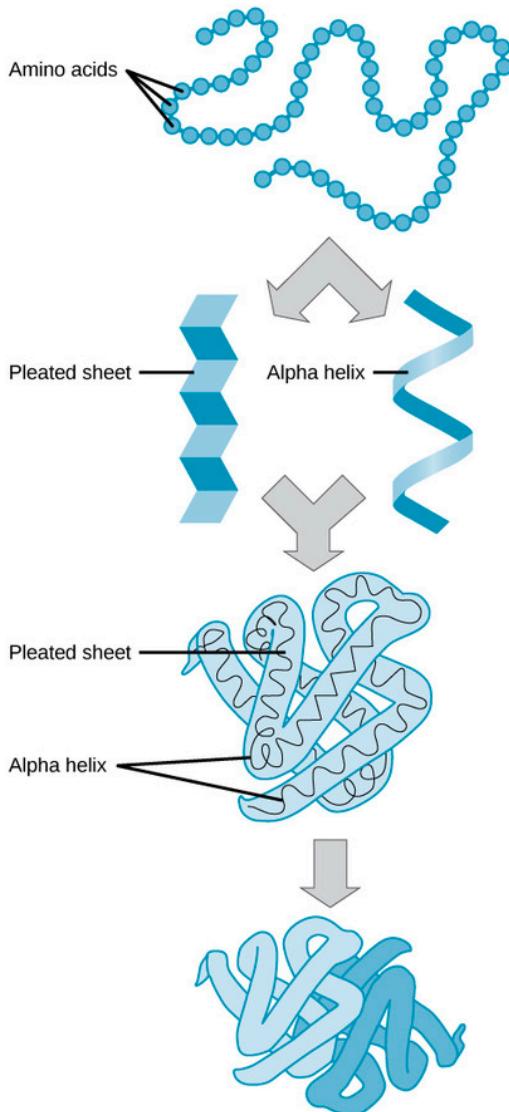


Sequence – Structure – Function – Interactions
Each is a different field

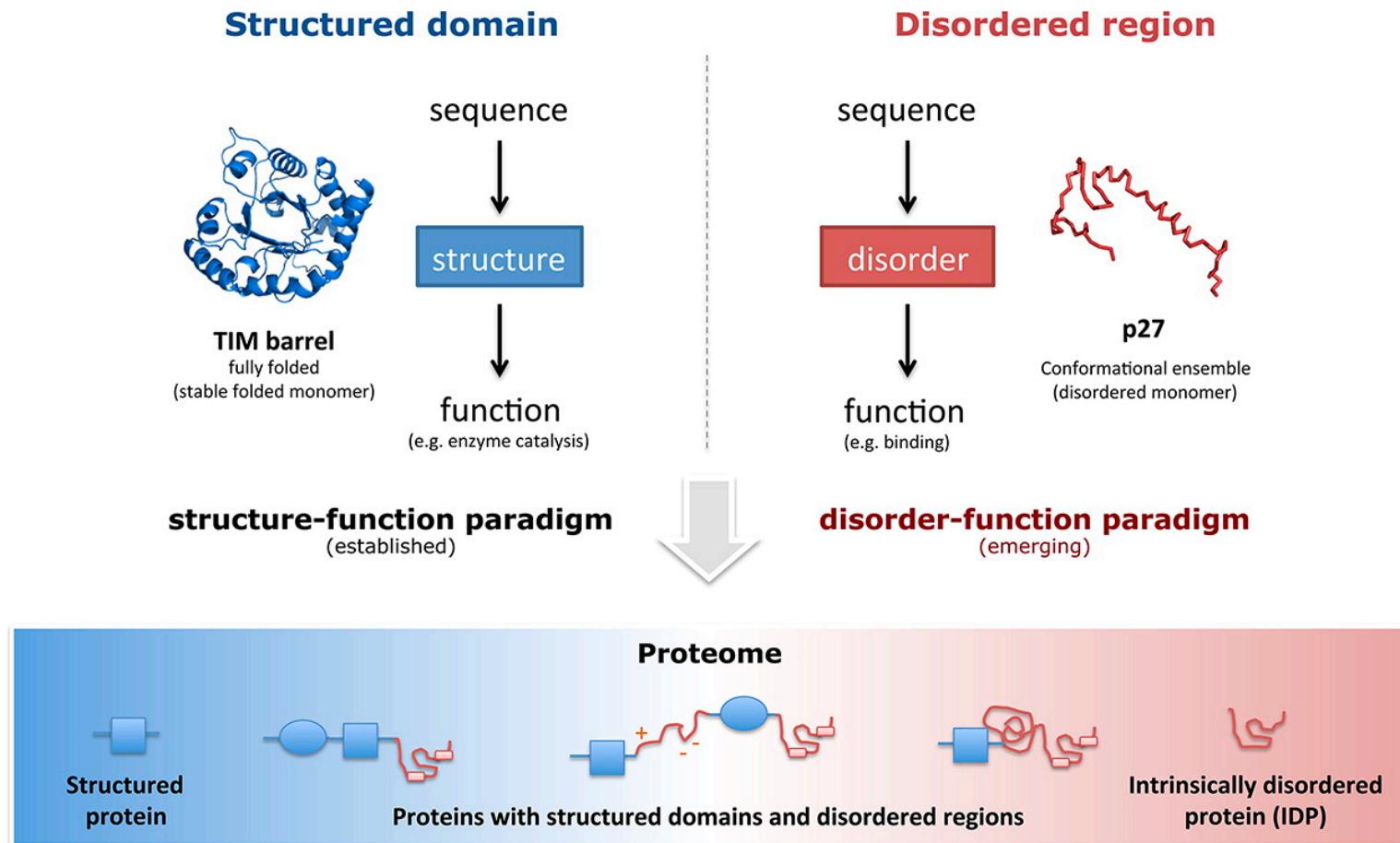
Protein sequence determines structure

Sequence analysis:

- Secondary structure
- Domains
- Disorder
- Homology
- Phylogeny
- Etc.



Protein disorder is widespread



(Van der Lee, 2014)

Can you think of any protein
with functional disorder regions?

EMBL



Practical

How do we collect and compare proteins?

Databases

