A Structured-based Model for the Decreased Activity of Ala222Val and Glu429Ala Methylenetetrahydrofolate Reductase (MTHFR) Mutants

Khuram Shahzad[1,2], Abdul Hai[3], Asifa Ahmed[2], Nadeem Kizilbash[3]* & Jamal Alruwaili[3]

[1]Illinois Informatics Institute, University of Illinois, Urbana-Champaign, Illinois, U.S.A; [2]Department of Biosciences, COMSATS Institute of Information Technology, Islamabad, Pakistan; [3]Department of Biochemistry, Faculty of Medicine & Applied Medical Sciences, Northern Border University, Arar-91431, Saudi Arabia; Nadeem Kizilbash – Email: fsd707@gmail.com; *Corresponding author

Abstract:
The structure of human Methylenetetrahydrofolate Reductase (MTHFR) is not known either by NMR or by X-ray methods. Phosphorylation seems to play an important role in the functioning of this flavoprotein. MTHFR catalyzes an irreversible reaction in homocysteine metabolism. Phosphorylation decreases the activity of MTHFR by enhancing the sensitivity of the enzyme to S-Adenosylmethione. Two common polymorphisms in MTHFR, Ala222Val and Glu429Ala, can result in a number of vascular diseases. Effects of the Glu429Ala polymorphism on the structure of human MTHFR remain undetermined due to limited structural information. Hence, structural models of the MTHFR mutants were constructed using I-TASSER and assessed by PROCHECK, DFIRE and Verify3D tools. A mechanism is further suggested for the decreased activity of the Ala222Val and Glu429Ala mutants due to a decrease in number of serine phosphorylation sites using information gleaned from the molecular models. This provides insights for the understanding of structure-function relationship for MTHFR.

Key Words: Methylenetetrahydrofolate Reductase, *in silico* modeling, hyperhomocysteinemia, 5,10-Methylenetetrahydrofolate, Methylenetetrahydrofolate Reductase phosphorylation.

Background:
Methylenetetrahydrofolate reductase (MTHFR) plays an important role in processing of amino acids. It converts 5, 10 Methylenetetrahydrofolate to 5-Methyltetrahydrofolate which provides the methyl group for methylation of homocysteine to methionine by the enzyme Methionine Synthase **[1]**. High plasma homocysteine levels (hyperhomocysteinemia) are associated with increased risk for cardiovascular disease and neural tube defects in humans. Treatment with Folic Acid decreases homocysteine levels and dramatically reduces the incidence of neural tube defects. The flavoprotein, MTHFR is a likely target for the action of Folic Acid. MTHFR is a

homodimer. Each domain can be separated by partial digestion of the native enzyme with trypsin. It contains a bound FAD prosthetic group on each domain and uses NADPH as reducing agent. An S-adenosylmethione-binding site is present in the regulatory domain. The binding of S-adenosylmethione allosterically inhibits the activity of the enzyme.

MTHFR gene codes for a protein of 656 amino acids with a predicted molecular mass of 74.5 kDa **[2]**. MTHFR uses NADPH as a reducing agent and FAD as a cofactor in mammalian cells **[3]**. Eukaryotic MTHFR comprises a catalytic domain of about 40 kDa at the N-terminus and a C-terminus regulatory domain

of about 37 kDa. The C-terminus of human MTHFR contains the S-Adenosylmethionine-binding site which regulates the Methionine level in cells [4]. This C-terminus domain is not present in MTHFR from E. coli. Hence, NADH-dependent MTHFR from *E. coli* is insensitive to S-Adenosylmethionine [5].

Both human and other eukaryotic MTHFR exist as homodimers, whereas MTHFR of *E. coli* is a homotetramer [6-7]. The homodimers can be separated by partial digestion of the native enzyme with trypsin [6]. Each homodimer contains two non-covalently bound FAD molecules per dimer. S-Adenosylmethionine plays an important role in keeping the enzyme in dimeric form [8]. MTHFR can also be phosphorylated which decreases its activity by ~20% and allows it to be more easily inhibited by the S-Adenosylmethionine [9]. As mentioned before, MTHFR contains a bound flavin cofactor and uses NADPH. A→V mutation in human (A222V) and in *E. coli* (A177V) increases the rate of dissociation of FAD leading to the structural changes in the enzyme [10]. Structural changes in the MTHFR enzyme lead to many diseases such as Homocystinuria [11], Coronary artery disease [12], Neural tube defects [13], thyroid dysfunction [14] and colorectal cancer [15], etc. Protein structure prediction is one of the most important areas in the fields of structural bioinformatics, computational biology and drug design. The current study was carried out to predict the structure of human MTHFR for a better understanding of the mechanism of the enzyme. The structure of human MTHFR was predicted using different online servers and then evaluated for further analysis.
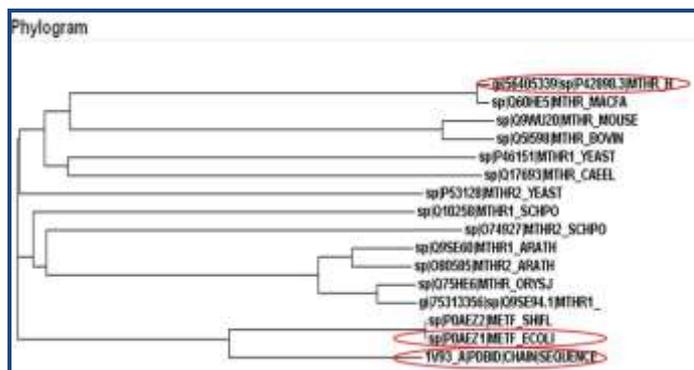
Figure 1: Phylogenetic relationship of Human MTHFR protein sequence (P42898) with MTHFR protein sequence of 15 selected species. The phylogram shows that Human MTHFR posseses greater similarity with *Macaca fascicularis* (SwissProt ID: Q60HE5). Only two structures of MTHFR from the selected species are experimentally determined having their structures available in the Protein Data Bank. These two species are *E. coli* (PDB code: 1b5t) and *Thermus thermophilus* (PDB code: 1v93).

