

Testcases for 2DProt

...

2. 11. 2018

Methods

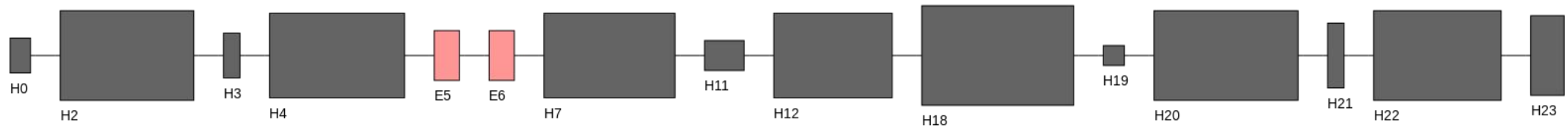
- Select 6 test-families
 - Should cover all classes: (class = first number in CATH code)
 1. Mainly α
 2. Mainly β
 3. Mixed α/β(we ignore class 4. Few SSEs)
- Select a few representants (~ 6) for each family
 - Some very similar (low RMSD, high Q-score)
 - Some very different (high RMSD, low Q-score)

$$Q = \frac{N_{\text{algn}} * N_{\text{algn}}}{(1 + (\text{RMSD}/R_0)^2) * N_{\text{res1}} * N_{\text{res2}}}$$

http://www.ebi.ac.uk/msd-srv/ssm/rl_qscore.html

GPCR 1.20.1070.10

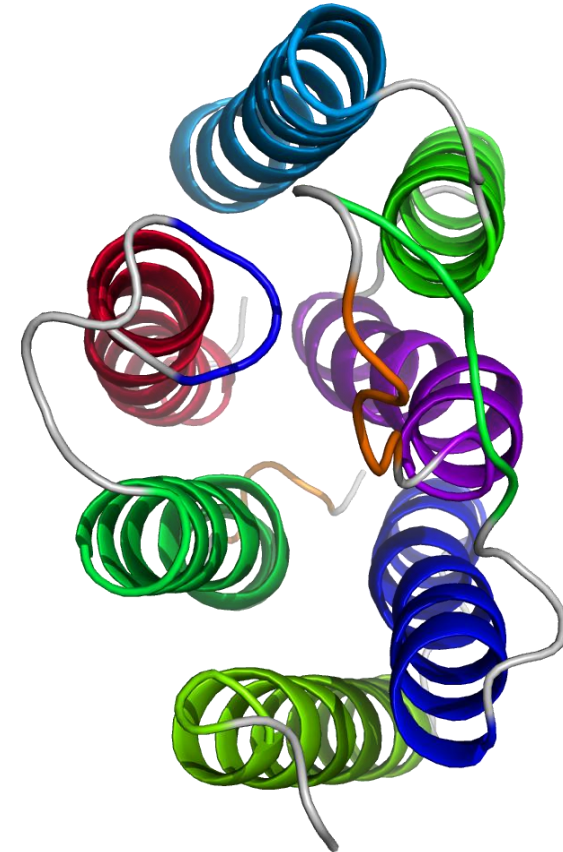
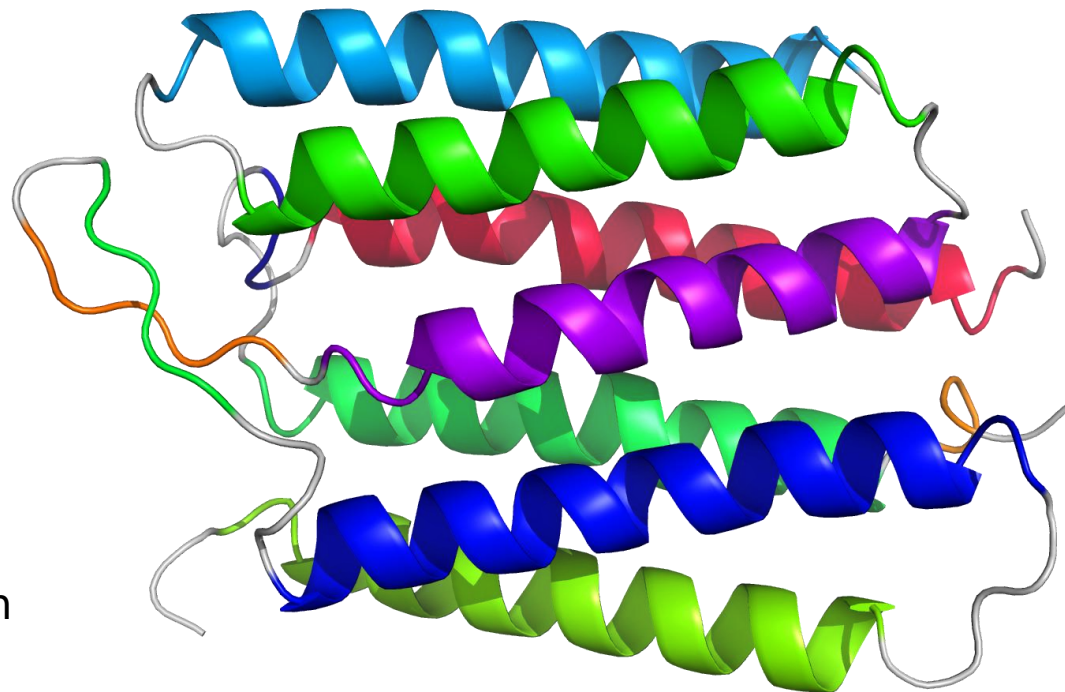
- Up-down helix bundle: 7 parallel helices