Software

Ontologies

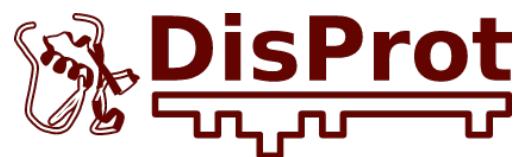
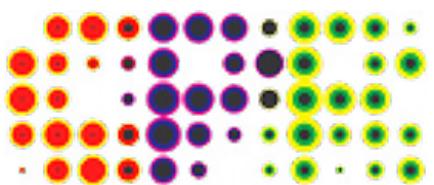


Software and resources



Pfam

A grey molecular network icon consisting of several nodes connected by lines, followed by the text "STRING".



GROMACS
FAST. FLEXIBLE. FREE.



UniProt



EMBL

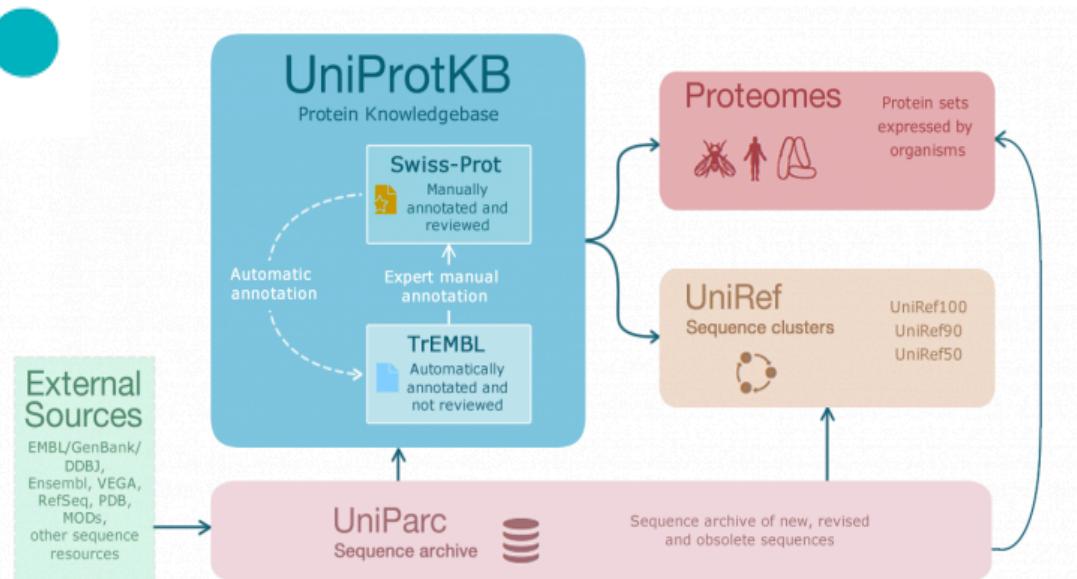


A universe of protein information

<https://www.uniprot.org/>



The database does not contain everything that is known about proteins!!!

