Comparative modeling of methylentetrahydrofolate reductase (MTHFR) enzyme and its mutational assessment: *in silico* approach

Dowlathabad Muralidhara Rao[1]*, Anuraj Nayarisseri[2], Mukesh Yadav[2], Showmy K.S.[3], Disha Patel[4]

[1]Department of Biotechnology, Bioinformatics Laboratory Sri Krishnadevaraya University, Anantapur, AP, India, muralidhararao@yahoo.com Ph.08554 255544 Fax.08554 255720
[2]Eminent Biosciences, Bioinformatics Research Laboratory, Indore, Madhyapradesh, India
[3]Sree Narayana Guru Institute of Science and Technology, North Paravur, Cochin, Kerala, India
[4]Softvision College, Viajaynagar, Indore, Madhyapradesh, India

Abstract- DNA-synthesis, DNA-repair, and DNA imprinting processes require efficient conversion of homocysteine to methionine. This methylation is catalyzed by methylentetrahydrofolate reductase through reduction of 5,10-methylenetetrahydrofolate into 5-methyltetrahydrofolate. Normal DNA synthesis is considered critical for physiological functions of body. The enzyme is coded by the gene with the symbol MTHFR on chromosome 1 location p36.3 in humans. At least 24 mutations in the MTHFR gene have been identified in people with homocystinuria. There is DNA sequence variants (genetic polymorphisms) associated with this gene. Two of the most investigated are C677T (rs1801133) and A1298C (rs1801131) single nucleotide polymorphisms (SNP). Mutations at C677T and A1298C which confer amino acid substitution Ala222Val and Glu429Ala respectively with a considerable reduced activity. This polymorphism and mild hyperhomocysteinemia are associated with neural tube defects in offspring, arterial and venous thrombosis, and cardiovascular disease. 677TT individuals are at a decreased risk for certain leukemia and colon cancer. The MTHFR gene could be one of the factors of overall schizophrenia risk. *In silico* analysis now has added important and wide range applications to proteomics from structure modeling to its functional levels. Several algorithms have been suggested from many authors to bring an accurate modeling at its best but ultimately every protein has its own variant features to be treated by the same algorithm. Studies in proteomics through computational techniques need complements between critical requirement for a protein and features available in an algorithm. Comparative modeling is now bridging the gap between available sequences and structures modeled with accuracy. Effective refinement techniques made it capable of driving models toward native structure. Structure of MTHFR can assist the study of involvement of this enzyme in the disorders and can provide better level of understanding about structural aspects of it. We have modeled wild type and mutated type MTHFR using comparative modeling and structure validation has given appreciable values. This work can further account for the structure based drug design community in the search of MTHFR inhibitors.

Keywords- MTHFR, Comparative modeling, PDB, Mutation, Toxicity

Introduction

The enzyme 5, 10-methylentetrahydrofolate reductase (MTHFR) regulates reductive parameters in conversion of 5, 10-methylentetrahydrofolate to 5-methylentetrahydrofolate availing methionine synthesis for DNA methylation. Activity of MTHFR enzyme is associated with polymorphism in MTHFR gene. Mutations at C677T and A1298C which confer amino acid substitution Ala222Val and Glu429Ala respectively with a considerable reduced activity [1].Two common allele variants of the MTHFR gene have been described, C677T (NCBI SNP ID: rs1801133) and A1298C (rs1801131), that lead to amino acid substitutions, Ala222Val and Glu429Ala, and to decreased enzyme activity [2-3]. The effect of the 1298C allele variant is less severe and homozygous carriers of this allele have a more moderate 30–40% reduction of the enzyme activity, yet its function remains controversial. Furthermore, people who are heterozygous at both loci, C677T and A1298C, experience an intermediate activity loss of 40–50%. The 677T variant increases the plasma homocysteine concentration in humans and reduces DNA methylation in cancer patients [4-5]. Although direct influence of polymorphism is not confirmed with breast cancer development in studies so far but chemo sensitivity of cancer cells towards some known drugs is found to be modulated by it. High toxicity has been reported towards chemotherapeutic agents like cyclophosphamide, methotrexate and flurouracil [6-7]. Also some studies suggest 2.8 fold increased risk in endometrial cancer and 2.9 in case of cervical intraepithelial neoplasia [8-9]. Mutant MTHFR gene at C677T is also involved in Neural tube defects and cardiac anomalies. To its broader spectrum as causing factor some other disease such as preovulatory overripeness ovapathy (PrOO). A Comparative study reveals no difference between Glu429Ala mutant protein and the wild-type protein however there appears a remarkable properties difference between Ala222Val mutant and wild type enzyme [10]. In our present work we have modeled wild type MTHFR and mutant with Ala222Val MTHFR