Methodology:
Sequence of MTHFR and Template Selection
The sequence of MTHFR (P42898) was retrieved from Expasy server [16] in FASTA format. PSI BLAST [17] was used to retrieve homologous sequences of existing protein structures in the Protein Data Bank. Fifteen sequences of different species showing some homology were retrieved using PSI-BLAST. Phylogenetic analysis was carried out to obtain information about evolutionary relationship between the different species and as well as to find out if structural homologs were available

in PDB for template-based modeling using the ClustalW server [18].

Homology modeling
For tertiary structure prediction of human MTHFR, different methodologies were adopted. Homology modeling was performed using different online servers for 3D structure prediction such as CPH Models [19], EsyPred3D [20], HMMstr/Rosetta [21-22], Swiss Model [23] and Protein Structure Prediction Server (PS) [24]. Then I-TASSER [25-27] server was used for threading purposes. In this method the target sequences are first threaded using a representative PDB structure library to search for the possible folds by Profile-Profile Alignment (PPA), Hidden Markov Model, PSI-BLAST profiles, Needleman-Wunch and Smith-Waterman alignment algorithms. I-TASSER server predicted the human MTHFR protein models from which five best models were selected after refinement.

Model Validation
For evaluation purposes, we used DFIRE [28], Verify3D [29], and Procheck [30-31] programs. DFIRE is a statistical, potential-based program that uses a distance-scaled finite ideal-gas reference state. DFIRE is used to assess non-bonding interactions in the protein model. A lower energy indicates that a model is closer to the native conformation. The Verify3D method analyzes the compatibility of an atomic model of the protein with its amino acid sequence. The scores range from -1 to +1 (-1 being unacceptable and +1 being an acceptable score). The PROCHECK suite of program was used to assess the stereochemical quality of a given protein structure. The results were also confirmed using the link: http://nihserver.mbi.ucla.edu /SAVES/.

Model Analysis
For the predicting the number of domains, DOMAC server [32-35] was used. The human MTHFR was predicted to have two domains. To predict phosphorylation sites, Netphos 2.0 [36] sever was used. Also identified in the structure are Tryptic cleavage site, FAD binding site at the N-terminus, NADPH and S-Adenosylmethionine-binding sites. Amino acid sequences of mutant proteins were obtained by introducing mutations in the sequence of the wild type manually. The three mutant sequences were constructed: (a) by replacing Alanine with Valine at position 222 in the protein sequence; (b) by replacing Glutamate with Alanine at the position 429 in the protein sequence; and (c) by replacing both Alanine at position 222 and Glutamate at position 429 with Valine and Alanine, respectively. After introducing the mutations in the sequences, the tertiary structures were predicted by I-TASSER. The protein models were visualized by the use of VMD software [37].

Result:
MTHFR protein sequence was used for multiple sequence alignment and retrieval of homologous sequences of experimentally determined 3-D structures present in the Protein Data Bank. **Table 1 (see supplementary material)** shows the percentage identity of the similar sequences and their PDB (Protein Data Bank) ID codes of known structures. It was noticed that experimentally determined structures were available for proteins sequences from *E. coli* and *Thermus thermophilus* only in the Protein Data Bank. **Figure 1** shows the

Phylogenetic relationship between the selected species. It was found out that the acid acid sequences of both *E. coli* and *Thermus thermophilus* MTHFR are very different from the human MTHFR sequence. The complete models were not predicted by homology modeling-based protein structure prediction servers because of the unavailability of the required templates in the Protein Data Bank (PDB). These servers were able to predict the structure of MTHFR from residue 1 to 340, but the rest of the structure was not predicted. Five models of the protein structure were generated by the I-TASSER server after refinement using the threading technique and were visualized by VMD software **(Figure 2).**
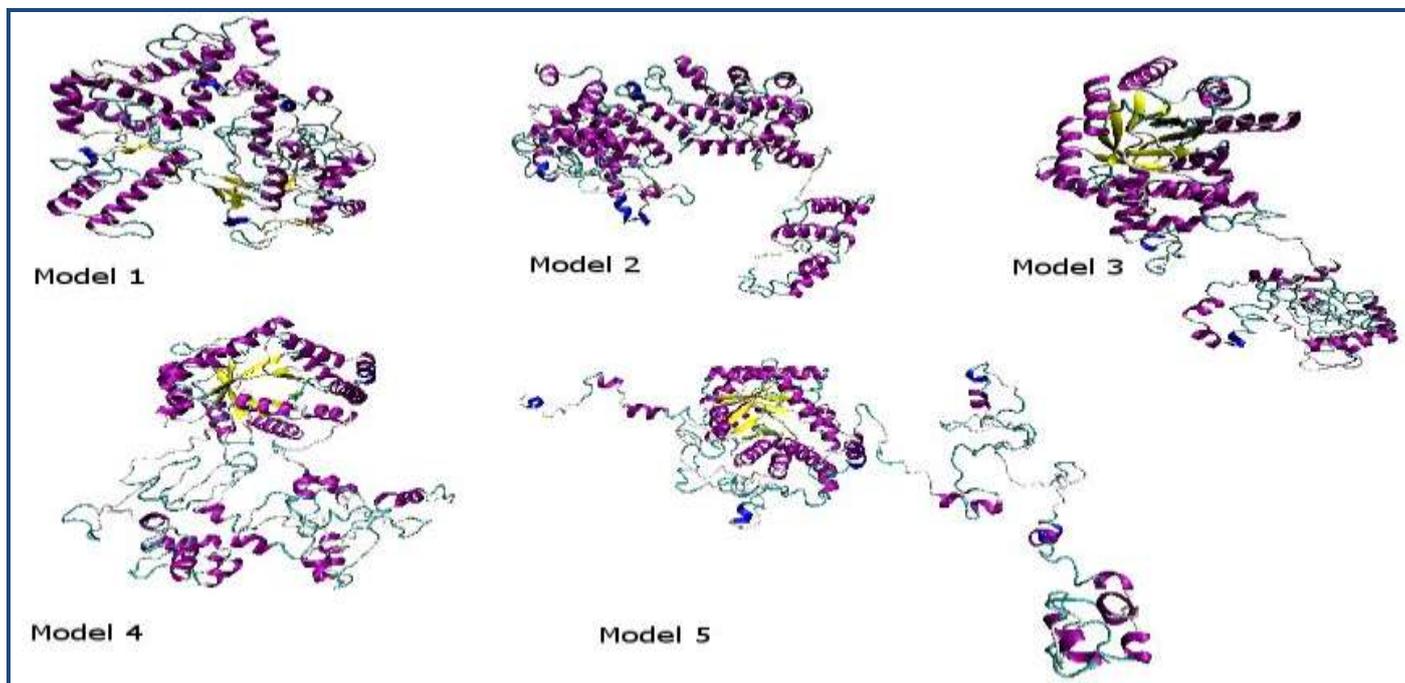

