

Original Article

STRUCTURAL AND FUNCTIONAL ANALYSIS OF AF9-MLL ONCOGENIC FUSION PROTEIN USING  
HOMOLOGY MODELING AND SIMULATION BASED APPROACH

MEDHA DAVE<sup>1</sup>, ADITI DAGA<sup>2</sup>, RAKESH RAWAL<sup>3\*</sup>

<sup>1</sup>Department of Bioinformatics, Maharaja Krishnakumarsinhji Bhavnagar University, Bhavnagar, Gujarat, India, <sup>2</sup>Department of Microbiology, MVM Science Collage, Saurashtra University, Rajkot, Gujarat, India, <sup>3</sup>Department of Cancer Biology, The Gujarat Cancer and Research institute, Ahmedabad, Gujarat, India.  
Email: rakeshmrawal@gmail.com

Received: 26 Aug 2015 Revised and Accepted: 27 Oct 2015

ABSTRACT

**Objective:** AF9-MLL has been implicated in the pathogenesis of AML, New Therapeutic regimens are prerequisite for this category of hematological malignancy due to the poor prognosis. The experimental 3D structure of AF9-MLL is not available. Therefore, present study aims in developing the homology model and evaluating the best model through Energy Minimization and MD simulation. The structure further analyzed for functional Annotation.

**Methods:** To the best of our knowledge, our study is novel in terms of predicting homology based 3D model of AF9-MLL leukemogenic fusion protein, facilitated by I-TASSER. The 3D modeled structure was subsequently optimized with MD simulation for 2 ns. Further stereo-chemical analysis and verification of the best structure so obtained were undertaken by different computational programs including PROCHECK, PROVE, Verify3D and ERRAT.

**Results:** Homology model predicted from I-TASSER and refined by YASARA showed results with 86.5% residues in the most favorable region, 14.7% in the allowed region, 0.8% in the generously allowed region and 0.3% in the disallowed region. The RMSD between the modeled and the refined structure was found to be 2.37 Å. The results of ERRAT, Verify\_3D, Prove and ProSA confirmed that the simulated model and energy minimized model is very good then the predicted raw model. The final structure was successfully submitted in Protein Model Database (PMDb) under ID: PM0080061.

**Conclusion:** In this study, homology model was developed and Validated for MLL-AF9 using bio-informatics tools. These analyses validated that the simulated model is best, robust as well as reliable enough to be used for future study and the functional analysis shows the presence of CXXC domain. Eventually, these molecular and structural studies result in advancement of newer therapies.

**Keywords:** MLL, Fusion Protein, Molecular modeling, Simulation, Structure Prediction.

INTRODUCTION

Chromosomal anomalies are regarded as one of the major hallmark of neoplastic cells, and the continual occurrence of chromosomal instability has been manifested in human neoplasia. Amongst these, recurrent reciprocal chromosomal translocations between non-homologous chromosomes are exclusively found to be implicated in the etiology of numerous hematological malignancies [1]. Balanced chromosomal rearrangements are crucial cellular mechanism, which lead to malignant transformation of normal cell via formation of chimeric fusion protein. 5–6% cases of Acute Myeloid Leukemia (AML) and 5–10% of acute lymphoblastic leukemia's (ALLs) cases are observed with the presence of chromosomal translocations involving the long arm (q23) of chromosome 11 [2]. Remarkably, the occurrence of 11q23 rearrangements is appreciably higher in pediatric AML and infant ALL. The Mixed-Lineage Leukemia (MLL) gene encodes the complex transcription factor that leads to the formation of unique hybrid genes, whose protein products are believed as critical elements in initiation of leukemogenesis. This multi exonic segment contains cluster of translocation breakpoints around exon 8 and various translocations partner genes combine with MLL gene yielding specific fusion protein responsible for development of a specific subtype of leukemia [3-8].