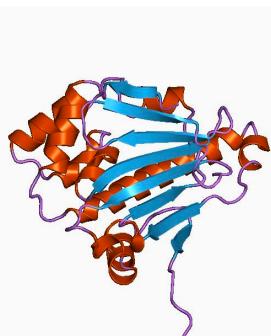


Protein comparison

Organize in teams of two



HtpG
E. coli
624 aa



Go to **Uniprot**

Search for: **P07900**

Search for: **P0A6Z3**



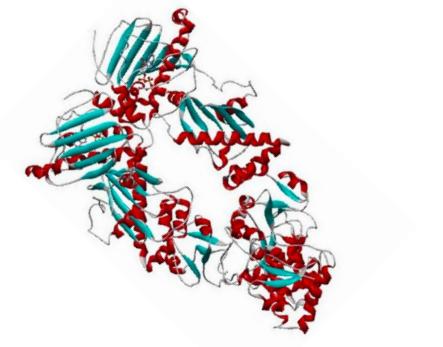
Hsp90
H. sapiens
732-854 aa

Which protein is it?
To which organism does it belong?

Does it has a determined structure?
How many? Which methods?

How many domains does it have?

In protein entry:
Family & Domains > **Pfam**



What are the main structural differences?

Sequence comparison



P07900
HtpG – *E. coli*

P0A6Z3
Hsp90 – *H. sapiens*



In Uniprot protein entry:

UniProtKB - PC
Display **BLAST**

BLAST
Filter by
**Reviewed (56)
Swiss-Prot**

Map to
UniProtKB
UniRef
UniParc

Download results
Align **Download**

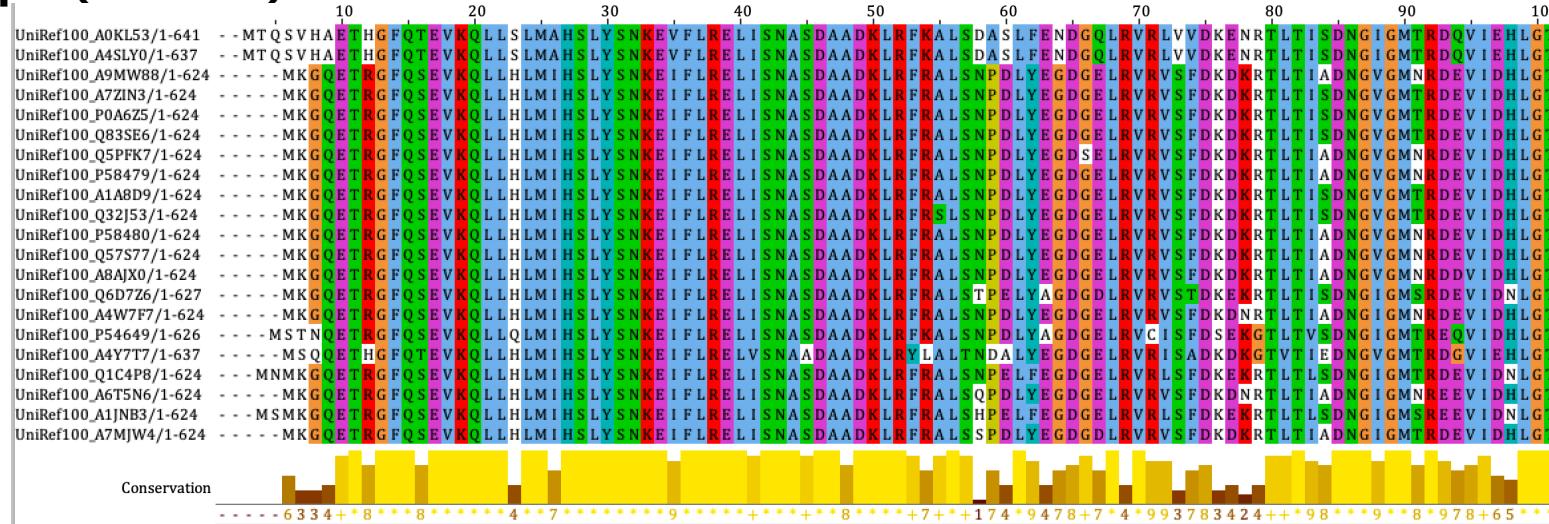
- Open Jalview
- Load/drag file with fasta sequences
- Align sequences with ClustalO