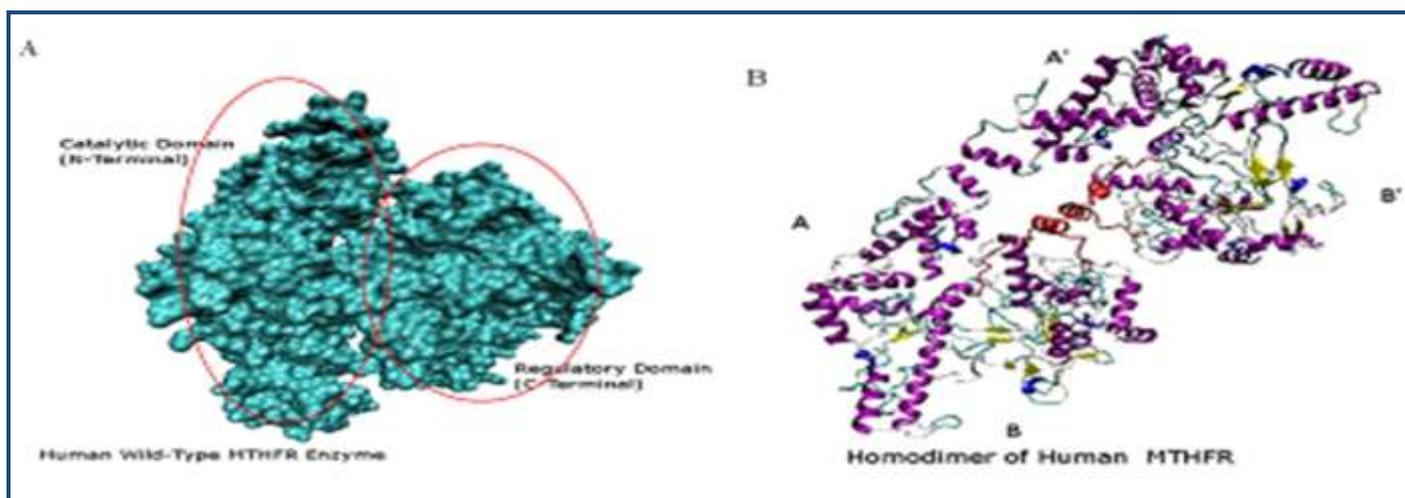**Figure 2:** The models 1-5 are generated from the I-TASSER server and visualized by the VMD.

Figure 3: **(A)** Two domains (Surf view) of wild-type Human MTHFR structure. **(B)** Homodimer view of Human MTHFR.

The five complete models obtained from I-TASSER server were evaluated for the one best model. **Table 2 (see supplementary material)** shows the structural evaluation results using DFIRE, Verify3D and Procheck. The DFIRE program uses energy value to evaluate the structures. According to these results the lowest energy is of model number 1. The Verify3D program assigned the highest positive score to model number 4. To analyze the stereochemical quality of the predicted structure Procheck software was used. According to Procheck results, model number 1 is the most stable because it has less area in the disallowed region (2.3% as compared to the other models).

Model number 1 was finally used for further analysis based on energy, stereochemical results and I-TASSER structure prediction server ranking.

Two domains were predicted by the DOMAC server. The domain prediction was made by the template-based methods using the protein templates available in Protein Data Bank **(Figure 3A). Figure 3(B)** shows the protein as a homodimer. The first domain stretches from amino acid 1 to 342 while the second domain is from amino acid 343 to 656. The phosphorylation sites predicted by the Motif Scan servers are shown in the figure

4 (A-B). **Figure 5 (A, B)** shows the FAD-binding site of the catalytic domain, S-Adenosylmethione-binding site, Trypsin cleavage site (KRREED) and a conserved amino acid motif (VRPIFW) in the regulatory domain. Two possible NADPH-binding sites are shown in **(Figure 6)**.