Till date, there have been more than 50 fusion gene partners reported for MLL. Amongst all MLL translocations, around 50% of infant AML cases comprises of t(9,11)(p22,q23) rearrangement. AF9 gene also known as LTG9 or MLLT3 is located at short arm p22 of chromosome 9 [9-11]. From several experimental studies, it was evident that leukemogenesis is caused by formation of MLL-AF9 fusion protein but still the mechanism of these partner genes is anonymous. In contrast, few other *in-vitro* and *in-vivo* analysis revealed that MLL-AF9 alters myeloid progenitor cells and

suppresses specific HOX gene e. g. mice with knock-in AF9-MLL fusion gene demonstrated anomalous proliferation of hematopoietic cell and developed AML identical to patient with t (9; 11) translocation [12-14]. Also, MLL and AF9 wild protein forms participate indispensably during hematopoiesis/embryogenesis process and are elements of protein complexes resulting in target gene transcriptional initiation (MLL) and elongation (AF9). Therefore it is hypothesized that MLL-AF9 fusion combines these characteristics, resulting in increased activation of target genes which may be interrupt hematopoietic cell differentiation and ultimately leads to leukemogenesis [15-19]. As the occurrence of 11q23 translocations is associated with extremely poor prognosis, novel therapeutic strategies are needed to be explored for this category of hematological malignancy. In spite of tremendous interest concerned with designing of target specific drug like molecules against this fusion protein. However, blocked by the unavailability of pertinent structural data. Additionally, structural & functional analysis of this chimeric gene (AF9-MLL) is required to be profoundly studied to get better insight into the causal mechanism leading to leukemogenesis. To resolve these problems, development of three dimensional molecular structure of AF9-MLL fusion protein is of prime importance with aim to discover newer alternative drug like compounds that precisely targets MLL-AF9 positive AML.

The fig. 1 shows the reciprocal chromosomal translocation between chromosome 9 and 11. Due to these translocation two genes fused and codes for oncogenic Fusion protein. To the best of our knowledge, our study is novel in terms of predicting the homology based 3D model of AF9-MLL leukemogenic fusion protein, undertaken by I-TASSER. The 3D model structure was subsequently optimized with MD simulation and further stereo-chemical validation studies and functional analysis of the best structure so obtained were executed by means of different computational

programs including PROCHECK, PROVE, Verify3D, ERRAT, NCBI-CDD, Pro Know and Inter Pro Scan.

## MATERIALS AND METHODS

### Sequence retrieval

The amino acid sequence of AF9-MLL fusion protein was retrieved from Uni Prot Database (<http://www.uniprot.org/>) submitted under the name of HUMAN putative AF9-MLL fusion protein with sequence ID of Q6TU33 and Entry name of Q6TU33\_HUMAN having total sequence length 107 amino acid residues [20]. This sequence so retrieved in FASTA format was utilized for further structural characterization and functional analysis.

### Protein structure prediction

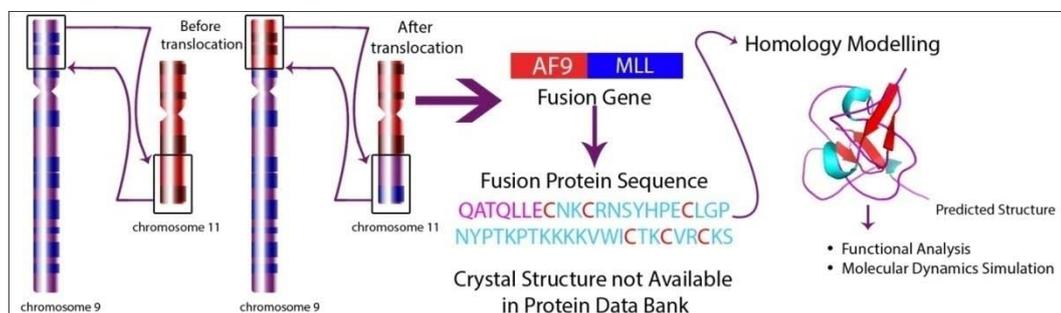
Full length AF9-MLL fusion protein sequence was uploaded to the I-TASSER (Iterative Threading Assembly Refinement) server (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>) for three dimensional structure predictions with default parameters. I-TASSER utilizes total four step protocol to unite alignment based model of existing protein structure with ab-initio model of unaligned regions in query protein to eventually provide alternatives of best scoring protein models [21]. The protein model was built from multiple sequence alignment of the query protein sequence with the template sequence with known structure and function [22]. The modeled structures were chosen on the base of sequence similarity with the Protein Data Bank

(PDB) templates and the energy minimization step was performed using YAMBER force field of YASARA plugin server [23].

