Methylenetetrahydrofolate reductase isoform 1 [*Homo sapiens*] Protein Structure

Introduction to Bioinformatics

Group 2: Jane Ryu, Joe Sparks, Julia Garcia, and Katharine Jenike

Milestone D

[7] Compare the predicted structure of your protein to that of a known structure.

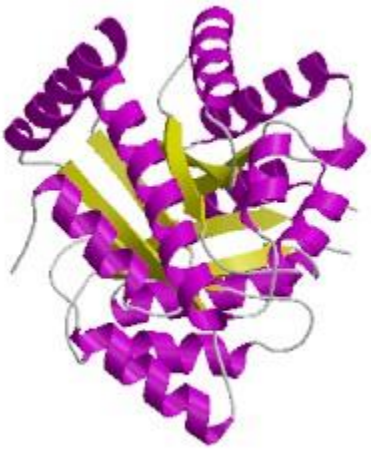
> Novel Protein

VRAGADLCITDVFYDTNAYAKFIKECREAGIARTFPIVPGILPIHSFKSFEGIVDHLGINVP
ASIREAIEPIKEDDAAMQEYGISLAESMCLELLNSGLAQGMYFYTFNLEYSVRHLLEERL
KVTPKSQLPWRPSANPKRIEEDVRPIFWANRPKSYLIRTESWNEFPSGRWGSAVESASFS
ELKDSTLFARETFFERDDIKKKAWGEAPQTREEVFEVFAGFVEGRVQFLPWCEESLHLE
TSVIRDKLVQV

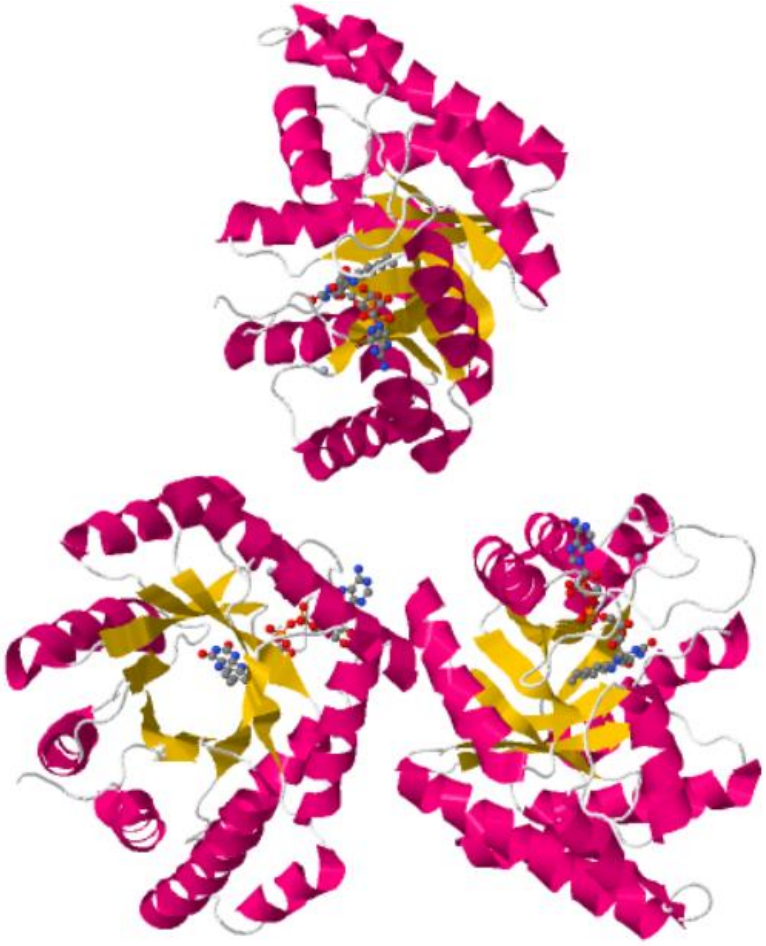
Submitting our novel protein sequence to PHD, we found that the secondary structure was 42.46% alpha helices (13 of them), 41.67% random coil, and 15.87% extended strand. TMHMM and HMMTOP found zero transmembrane domains in the protein. Using PROSITE, we found zero motifs for our novel protein when excluding motifs with a high occurrence rate. Unselecting this option showed several commonly occurring motifs, 4 N-myristoylation sites, 5 protein kinase C phosphorylation sites and 7 casein kinase II phosphorylation sites.

InterPro characterized the protein as a member of the methylenetetrahydrofolate family, as expected. Pfam and the CATH database found the MTHFR functional domain between amino acids 1-121. CATH also found several other possible domains, including a galactose-binding domain, a molybdenum cofactor biosynthesis domain, and a NAPD-dependent oxidoreductase domain.

Below is a photo from PDB of the domain that characterizes it's methylenetetrahydrofolate functionality.