After selecting the best 3-D model of the wild type MTHFR, its amino acid sequence was mutated and then the structures of the three mutated structures were predicted using the I-TASSER server. After this step, different core sites were highlighted which can be affected due to mutations **(Figure 7).** A total of 22 Serine, 6 Threonine and 3 Tyrosine phosphorylation sites were predicted by NetPhos 2.0 server for the wild-type Human MTHFR. There were differences found in Serine phosphorylation sites, which were 21 in both mutated sequences having a single amino acid mutation and 20 sites were predicted in MTHFR sequence having both the mutations. **Figure 8** (A-D) shows the phosphorylation sites in normal and mutated structures predicted by NetPhos 2.0 server. The FAD-binding sites in both wild type and mutant are shown in the **(Figure 9).** S-Adenosylmethione-binding site mutations are also visible in the structure **(Figure 10).** The complete view of the model is shown in **(Figure 11).**
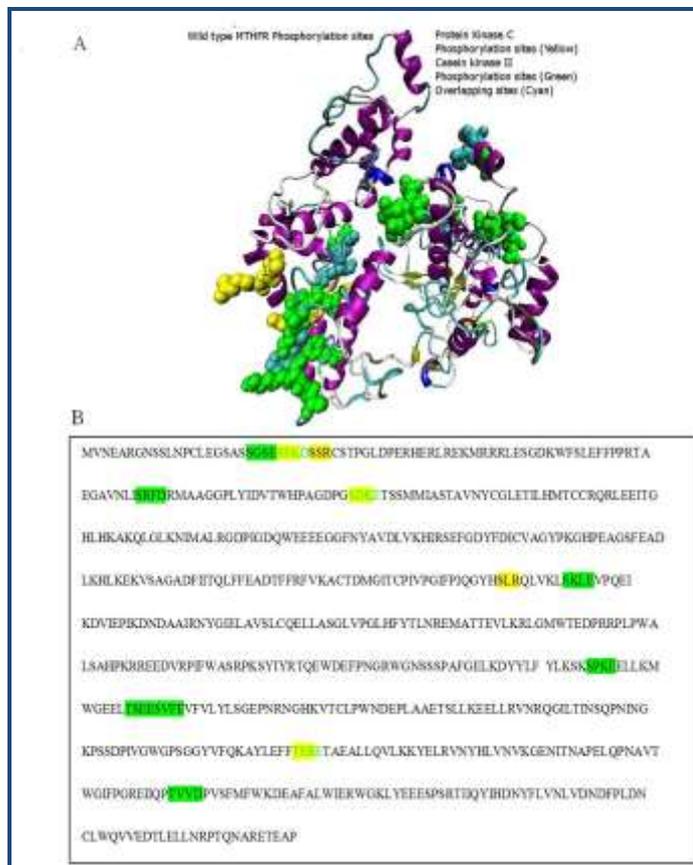
Figure 4: Phosphorylation binding sites: **A)** Protein Kinase C phosphorylation sites are colored in yellow, Casein Kinase II phosphorylation sites are colored in green, while the overlapping phosphorylation sites of the both enzymes are colored in cyan as show in the figure A; **B)** Protein Kinase C phosphorylation sequence is colored in yellow, Casein Kinase II phosphorylation sequence is colored in green, while the green colored sequence highlighted with yellow background is the overlapping sequence of both the enzymes.
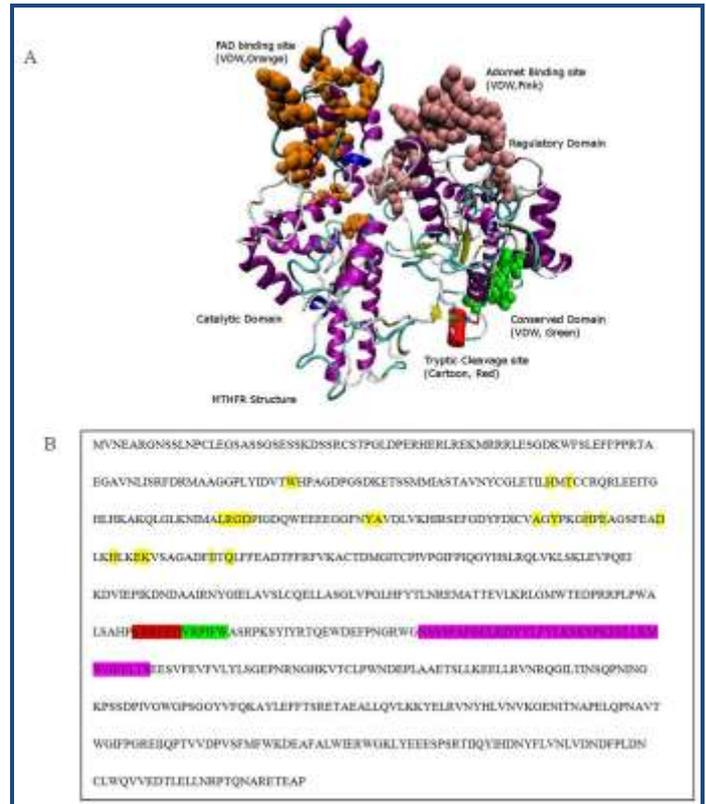

Figure 5: Validation by Experimental Evidences: **A)** FAD binding site at the N-terminus, S-Adenosylmethione-binding site in the regulatory domain, tryptic cleavage, and a conserved domain is shown in the wild type MTHFR; **B)** MTHFR sequence showing the FAD binding site in yellow color, Adomet binding site in the pink color, tryptic cleavage site in the red color and the conserved domain in the green color.
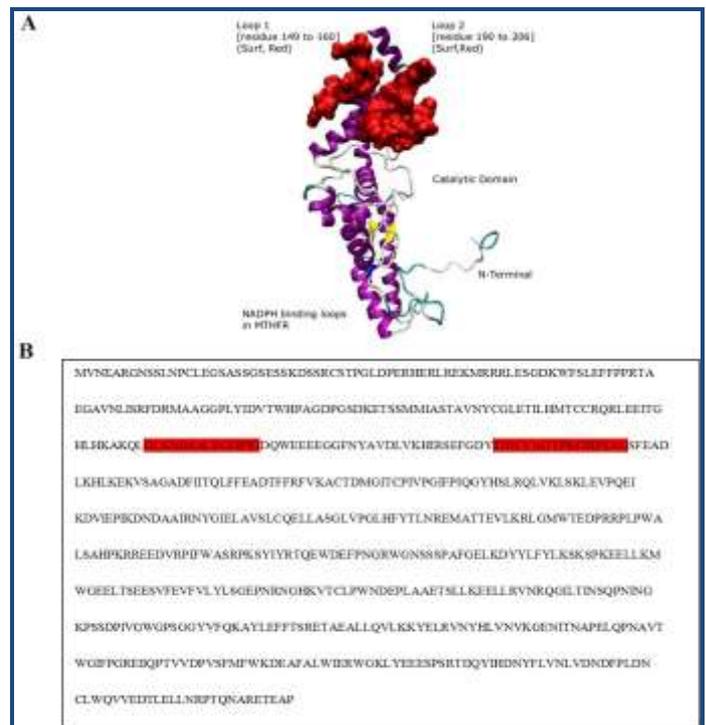