Web Service -> Alignment -> ClustalO(default)

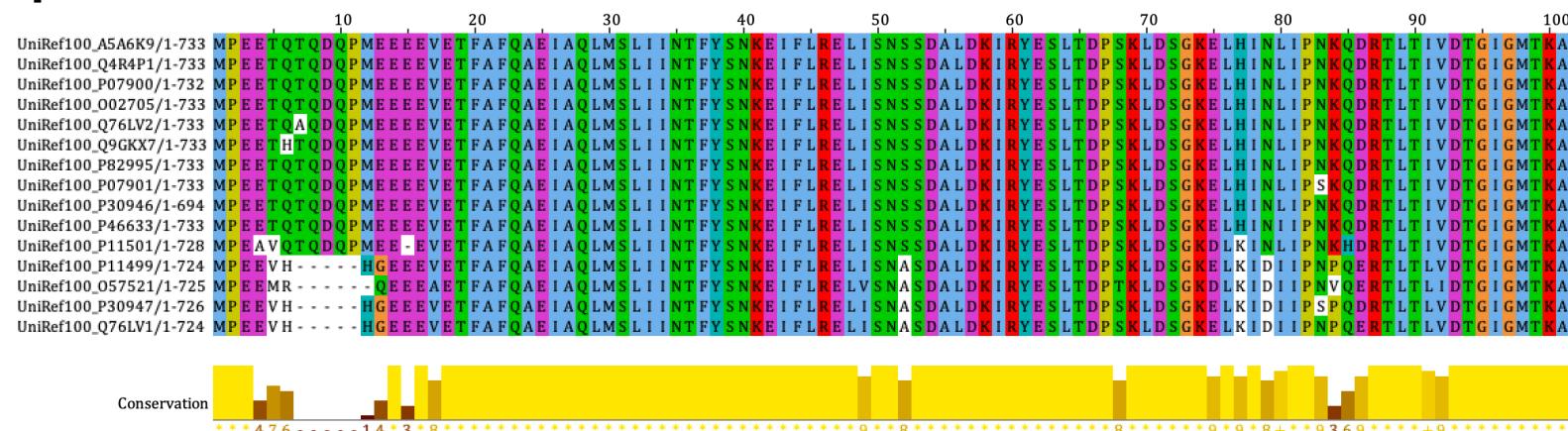


Alignment comparison

HtpG (P0A6Z3)