	Q-score	RMSD
5b6xA00	1.000	0.00
5b6yA00	0.999	0.09
1kg9A00	0.948	0.39
2jagA00	0.666	1.90
5g28A00	0.291	4.16
2vt4A00	0.135	5.90

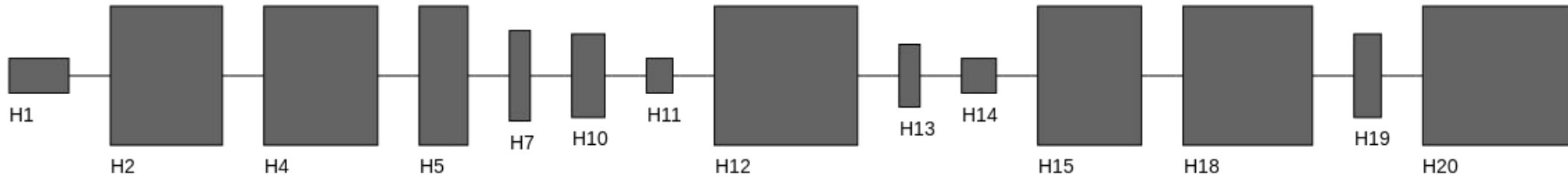
Q-score against
the first domain

RMSD against
the first domain



Globins 1.10.490.10

- Orthogonal helix bundle: non-parallel helices

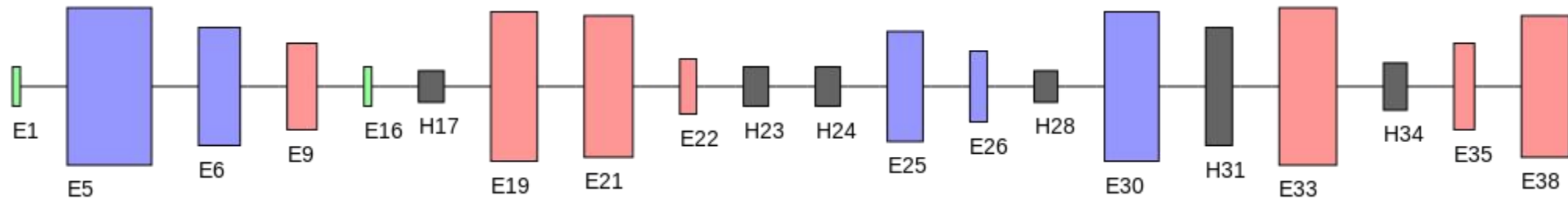


	Q-score	RMSD
111mA00	1.000	0.00
1co9A00	0.997	0.16
5jomA00	0.960	0.60
5hu6A00	0.705	1.55
2lh5A00	0.441	3.06
3zjqA00	0.232	4.08