structures. Structure validations have been achieved with significant stereo chemical parameters. Established drugs show low binding affinity and high toxicity towards mutated MTHFR in chemotherapy. In present studies on MTHFR relation between toxicity on established drugs and mutated structure has been targeted with future prospective to develop efficient inhibitors for mutated MTHFR.

Method

Location of gene

Cytogenetic Location: 1p36.3
Molecular Location on chromosome 1: base pairs 11,769,246 to 11,788,568.
The MTHFR gene is located on the short (p) arm of chromosome1at position 36.3. More precisely, the MTHFR gene is located from base pair 11,769,246 to base pair 11,788,568 on chromosome1.

Sequence Retreival

Methylenetetrahydrofolate reductase sequence has been retrieved from SwissProt / Uniprot, which is curated and annotated database. SwissProt provides descriptions of a nonredundant set of proteins, including their function, domain structure, posttranslational modifications and variants from http://www.genome.gov/page.cfm?pageID=10005283 [11]. There are total 565 amino acid residues found in Methylenetetrahydrofolate reductase sequence. Molecular weight of the protein calculated to be 74596.5 by Emboss.

Template for MTHFR modeling

Structure similarity searching was performed by standalone blastp against PDB database which was downloaded from ftp://ftp.ncbi.nih.gov/blast/db/FASTA/pdbaa.gz.
Results when analyzed gave a protein structure 1V93 A chain showing 39 % identity and 56% similarity, which is considered to be good (>than 30%) for homology or comparative modeling of Methylenetetrahydrofolate reductase. 1V93 A chain is a structure of 5,10-Methylenetetrahydrofolate reductase from *Thermus thermophilus* HB8 obtained by X-ray diffraction at 1.90 Å resolution determined by Nakajima et al available on http://www.pdb.org/pdb/explore/explore.do?structureId=1V93 [12].

Comparative Modeling of MTHFR

We used ICM Molsoft for comparative modeling of MTHFR. The ICM Comparative modeling algorithm has come up with highly accurate and more robust modeling tools. All side chain torsion angles coming from non-identical residues are predicted using global energy minimization protocols [13-15]. ICM Biased Probability Monte Carlo (BPMC) optimization in combination of internal coordinates is applied to conformational modeling of protein side chains and loops [16]. The ICM homology modeling algorithm has demonstrated excellent accuracy in blind predictions at the CASP2 competition 6 and in several protein engineering applications. Loop prediction in ICM is enriched by loop PDB database to confer structure quality of comparative modeling.

WILD TYPE MTHFR modeling

Before modeling the structure position of Ala 222 and Glu429 confirmed in the sequence in wild type to assure further mutagenesis studies. Alignment of MTHFR sequence with 1v93 template and loop modeling with fast and efficient ICM algorithm provided a good structure of MTHFR (Figure-1). After energy minimization it was calculated to be -7447.150 KJ/mol. Lower energy provides thermodynamically stable structure for further validation Ramachandran plot has been taken under study. It shows 94.6% (226) of residues in most favored region, 4.6% (11) of residues in additional allowed region, 0.4% (1) in generously allowed region and 0.4% (1) residue in disallowed region. Following the definition of standard protein structure according to Ramachandran plot at least 90% residues would cover most favored region [17]. Stereo chemical aspects are demonstrated by PROCHECK studies for structure modeled. Main chain parameters and side chain parameters occupy better region and G-factors for normal probability also exhibit appreciable numerals for dihedral and covalent interaction among amino acids. Planar groups are under ideal values although structure seems to suffer some bad contacts (6) in PROCHECK analysis. Ramachandran Plot and PROCHECK analysis for wild type MTHFR structure are shown in figure -2.