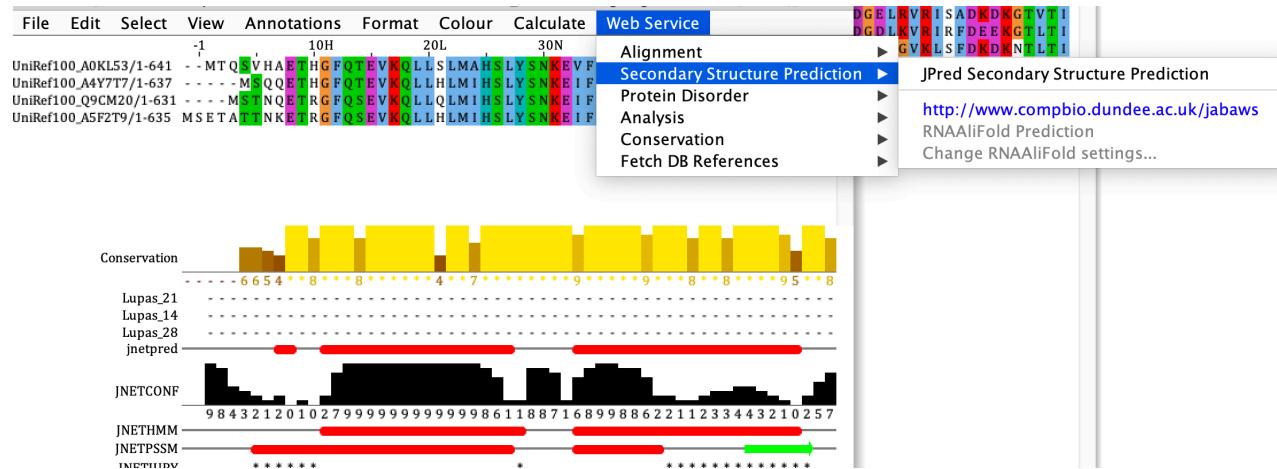


Hsp90 (P07900)