Figure 6: NADPH-binding sites: **A)** Structural view of supposed NADPH binding sites Loop 1 and Loop 2 are colored in red to

indicate the NADPH-binding sites; **B)** The MTHFR sequence from 149 to 160 residues indicates the loop 1 region while sequence from 190 to 206 residues (colored in gray) indicates the loop 2 region.
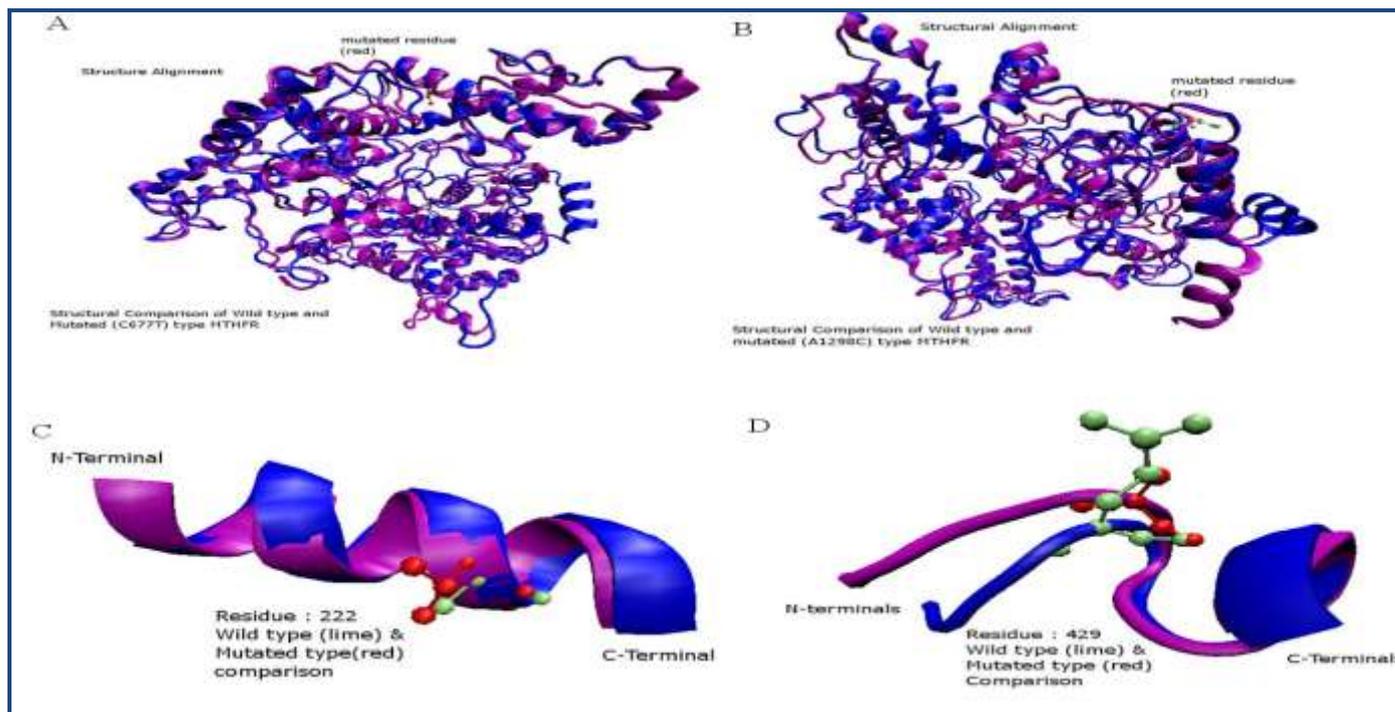
Figure 7: A) Structural alignment of Normal (Blue) and Mutated (Purple) amino acids of MTHFR. Residue number 222 of both types is shown (CPK view). Normal amino acid *Alanine* is shown in the lime color while mutated amino acid *Valine* is shown in the red color; **B)** Structural alignment of the normal and mutated amino acids of MTHFR. *Glutamic Acid* is shown in the lime color which is replaced by *Alanine* in the mutated structure; **C)** Close look of A222V mutated structure. Blue colored helix shows the normal structure while purple colored helix represents the mutated structure; **D)** Close look of E429A mutated structure. Blue colored helix shows the normal structure while purple colored helix represents the mutated structure.
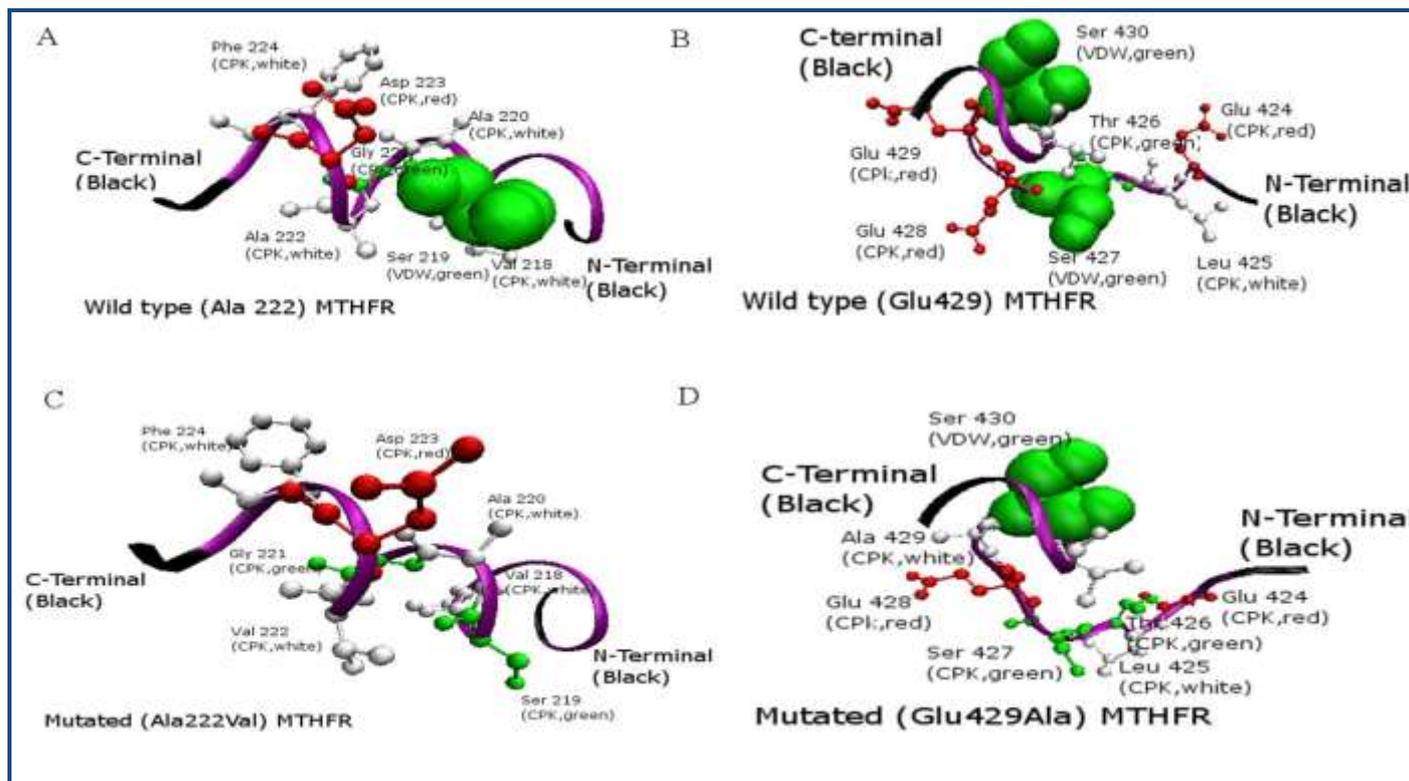
Figure 8: Comparison of Phosphorylation sites of normal and mutated sites: **A)** Wild-type Human MTHFR model with active Serine Phosphorylation residue at its position 219; **B)** Wild type MTHFR with active Serine residue at its position 427 and 430; **C)**