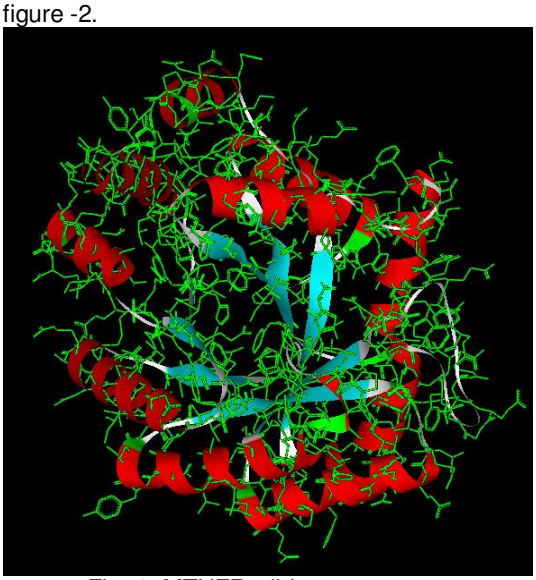


Fig. 1- MTHFR wild type structure

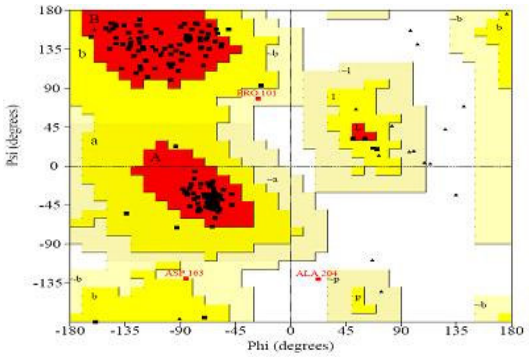
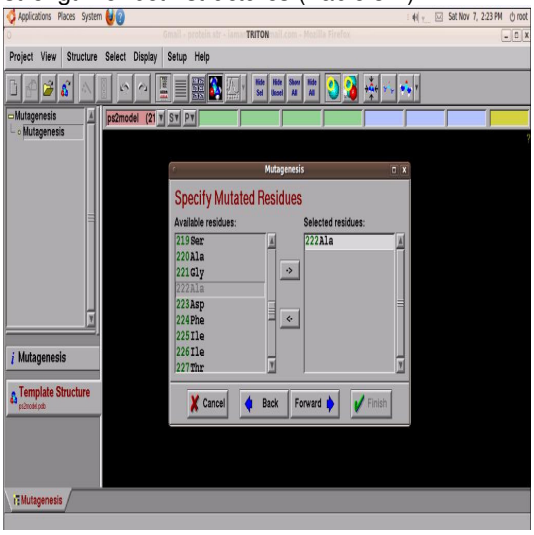


Fig. 2- Ramachandran Plot for wild type MTHFR

Ala222Val Mutation in MTHFR

The program TRITON is a graphical tool for computational aided protein engineering. It implements methodology of computational site-directed mutagenesis to design new protein mutants with required properties. New site directed protein mutants are modeled from their wild type using an external program MODELLER [18-20]. A comparative study show no difference Glu429Ala mutant protein and the wild type therefore with Ala222Val MTHFR is targeted in present work to model and study for activity with wild type. Mutated structure processed for energy minimization and its energy calculated to -9048.897 KJ/mol. This energy of mutated structure shows thermodynamically favorable change to be submitted to wild type. This study is also approved by the bad contacts of mutated form (4) as in comparison with bad contacts of wild type (6). In many aspects of Bond length, bond angles and bad contacts Ramachandran plot and PROCHECK analysis appear improved for mutant MTHFR. Consideration of main-chain parameters also proves better structural features of mutant structure over wild type (Table 1-2). Side-chain parameters appear to be of equal strength for both structures (Table 3-4).