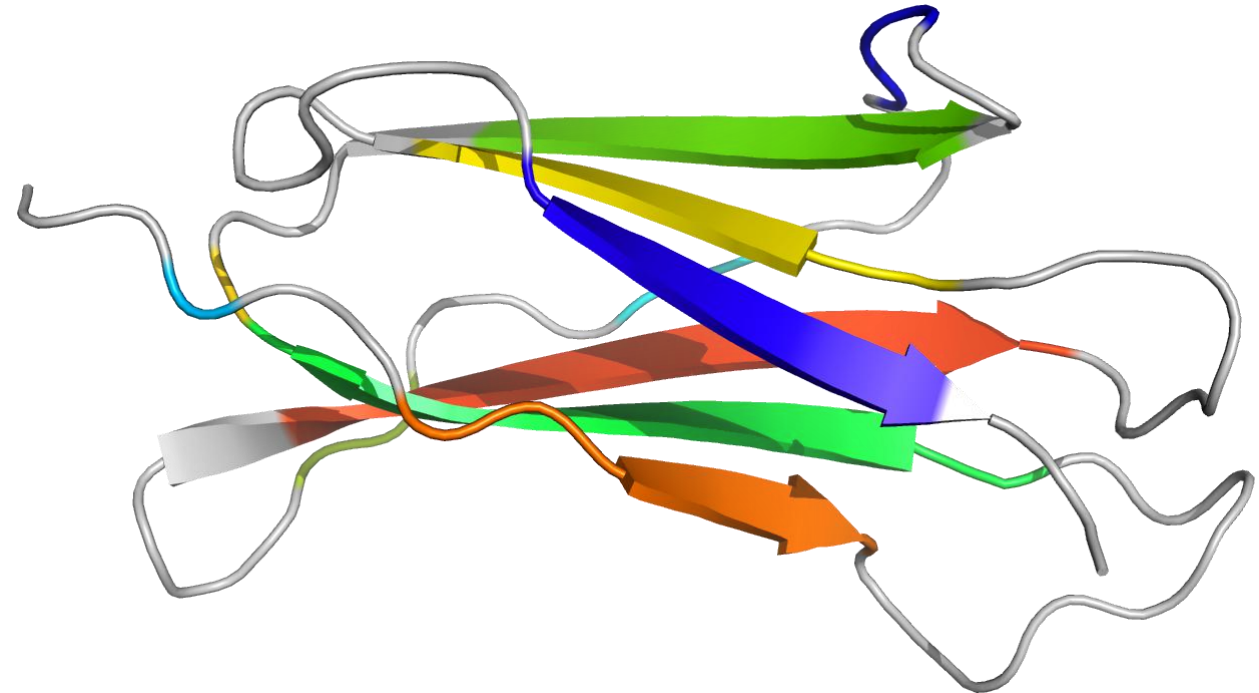


Immunoglobulins 2.60.40.10

- Sandwich = 2 sheets

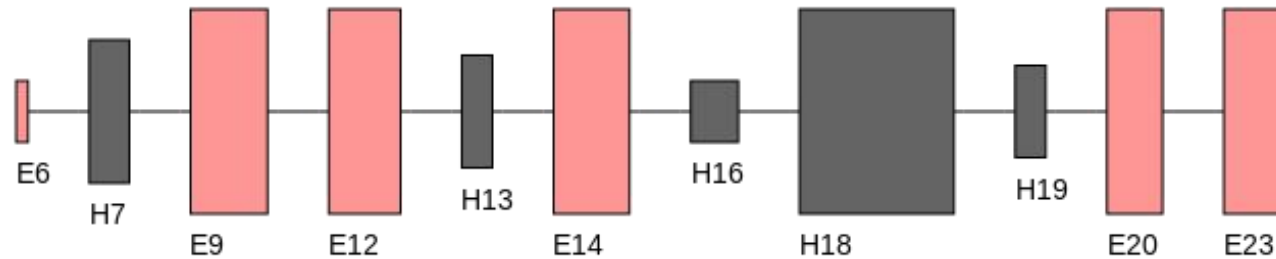


	Q-score	RMSD
1aqdA02	1.000	0.00
4p5mA02	0.978	0.45
1p7qA02	0.603	2.17
1t0pB00	0.362	3.42
4urtB01	0.166	5.35