Mutated (Ala222Val) MTHFR showing that there is no active Serine residue for Phosphorylation at position 219; **D)** Mutated (Glu429Ala) MTHFR model having non-phosphorylated serine residue at position 427.

Discussion:

In silico methods were used to determine the structural difference between the wild type and mutant human MTHFR enzyme. The predicted structures show β8α8 barrel catalytic domains that bind FAD. In order to understand the underlining factors responsible for the malfunction of mutants of MTHFR, structural studies were carried out by using different bioinformatics tools and software. The sequence alignment and phylogenetic analysis was carried out using PSI-BLAST and ClustalW software. Only two experimentally determined structures were present in the Protein Data Bank (PDB), namely from *E. coli* (PDB code: 1b5t) **[38]** and from *Thermus thermophilus* (PDB code:1v93) **[39, 40].** Using these two known structures, homology modeling for tertiary structure prediction was carried out. Only the structure of first domain of Human MTHFR was predicted from homology modeling results based on the available templates of *E. coli* and *Thermus thermophilus* in the Protein Data Bank. To predict the complete structure of Human MTHFR, the threading technique was applied. I-TASSER server predicted the complete structural results by threading. The models predicted by this server were used for further analysis.

The X-ray analysis of *E. coli* MTHFR provides a model for the catalytic domain that is shared by all MTHFRs. This domain is a β8α8 barrel that binds FAD in a novel fashion **[38]**. Ala177, which corresponds to Ala222 in human MTHFR, is located near the bottom of the barrel and is distant from FAD. The mutation A177V does not affect Km or kcat but instead increases the propensity for bacterial MTHFR to lose its essential FAD cofactor. Folate derivatives protect the wild-type and mutant *E. coli* enzymes against the loss of FAD and also against thermal inactivation.

According to the known structure of bacterial MTHFR enzyme, FAD-binding residues are located near the N-terminus of MTHFR **[38].** It was observed that these FAD-binding residues have a compact structure for binding of FAD cofactor in the wild-type human MTHFR. When these residues were color-coded in the predicted structure of Ala222Val mutant, it was observed that the FAD-binding groove appeared broadened. This result indicates that FAD binding in the catalytic domain was disturbed which resulted in the distortion of the original structure of wild-type human MTHFR. Trypsin cleavage of native enzyme was also shown in the wild-type human MTHFR structure, which also confirmed the two domains in the structure.

According to Yamada *et al* **[9],** phosphorylation is crucial for the regulation of human MTHFR. It acts as an on and off switch for the expression of MTHFR. NetPhos 2.0 server was used for the prediction of phosphorylation sites and it predicted a total of 22 Serine, 6 Threonine and 3 Tyrosine sites. It was analyzed that there was no change in the phosphorylation sites of Threonine and Tyrosine residues and only the Serine phosphorylation sites were perturbed. In mutated Ala222Val and Glu429Ala mutants, a total of 21 Serine active sites were predicted which were one less than the total sites predicted in the wild type MTHFR enzyme. It was found that double mutants, containing both Ala222Val and Glu429Ala mutations, a lower number of serine

