



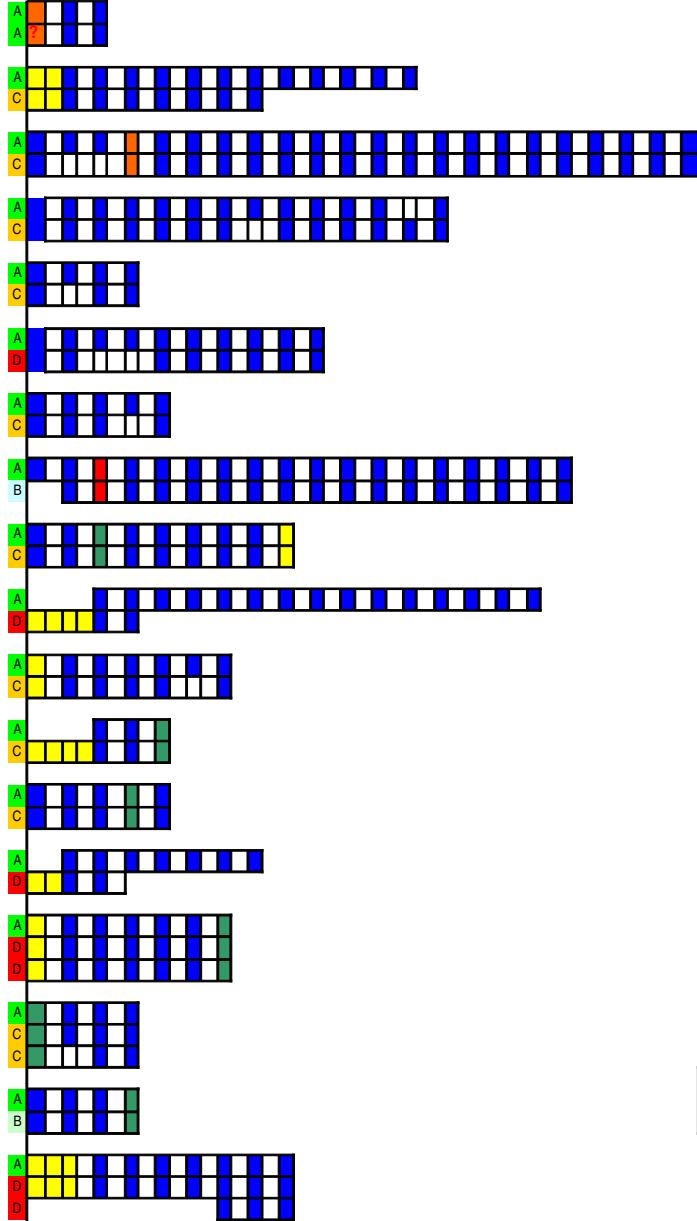
# Structure based analysis

- All proteins with a template with almost total coverage
- 20 families: 41 transcripts
- All models deposited in [www.caspur.it/PMDB](http://www.caspur.it/PMDB) (keyword: encode)



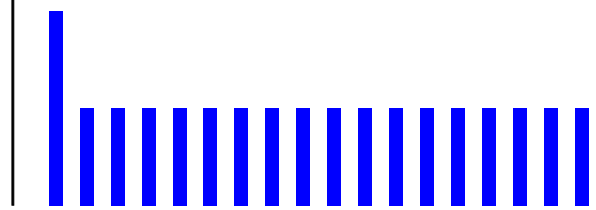
# Structure based analysis

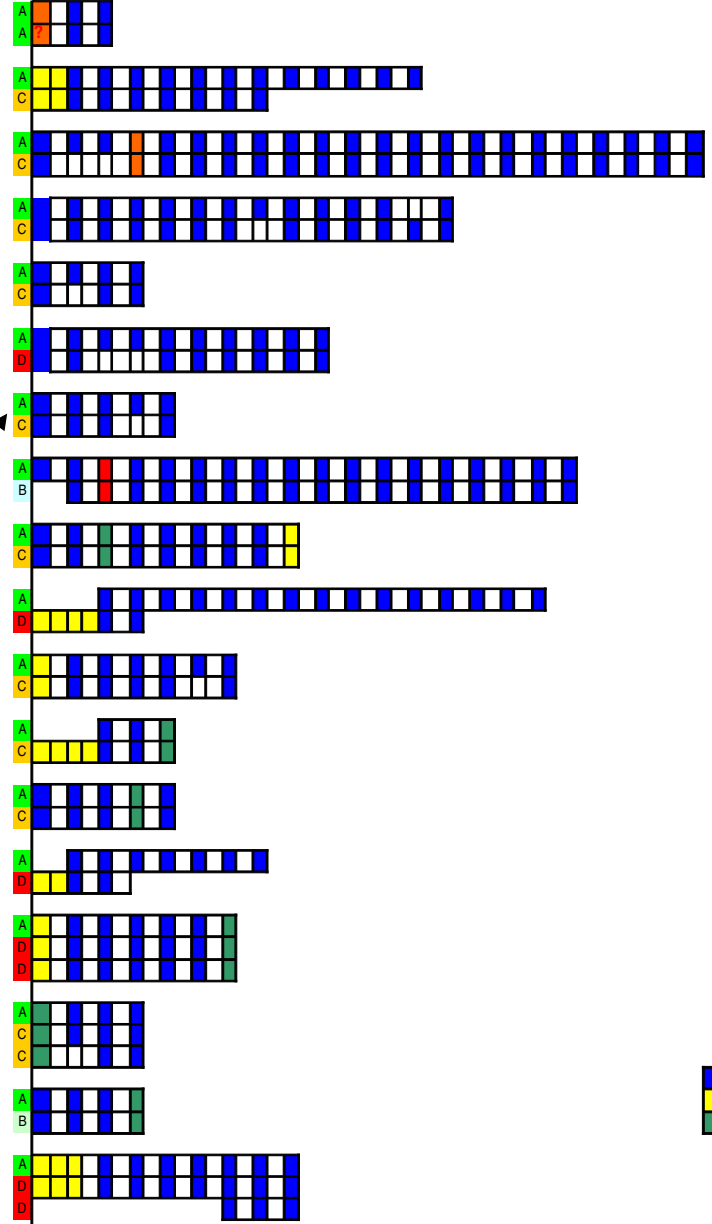
- 3 families: N-terminal alternative splicing
- 9 families: C-terminal alternative splicing
- 8 families: internal alternative splicing