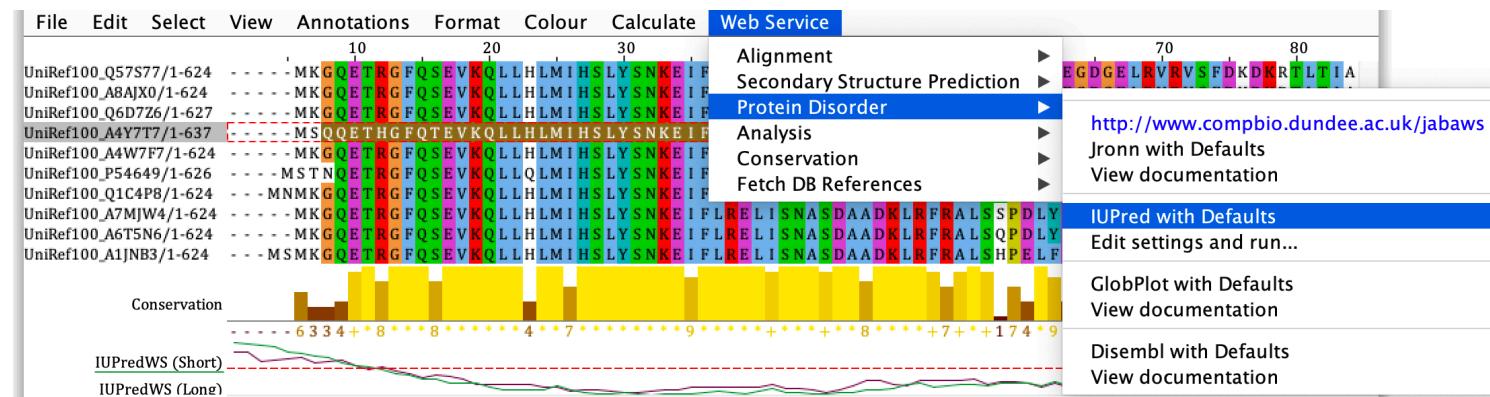


Alignment comparison

Secondary structure prediction



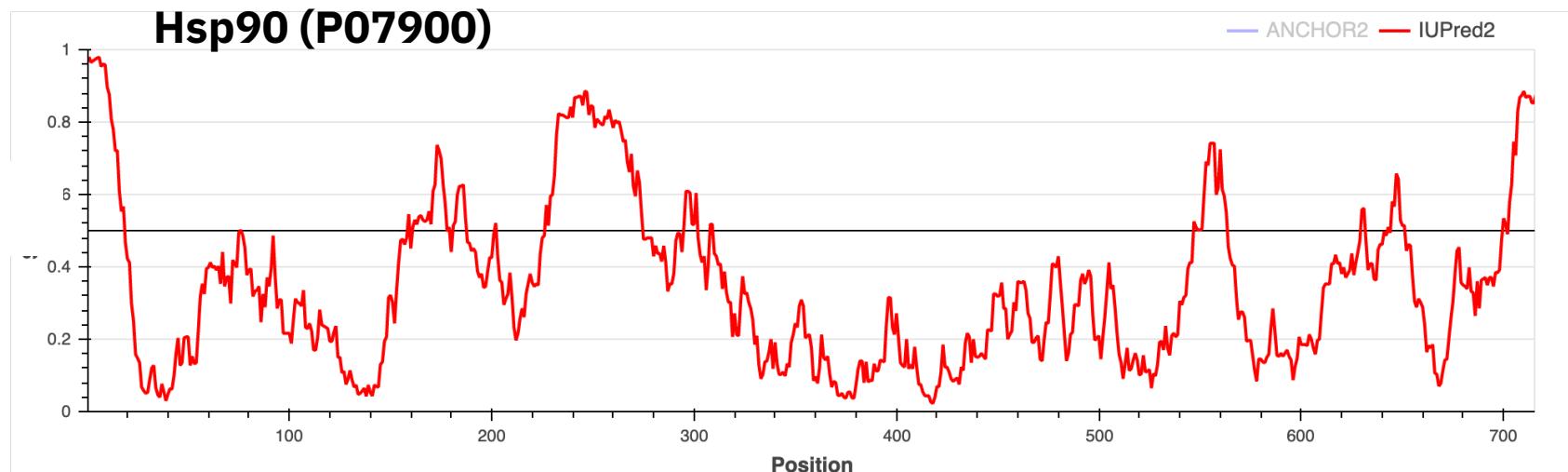
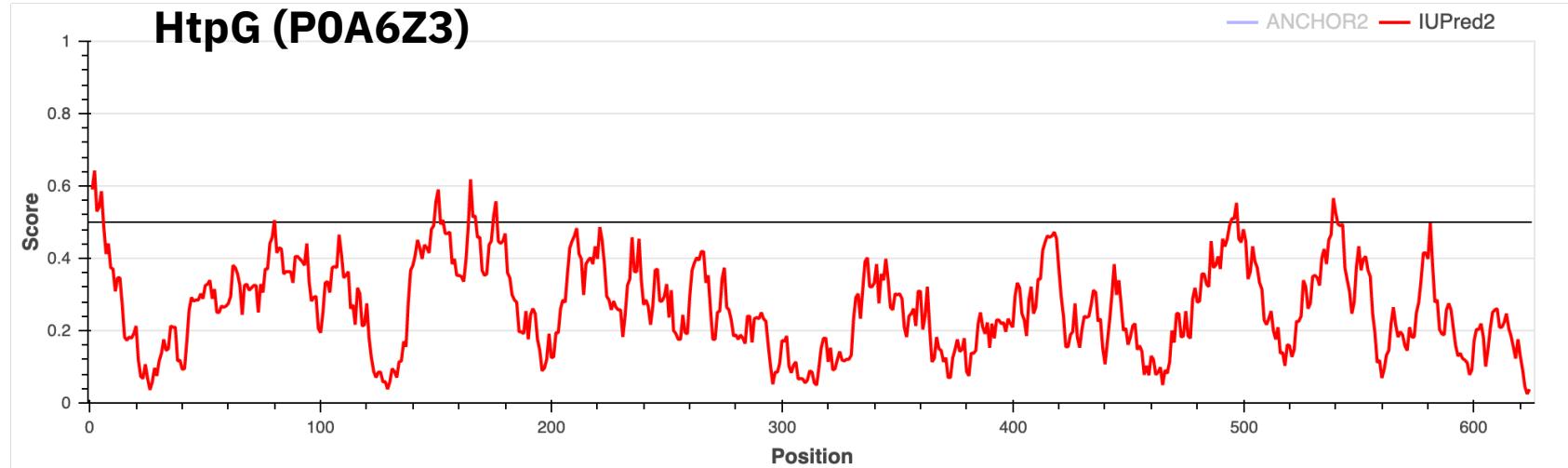
Disorder prediction