phosphorylation sites are present as compared to the singly mutated structures mentioned before. Cysteine phosphorylation sites were also affected by mutations. These sites were absent in the mutated region.
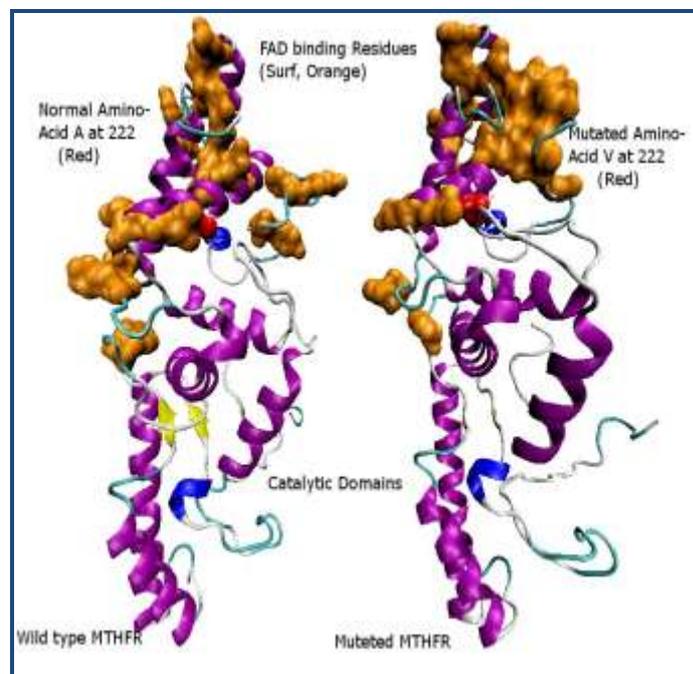
Figure 9: Comparison of FAD-binding sites in normal and mutated structures. FAD-binding region is highlighted in orange color present in the catalytic region. The structural perturbance can be seen very clearly in the figure due to the effect of mutations.
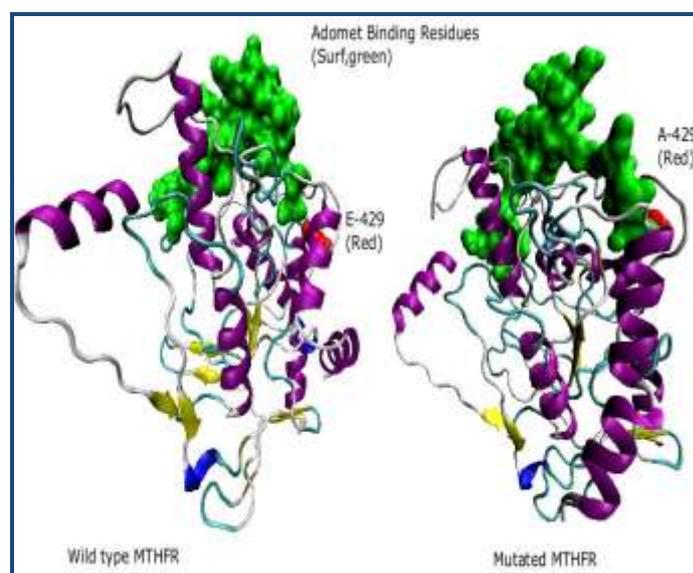
Figure 10: S-Adenosylmethione-binding sites of both mutatnats are shown in the figure. The difference between the two sites can be visualized clearly. S-Adenosylmethione-binding site in the mutated structure is distorted as compared to the normal structure.

BIOINFORMATION
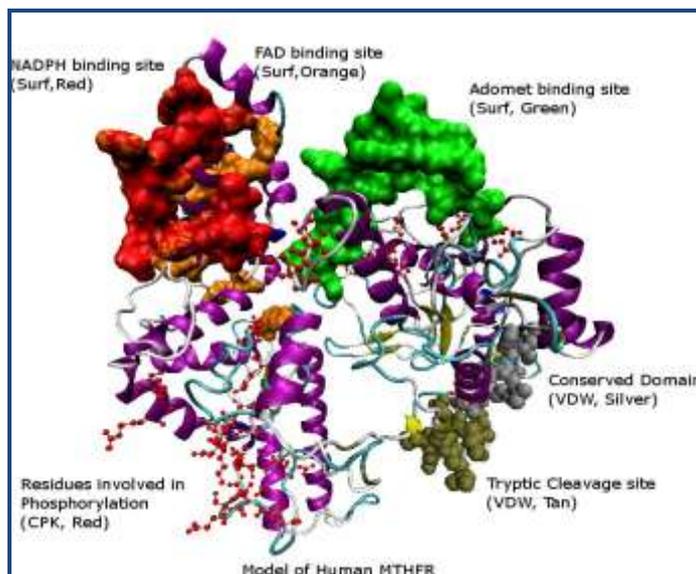
Figure 11: A model of the Human MTHFR structure showing different sites by color-coding in the structure.

Conclusion:
It has been shown that the mutated amino acids of human MTHFR affect the phosphorylation sites in the protein structure by affecting the FAD-binding sites. Phosphorylation seems to play a significant role in the functioning of the wild-type enzyme. However, the structural basis for this observation is not clear. Hence, molecular models of the MTHFR mutants were developed to elucidate the thermodynamically unfavorable observations in comparison to the wild type. This provides insights to the understanding of structure-function relationship of MTHFR mutants and its wild type for developing treatments in hyperhomocysteinemia, homocystinuria, thyroid dysfunction and colorectal cancer.