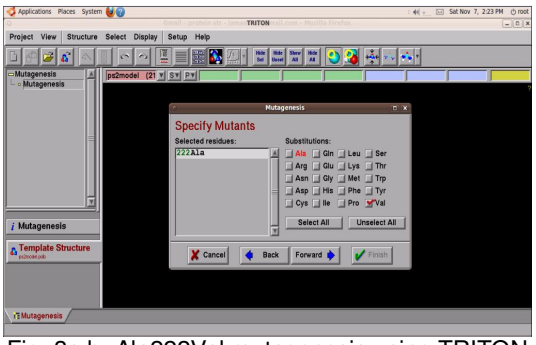


Fig. 3a,b- Ala222Val mutagenesis using TRITON programme

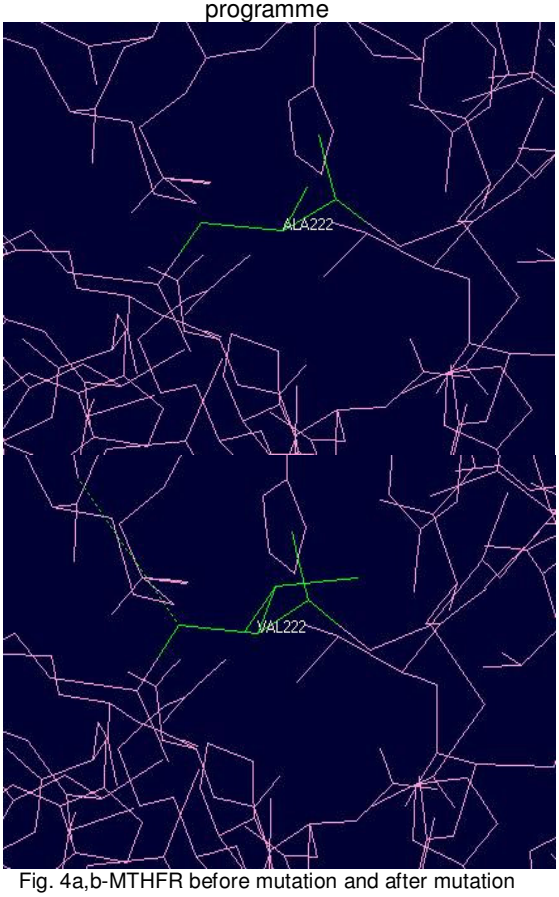


Fig. 4a,b-MTHFR before mutation and after mutation

Results and Discussion

Comparative modeling and mutational analysis for MTHFR yielded results in support of favorable mutational events occurring in enzyme to comprehensive stability of mutated form over wild enzyme structure. This result can also be referred towards toxicity imparted by mutated form against established drugs brought in use to treat disorders in chemotherapy. Comparative modeling of MTHFR reveals that mutations brought thermodynamically favorable change in the structure over wild type. Thus energetic of mutation confirms stability of Ala222Val mutant of MTHFR. Structural features explain existence of better main-chain parameters for MTHFR in its mutant structure and can be further studied in

relation with other mutations occurring in enzymes it terms of their energy and stereo chemical aspects. Present *in silico* study of has provided structures of both wild and mutant MTHFR which would definitely assist structural based drug design community to accelerate the search of suitable inhibitors for it. The main aim is to identify, exploit and analysis of new molecular drug targets at structural level. This computational approach will lead to the discovery and structural development of novel drug targets. Computational community can further explore active site of MTHFR for binding of drug and apply docking studies to indentify amino acids involved in electrostatic, hydrophobic and hydrogen bond formation with inhibitors of this enzyme with special distinction to wild and mutated form.

References

[1] Frosst P., Blom H.J., Milos R., Goyette P., Sheppard C.A., Matthews R.G., Boers G.J., den Heijer M., Kluijtmans L.A., van den Heuvel L.P., Rozen R.A. (1995) *Nature Genet*., 10:111-113.

[2] Weisberg I., Tran P., Christensen B., Sibani S., Rozen R. (1998) *Mol Genet Metab.,* 64:169-172.

[3] Weisberg I.S., Jacques P.F., Selhub J., Bostom A.G., Chen Z., Curtis E.R., Eckfeldt J.H., Rozen R. (2001) *Atherosclerosis*, 156:409-415.

[4] Yamada K., Chen Z., Rozen R., Matthews R.G. (2001) *Proc Natl Acad Sci U S A*, 98:14853-14858.

[5] Paz M.F., Avila S., Fraga M.F., Pollan M., Capella G., Peinado M.A., Sanchez-Cespedes M., Herman J.G., Esteller M. (2002) *Cancer Res*., 62:4519-4524.

[6] Justenhoven C., Hamann U., Pierl C.B., Rabstein S., Pesch B., Harth V., Baisch C., Vollmert C., Illig T., Bruning T., Ko Y., Brauch H. (2005) *Cancer Epidemiol Biomarkers Prev*., 14:3015-3018.

