Molecular modeling and identification of transmembrane protein (MLH1) related to colon cancer using *Insilico* techniques

Hepzibah. W¹, John William S¹ and Balaji.M²

¹P.G & Research Dept. of Advanced Zoology and Biotechnology, (SECNARM), Loyola College, Chennai-600034.  
²Managing Director, Akshaya Computational Medicine Pvt, Ltd., Chennai-600005.

**Abstract**

Worldwide more than 1 million people get colorectal cancer every year resulting in about 715,000 deaths as of 2010 up from 490,000 in 1990. In this research investigation, we use *insilico* tools for protein modeling and analysis. The gene MLH1 protein sequence was modeled and validated using automated homology modeling servers. Finally, we find out the transmembrane regions present in the MLH1 protein. These transmembrane regions play a major role in structure-based drug designing process.

**Keywords:** Colorectal Cancer, MLH1 protein and Peptide modeling

**INTRODUCTION**

Globally more than 1 million people get colorectal cancer every year [1] resulting in about 715,000 deaths as of 2010 up from 490,000 in 1990 [2]. As of 2012, it is the second most common cause of cancer in women (9.2% of diagnoses) and the third most common in men (10.0%) with it being the fourth most common cause of cancer death after lung, stomach, and liver cancer. It is more common in developed than developing countries [3]. Globally incidences vary 10-fold with highest rates in Australia, New Zealand, Europe and the US and lowest rates in Africa and South-Central Asia.

**METHODOLOGY**

**Literature studies:** In these studies we used online literature database like PMC and OMIM to identify the molecular profiles of the colon cancer.

**Molecular modeling:** Based on molecular modeling studies we use CPH 3.0 model server [4] in order to perform 3D structure validation using Rampage tool (Rapper server). The modeled protein structure was viewed using molecular visualization tools like Discovery studio software. Finally, we identify the transmembrane surface of the (l) using SPLIT sever.

**RESULT AND DISCUSSION**

The length of the selected protein target is 756 aa which is retrieved from uniport database and which is located at 3rd chromosome [5]. Maliaka et al., [6] identified 6 different novel mutations in the MLH1 and MSH2 genes in Russian and Moldavian HNPCC families. Three of these mutations occurred in CpG dinucleotides and led to a premature stop codon, splicing defect, or an amino acid substitution in evolutionarily conserved residues. Analysis of a compilation of published mutations including the new data suggested to the authors that CpG dinucleotides within the coding regions of the MSH2 and MLH1 genes are hotspots for single base pair substitutions. In Fig (2 and 3) clearly represents that modeled protein target shows molecular structural regions. Fig (4) explains that the target protein structure MLH1 shows transmembrane regions.

![Fig 1. FASTA sequence of colon cancer protein - MLH1](image)

>sp|P40692-2|MLH1_HUMAN Isoform 2 of DNA mismatch repair protein Mlh1 OS=Homo sapiens GN=MLH1 MNGYISNANYSVKKCIFLLFINHRLVESTSLRKAIETVYAAYLPKNTHPFLYLSLEISPQ NVDVNVHPTKHEVHLHEESILERVQQHIESKLLGSNSSRMYFTQTLPLLGLAGPSGMVK

To whom correspondence should be addressed: 
Hepzibah. W
Email: hepsijoe123@gmail.com
Hepzibah. W et al., Molecular modeling and identification of transmembrane protein (MLH1) related to colon cancer using Insilico techniques

STTSLTSSTSGSSKVVYAMVRDSEQKLDAFLQPLSSQPAIVTEDKTDISS
GRARQDEEMLELPAVEAKNQSLGDTTGTSEMSEKRGPTSSNPRKRRHRSDDVEM
VEDDSRKMIAACTPARRRINLTSLQEEINEQGHEVRLREMHLHNSYICVNPQWALA
QHQTLYNLQNTTMLFQILYDFANFGLSPEPLFLDAMLALDSPEGWTEED
GPKGLAELYVEFLKKAEMADYFSEIDEGNKLGLDLIDNYVPPEGLPLIFILRLA
TEVNWDEEKFESLSKECAMFSIRKQYISESTLSGQSEVPNSWKTVHEIVY
KARSHILPPKHFDGNIQLANLPDLYKFerc

Fig 2. Protein modeling – MLH1 protein

Secondary structure view – red: helix, blue: sheets
green – turns and white: Coils

Fig 3. Protein modeling – MLH1 protein

CPK model view- Red – oxygen, Blue–nitrogen, Yellow–sulphur and white–hydrogen

Fig 4. SPLIT server – Transmembrane region prediction – MLH1

SPLIT SERVER THE PREDICTION RESULTS

Red line: Transmembrane helix preference. (THM index)
Blue line: Beta preference. (BET index)
Gray line: Modified hydrophobic moment index. (INDA index)
Violet boxes (below abscissa): Predicted transmembrane helix position. (DIG index)

Fig 5. Rapper server : Structure validation

Number of residues in favoured region (~98.0% expected): 250 (95.1%)
Number of residues in allowed region (~2.0% expected): 11 (4.2%)
Number of residues in outlier region: 2 (0.8%)
CONCLUSION
In this research investigation, we focus on two major areas. First, we focus on the molecular modeling of the target protein. Next, we identify the transmembrane surface of the modeled protein structure. The identified transmembrane protein play a major role in future structure based drug docking studies.

REFERENCES