References:
[1] Zhang Z *et al*. *Eur J Med Chem.* 2012 **58**: 228 [PMID: 23124219]
[2] Goyette P *et al*. *Nat Genet.* 1994 **7***:* 195 [PMID: 7920641]
[3] Kutzbach C & Stokstad EL, *Biochim Biophys Acta.* 1971 **250**: 459 [PMID: 4399897]
[4] Sumner J *et al*. *J Biol Chem.* 1986 **261**: 7697 [PMID: 3754872]
[5] Roje S *et al*. *J Biol Chem.* 2002 **277**: 4056 [PMID: 11729203]
[6] Matthews RG *et al*. *J Biol Chem.* 1984 **259**: 11647 [PMID: 6384210]
[7] Sheppard CA *et al*. *J Bacteriol.* 1999 **181**: 718 [PMID: 9922232]
[8] Jencks DA & Matthews RG, *J Biol Chem.* 1987 **262**: 2485 [PMID: 3818603]
[9] Yamada K *et al*. *Proc Natl Acad Sci USA.* 2005 **102**: 10454 [PMID: 16024724]
[10] Yamada K *et al*. *Proc Natl Acad Sci USA.* 2001 **98**: 14853 [PMID: 11742092]
[11] Jacques PF *et al*. *Circulation.* 1996 **93**: 7 [PMID: 8616944]
[12] Friso S *et al*. *Clin Exp Med.* 2002 **2**: 7 [PMID: 12049191]
[13] Peng F *et al*. *Int J Mol Med.* 2001 **8**: 509 [PMID: 11605019]
[14] Hustad S *et al*. *Am J Clin Nutr.* 2004 **80**: 1050 [PMID: 15447919]
[15] Le Marchand L *et al*. *Cancer Epidemiol Biomarkers Prev.* 2005 **14**: 1198 [PMID: 15894672]
[16] Gasteiger E *et al*. *Nucleic Acids Res.* 2003 **31**: 3784 [PMID: 12824418]
[17] Altschul SF *et al*. *Nucleic Acids Res.* 1997 **25**: 3389 [PMID: 9254694]
[18] Thompson JD *et al*. *Nucleic Acids Res.* 1994 **22**: 4673 [PMID: 7984417]
[19] Nielsen M *et al*. *In Press - NAR web issue.* 2010
[20] Lambert C *et al*. *Bioinformatics* 2002 **18**: 1250 [PMID: 12217917]
[21] Bystroff C & Shao Y, *Bioinformatics* 2002 **18**: S54
[22] Bystroff C *et al*. *J Mol Biol.* 2002 **301**: 173 [PMID: 10926500]
[23] Arnold K *et al*. *Bioinformatics* 2006 **22**: 195 [PMID: 16301204]
[24] Chen CC *et al*. *BMC Bioinformatic.* 2009 **10**: 366 [PMID: 19878598]
[25] Zhang Y, *Proteins.* 2007 **69**: 108 [PMID: 17894355]
[26] Wu S & Zhang Y, *Nucleic Acids Res.* 2007 **35**: 3375 [PMID: 17478507]
[27] Zhou H & Zhou Y, *Protein Sci.* 2002 **11**: 2714 [PMID: 12381853]
[28] Luthy R *et al*. *Nature.* 1992 **356**: 83 [PMID: 1538787]
[29] Vuister GW *et al*. *J Biomol NMR.* 2013 [Epub ahead of print] [PMID: 23877928]
[30] Morris AL *et al*. *Proteins.* 1992 **12**: 345 [PMID: 1579569]
[31] Cheng J, *Nucleic Acids Res.* 2007 **35**: W354 [PMID: 17553833]
[32] Tress M *et al*. *Proteins.* 2007 **69**: 137 [PMID: 17680686]
[33] Larson J *et al*. *J Virol.* 2007 **81**: 6957 [PMID: 17442718]
[34] Sadowski MI, *Proteins.* 2013 **81**: 253 [PMID: 22987736]
[35] Blom N *et al*. *J Mol Biol.* 1999 **294**: 1351 [PMID: 10600390]
[36] Pagni M *et al*. *Nucleic Acids Res.* 2007 **35**: W433
[37] Humphrey W *et al*. *J Mol Graph.* 1996 **14**: 33 [PMID: 8744570]
[38] Guenther B D *et al*. *Nat Struct Biol.* 1999 **6**: 359 [PMID: 10201405]
[39] http://www.pdb.org/pdb/explore/explore.do?structureId=1V93
[40] Pejchal R *et al*. *Biochemistry* 2006 **45**: 4808 [PMID: 16605249]

BIOINFORMATION

Supplementary material:

Table 1: The comparison of Human MTHFR sequence with MTHFR sequences of selected species. Residue numbers, SwissProt ID, Percentage Identity and PDB codes are shown in the table.

No.	Species [a]	Residues	SwissProt ID	Percentage Identity	PDB ID
1.	MTHR_MACFA	656	Q60HE5	(639/656) 97%	---
2.	MTHR_BOVINE	655	Q51598	(591/655) 90%	---
3.	MTHR_MOUSE	654	Q9WU20	(589/657) 89%	---
4.	MTHR_CAEEL	663	Q17693	(298/614) 48%	---
5.	MTHR_ORYSJ	594	Q75HE6	(264/596) 44%	---
6.	MTHR1_MAIZE	593	Q9SE94	(263/596) 44%	---
7.	MTHR1_ARATH	592	Q9SE60	(260/597) 43%	---
8.	MTHR2_ARATH	594	O80585	(259/597) 43%	---
9.	MTHR1_SHPO	641	Q10258	(248/584) 42%	---
10.	MTHR2_YEAST	600	P53128	(251/599) 41%	---
11.	MTHR1_YEAST	657	P46151	(114/292) 39%	---
12.	THERMUS THERMOPHILUS	296	---	(110/282) 39%	1v93
13.	E.COLI	275	P0AEZ1	(96/279) 34%	1b5t
14.	MTHR2_SHPO	641	O74927	(228/653) 34%	---
15.	METF_SHIFL	296	POAEZ2	(100/300) 33%	---

[a] Species names of MTHFR enzyme were Selected from the PSI-BLAST Server

Table 2: Evaluation of the Models by DEFIRE, Verify3D and Procheck software.

Software		Model 1	Model 2	Model 3	Model 4	Model 5
DFIRE		- 44247.74	- 42550.36	- 41972.78	- 42053.35	- 40490.94
Verify3D		- 0.25 to 0.62	- 0.26 to 0.65	- 0.15 to 0.71	- 0.14 to 0.74	- 0.32 to 0.71
Procheck	**Core Region**	79.2%	82.9%	79.8%	78.0%	76.1%
	Allowed Region	12.3%	11.1%	12.7%	13.5%	13.9%
	Generous Region	6.2%	2.5%	4.2%	4.8%	6.5%
	Disallowed Region	2.3%	3.5%	3.3%	3.7%	3.5%