[7] Toffoli G., Veronesi A., Boiocchi M., Crivellari D. (2000) *Ann Oncol*., 11:373-374.

[8] Esteller M., Garcia A., Martinez-Palones J.M., Xercavins J., Reventos J. (1997) *Carcinogenesis*, 18:2307-2311.

[9] Piyathilake C.J., Macaluso M., Johanning G.L., Whiteside M., Heimburger D.C., Giuliano A. (2000) *Anticancer Res.,* 20:1751-1757.

[10] Kazuhiro Yamada, Zhoutao Chen, Rima Rozen, and Rowena G. Matthews. (2001) *PNAS*, 98(26), 14853–14858.

[11] http://www.genome.gov

[12] http://www.pdb.org

[13] Schapira M., Raaka B.M., Samuels H.H. and Abagyan R. (2001) *BMC Structural Biology* 1:1.

[14] Schapira M., Abagyan R. A.and Totrov M. M. (2002) *BMC Struct Biol* 2(1):1.

[15] Schapira M., Abagyan R. A. and Totrov M. M. (2003) *J Med Chem* 46(14):3045-3059.

[16] Abagyan R.A. and Totrov M.M. (1996) *Abstr. Pap. Am. Chem. Soc*. 211:35-COMP.

[17] http://www.biochem.ucl.ac.uk

[18] Prokop M., Adam J., Kriz Z., Wimmerova M., Koca J. (2008) *Bioinformatics* 24: 1955-1956.

[19] Prokop M., Damborsky J., Koca J. (2000) *Bioinformatics* 16: 845-846.

[20] Damborsky J., Bohac M., Prokop M., Kuty M., Koca J. (1998) *Protein Engineering* 11: 901-907.

Table 1- Plot statistics for main-chain parameters for Wild MTHFR

Stereo chemical parameters	No. of data pts.	Parameters value	Comparison typical value	Value band width	No. of band width from mean	comment
%-tage residues in A, B, L	239	94.6	83.8	10	1.1	BETTER
Omega angle st dev	278	4	6	3	-0.7	Inside
Bad contacts / 100 residues	6	2.2	4.2	10	-0.2	Inside
Zeta angle st dev	255	1.3	3.1	1.6	-1.1	BETTER
H-bond energy st dev	185	0.7	0.8	0.2	-0.4	Inside
Overall G-factor	279	0	-0.4	0.3	1.2	BETTER

Table 2- Plot statistics for main-chain parameters for mutated MTHFR

Stereo chemical parameters	No. of data pts.	Parameters value	Comparison typical value	Value band width	No. of band width from mean	comment
%-tage residues in A, B, L	239	93.3	83.8	10	1	Inside
Omega angle st dev	278	5.8	6	3	-0.1	Inside
Bad contacts / 100 residues	4	1.4	4.2	10	-0.3	Inside
Zeta angle st dev	255	1.8	3.1	1.6	-0.8	Inside
H-bond energy st dev	187	0.8	0.8	0.2	-0.1	Inside
Overall G-factor	279	0.1	-0.4	0.3	1.5	BETTER

Table 3- Plot statistics for Side-chain parameters for wild MTHFR

Stereo chemical parameters	No. of data pts.	Parameters value	Comparison typical value	Value band width	No. of band width from mean	comment
Chi-1 gauche minus st dev	43	6.1	18.1	6.5	-1.8	BETTER
Chi-1 trans st dev	91	8.4	19	5.3	-2	BETTER
Chi-1 gauche plus st dev	85	7.6	17.5	4.9	-2	BETTER
Chi-1 pooled st dev	219	7.8	18.2	4.8	-2.1	BETTER
Chi-2 trans st dev	55	9.1	20.4	5	-2.3	BETTER

Table 4- Plot statistics for Side-chain parameters for mutated MTHFR

Stereo chemical parameters	No. of data pts.	Parameters value	Comparison typical value	Value band width	No. of band width from mean	comment
Chi-1 gauche minus st dev	43	7.3	18.1	6.5	-1.7	BETTER
Chi-1 trans st dev	91	8.8	19	5.3	-1.9	BETTER
Chi-1 gauche plus st dev	86	9.6	17.5	4.9	-1.6	BETTER
Chi-1 pooled st dev	220	9	18.2	4.8	-1.9	BETTER
Chi-2 trans st dev	55	10.5	20.4	5	-2	BETTER