Order disorder comparison

IUPred2A

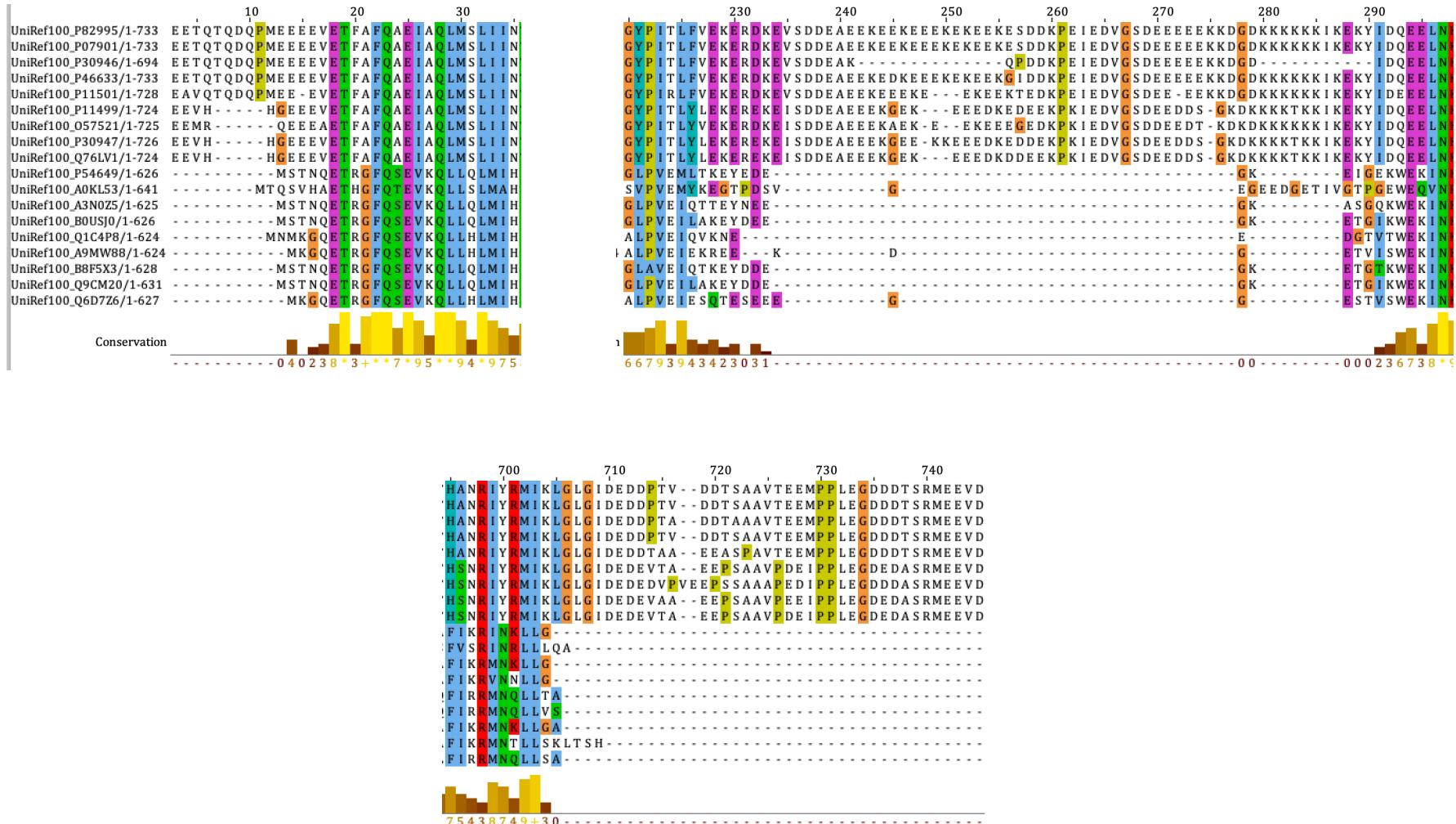
<https://iupred2a.elte.hu/>



EMBL



Alignment comparison





P07900
HtpG – *E. coli*

P0A6Z3
Hsp90 – *H. sapiens*



- Go to PDB

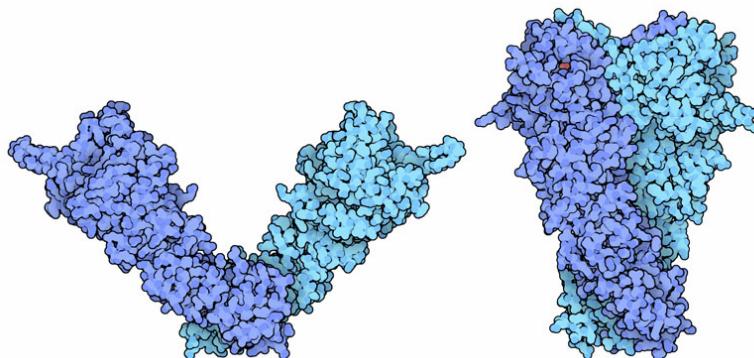


<https://www.rcsb.org/>

Look for **2IOP**

Look for **5FWK**

- How were they obtained?



Thanks

Questions?