exon  
untransl  
diff size

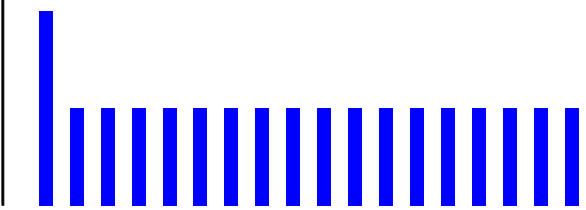
OK  
easy  
problem  
difficult





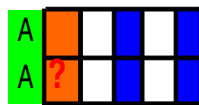
exon  
untransl  
diff size

OK  
easy  
problem  
difficult





# Example:



Locus AC006153.3: GTPase superfamily

Involved in protein synthesis, signal transduction and membrane trafficking.

Essential for growth and differentiation and usually conserved from bacteria to eukaryotes.

AC006153.3 has two variants, one of them lacking internal exons.

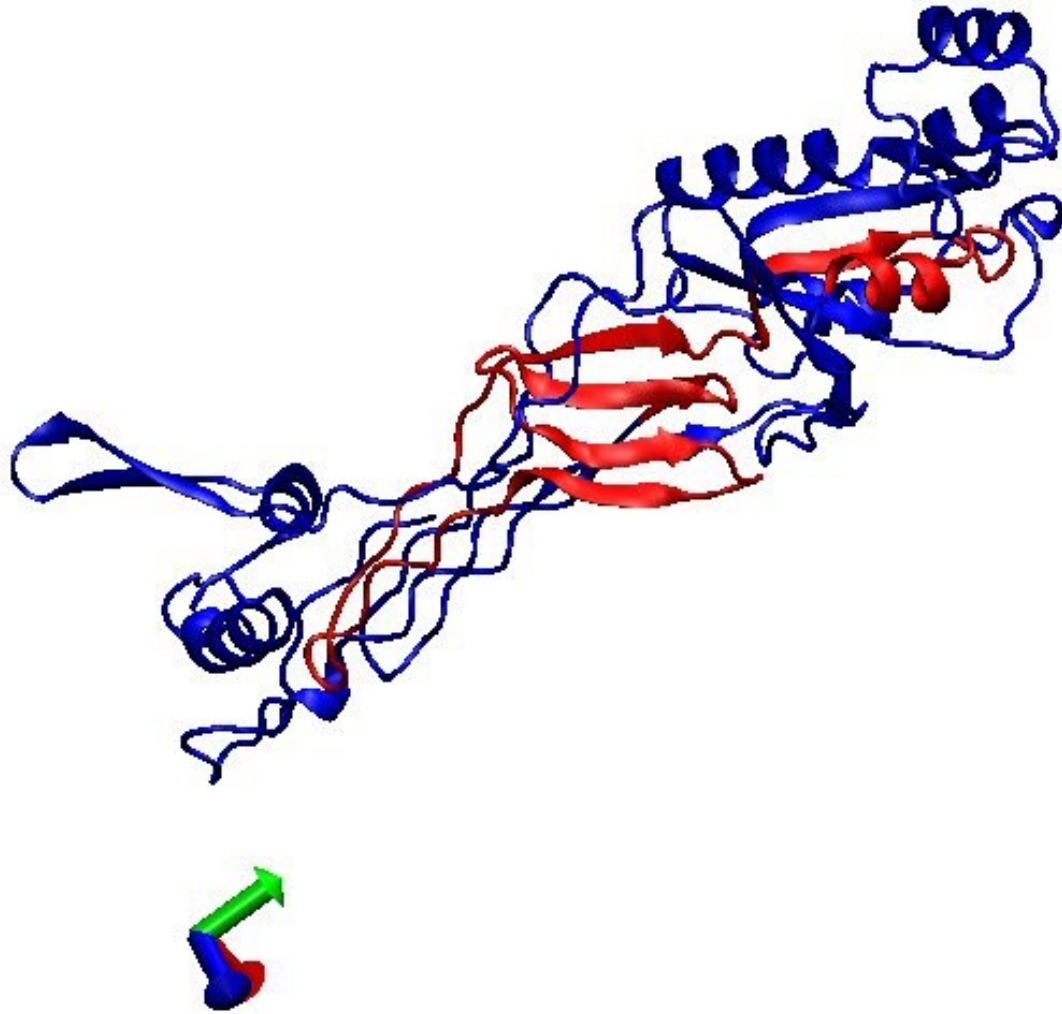


Analysis of the longest variant (001) using PFAM reveals the presence of two domains:

– The first is a GTP1-OBG fold proposed to be a GTP binding domain.

This domain is not present in the second variant (002).

The second is a six-stranded beta sheet and five alpha-helices, a GTPase domain of unknown function; member of the G-protein superfamily and present in both variants.





# ENCODE

- Alejandro Giorgetti
- Domenico Raimondo
- Massimiliano Orsini