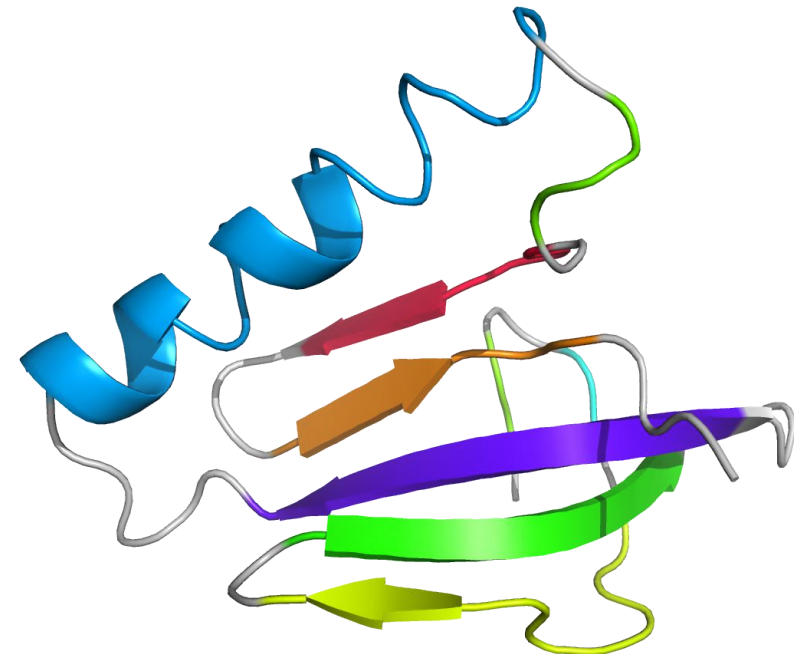


Protein of Death 3.30.200.20

- 2-layer sandwich = 1 sheet + 1 layer of helices

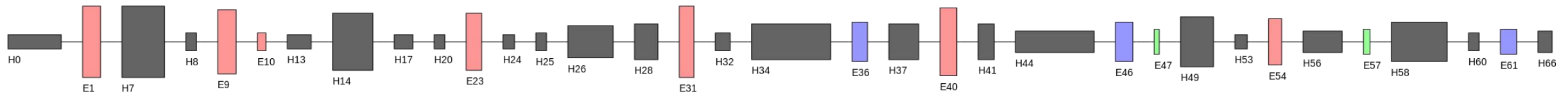


	Q-score	RMSD
1e9hA01	1.000	0.00
2wfyA01	0.903	0.98
5ap1A01	0.570	2.32
3gopA01	0.232	3.90
4fgrA01	0.131	6.11

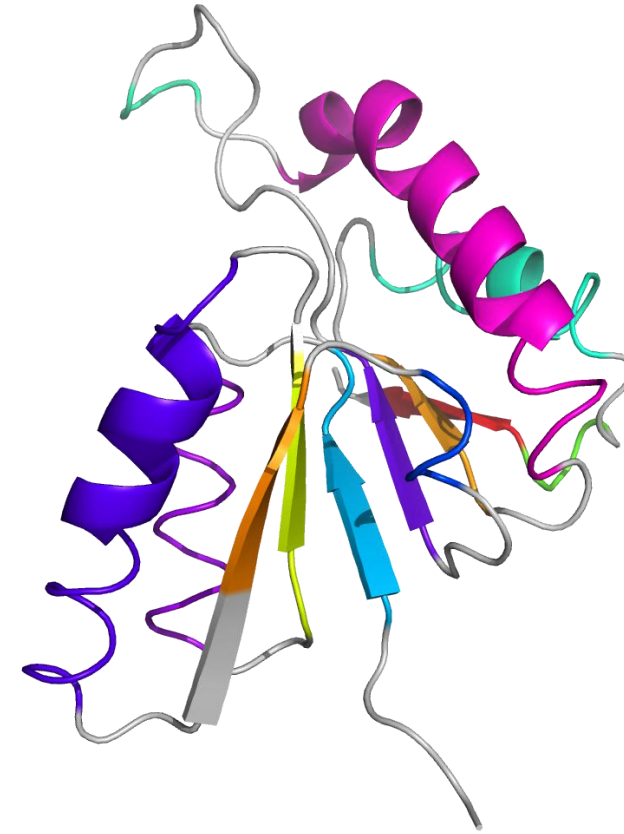


NAD(P)-binding Rossmann-like Domain 3.40.50.720

- Rossmann fold = 1 sheet + helices on both sides

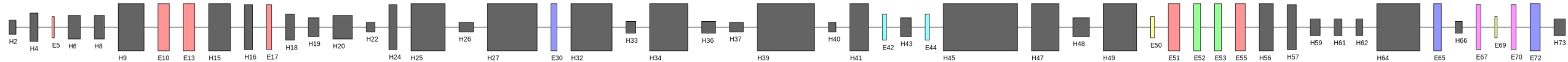


	Q-score	RMSD
1r0lA01	1.000	0.00
3iieA01	0.802	1.28
1rjwA02	0.259	3.64
3sc6A01	0.110	4.16
2bmaA02	0.088	6.04