In comparison to our novel protein, the following image is the known 3d structure of *E. coli* methylenetetrahydrofolate reductase[10] from protein databank ID 1B5T. According to pfam, this known structure contains the characterizable MTHFR domain at amino acids 22-291[11]. The structure is a homo-3-mer with 3 identical subunits. On visual observation of the models, the 3-dimensional structure of one subunit looks very similar to our predicted novel protein's structure. The TIM barrel with alternating alpha helices and beta strands is notable. The *E. coli* protein is 47% helical and 14% beta sheet[10]. This is also in line with our novel protein's predicted secondary structure.



[8] Show whether this gene is under positive or negative evolutionary selection.

Using SNAP2 from the Rost lab we examined the functional effect of mutations at each amino acid of the novel protein. SNAP2 showed that the amino acid at position 252 (typically a valine) is least likely to be effected by a change to another amino acid, with a SNAP score of -99 and 97% expected accuracy. Of the 5040 possible mutations, 1826 of them are expected to have a neutral effect and the remaining possible mutations are expected to have an effect on the protein structure. This tells us that non-synonymous substitutions outnumber synonymous substitutions. Since d_N is greater than d_s, the d_N/d_s ratio would be greater than 1 and the gene is likely under positive evolutionary selection.

The amino acids with a the highest SNAP score are likely under evolutionary pressure. The 104th amino acid, which is typically a Tyrosine, is likely under evolutionary pressure according to SNAP2, since the SNAP score is 98 with 95% expected accuracy. Additionally, mutations at positions 40, 62, and 106 are expected to have a functional effect and all have the highest SNAP scores of all amino acids in the novel protein.

Conversely, the 246th, 231st, 214th, and 201st amino acids all have the lowest expected accuracy at 53%. Changes to these amino acids are expected to be neutral.

[9] Discuss the significance of your novel gene. What have you learned about this gene/protein family?

MTHFR (methylenetetrahydrofolate reductase isoform 1) from Homo sapiens is known to be linked to a variety of diseases including vascular, neural tube defects, colon cancer, MTHFR deficiency, and acute leukemia. The 3677T polymorphism, in which up to 25% of the population is homozygous, is linked to increased homocysteine in the blood which is an independent risk factor for cardiovascular disease (Konrad et al., 2004). The prevalence of the polymorphism and clinical signficance has made the MTHFR protein a subject of interest. By studying other possible organism sequences that contain MTHFR, or homologs, we can shed light on the importance of MTHFR and its sequence. We chose Schizochytrium MTHFR from the organism Schizochytrium aggregatum to learn more about the MTHFR protein family.

Using multiple sequence alignment (MSA), the MTHFR protein sequence was highly conserved across species, this was shown through comparison of sequences with human, western lowland gorilla, pygmy chimpanzee, green monkey sequences. Schizochytrium MTHFR sequence showed ~50% similarity with the sequences, but high similarity in the mid region of the sequence. Comparing the predicted protein structure of Schizochytrium MTHFR to *E. coli* MTHFR as shown above, there was high similarity in structure. Both structures shared the Tim barrel and the MTHFR domain.

Like its homologs, our novel protein, Schizochytrium Methylenetetrahydrofolate reductase isoform 1 (MTHFR) most likely has methylenetetrahydrofolate reductase (NAD(P)H) activity, which catalyzes the reaction: 5-methyltetrahydrofolate + NAD(P)+ = 5,10-methylenetetrahydrofolate + NAD(P)H + H+ [8]. It may also be a part of the biological processes of methionine metabolism and the oxidation-reduction process [7].

References:

1. https://npsa-prabi.ibcp.fr/cgi-bin/secpred_phd.pl
2. http://www.cbs.dtu.dk/cgi-bin/webface2.fcgi?jobid=58F283FA00006B1BDB246360&wait=20
3. http://www.enzim.hu/hmmtop/server/hmmtop.cgi
4. http://cctop.enzim.ttk.mta.hu/?_=/jobs/submit

5. http://prosite.expasy.org/
6. http://pfam.xfam.org/search/sequence
7. https://www.ebi.ac.uk/interpro/sequencesearch/iprscan5-S20170415-215405-0774-52603372-es
8. http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0004489
9. http://www.cathdb.info/search/by_sequence
10. http://www.rcsb.org/pdb/explore/jmol.do?structureId=1B5T&bionumber=1
11. http://pfam.xfam.org/protein/METF_ECOLI#tabview=tab2
12. https://blast.ncbi.nlm.nih.gov/Blast.cgi
13. Konrad C., Muller G.A., Langer C., Kuhlenbaumer G., Berger K., Nabavi D.G., Dziewas R., Stoqbauer F., Ringelstein E.B., and Junker, R. (2004), Plasma homocysteine, MTHRF C677T, CBS 844ins68bp, and MTHFD1 G1958A polymorphisms in spontaneous cervical artery dissections, *J Neurol*, 251(10): 1242-8.
14. Bromberg, Y., & Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Research*, *35*(11), 3823–3835. http://doi.org/10.1093/nar/gkm238