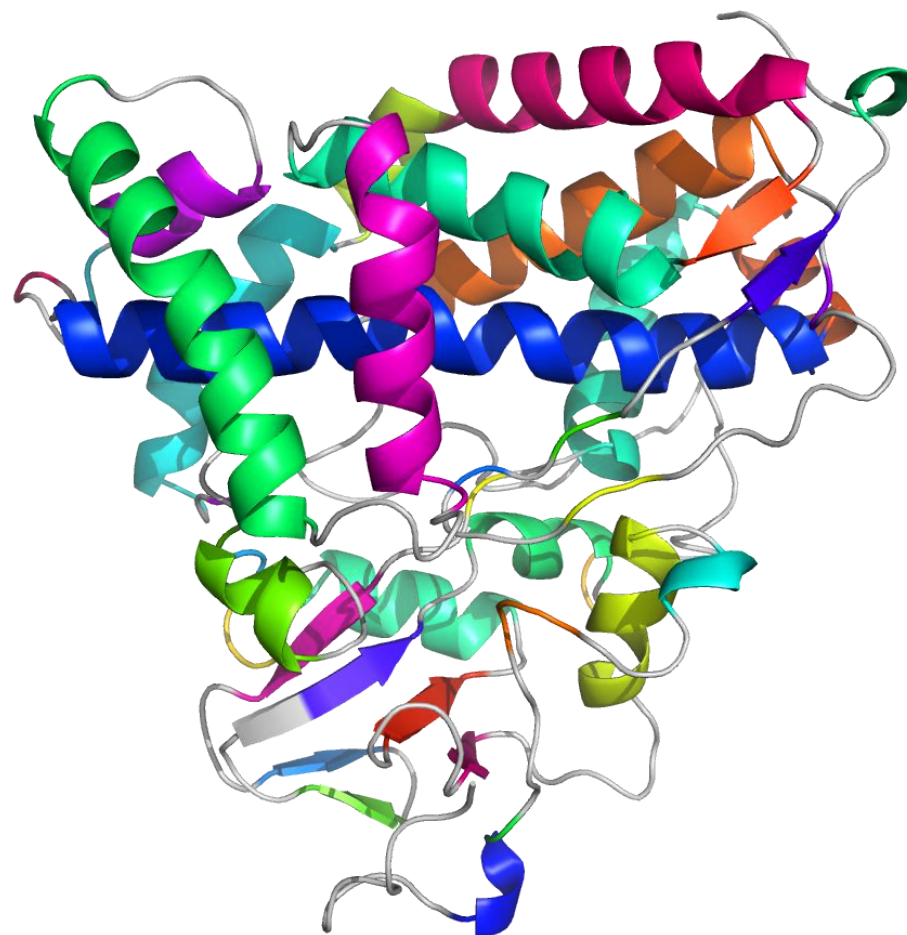


CYPs 1.10.630.10

- Big helix mess



	Q-score	RMSD
1iwkA00	1.000	0.00
4l4dA00	0.983	0.40
3p6xA00	0.720	1.80
4triA00	0.494	2.79
4uvrA00	0.255	4.04
4xryA00	0.208	4.75



Methods

Návrh metriky *Comp* (*comparativity*, nevymyslel som lepšie meno):

- $Diff3D(i,j)$ = miera odlišnosti proteínov i a j v 3D, napr.:
 - RMSD
 - Q-score
 - $\sum_{(X,Y)} [(vzdialenosť\ X\ a\ Y) + const * (1 - \cos(\text{uhol medzi } X\ a\ Y))]$
 - kde (X,Y) sú všetky spoločné SSEs (X je v i , Y je v j)
- $Diff2D(i,j)$ = miera odlišnosti proteínov i a j v 2D, napr.:
 - $\sum_{(X,Y)} [(vzdialenosť\ X\ a\ Y) + const * (1 - \cos(\text{uhol medzi } X\ a\ Y))]$
- $Comp(A) = corr(Diff3D(i,j), Diff2D(i,j))$
 - kde (i,j) sú všetky kombinácie proteínov z rodiny A
- $Comp \uparrow$ podobné vyzerajú podobne, rôzne vyzerajú rôzne, sme happy :)
- $Comp \downarrow$ niečo je špatne :(