### Refinement of modeled structure

The preliminary 3D model of AF9-MLL fusion protein acquired from homology modeling was further refined by Molecular Dynamic (MD) simulation in order to improve the accuracy of the structure. Here, MD simulation was accomplished by YASARA plugin which utilizes the molecular dynamics macro called md\_refine for enhancement of built model which consequently lessens the steric hindrances amongst the residues and thereby contributes towards overall stabilization of protein backbone. During simulation, model was solvated with water molecules and Conjugate Gradient protocol subsequent to steepest descent algorithm was undertaken in order to perform initial energy minimization steps.

Ultimately, the global minimization of model was attained by Simulated Annealing for eradication of redundant contact area among protein atoms & water molecules. Briefly, the predicted structure was then simulated inside trajectory box filled with 0.9% of NaCl ions (physiological condition) and water molecules by YASARA2 force-field using default parameter of macro and the NVT canonical ensemble. The pH was 7.4, temperature was 298°K and the density was 0.997 throughout the refinement. The data of protein model produced using YASARA got after simulation was investigated for trajectory projection [24].



**Fig. 1: Reciprocal translocation between chromosome 9 and 11 leads to AF9-MLL fusion gene codes for Novel Fusion Protein. Structure prediction of fusion protein through Homology Modeling**

### Validation of modeled structure

The predicted Models were further considered for accurate validation and verification by PROCHECK server [25] (<http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>) for stereo-chemical analysis of dihedral angles in modeled protein structure. PROCHECK analyzes overall residue by residue/structural geometry as determine by Ramachandran plot. VERIFY 3D ([http://services.mbi.ucla.edu/Verify\\_3D/](http://services.mbi.ucla.edu/Verify_3D/)) [26] decides similarity of model with its own amino acid sequence (1-D) by allocating structural class based on its location and environment, thus comparing results of superior structures. ERRAT (<http://nihserver.mbi.ucla.edu/ERRAT/>) [27] is a protein structure verification algorithm for assessing progression of crystallographic model building and refinement. The program scrutinizes the statistics of non-bonded interaction between different atom types which is useful to check structural reliability. PROVE (Protein Volume Evaluation) calculates the volume of atoms in macromolecules [28].

### Functional analysis of predicted structure

The domain analysis was performed to obtain accurate function of predict protein. The function domain of protein was predicted by means of different publicly accessible protein family databases. NCBI Conserved Domains Database (NCBI-CDD) which is a collection of sequence alignments and profiles representing protein domains conserved in molecular evolution. It also includes alignments of the domains to known 3-dimensional protein structures in the MMDB database [29]. Pro Know server and Inter-Pro Scan also predict the

function of proteins from the given structure, are also utilized for functional annotation of this fusion protein [30, 31].

## RESULTS AND DISCUSSION

The eventual objective of computational protein modeling is to anticipate a protein structure from its amino acid sequence with a precision that is analogous to the finest outcome accomplished by various sophisticated experimental techniques. This would permit use of *in silico* predicted structures in all the perspective where currently just experimental structures offer a concrete base for protein function annotation, structure based drug designing, interactions analysis and antigenicity, and rational designing of proteins with improved steadiness or unique functionality. Moreover, protein modeling is the only approach to gain structural information in case of failure faced during experimental methodologies and sometimes few proteins are too hefty for X-ray diffraction and NMR analysis. Amongst the three main methods of 3D structure prediction, homology modeling is comparatively reliable and easier than other approaches [32-34]. The current study centered on structural and functional analysis of AF9-MLL oncogenic fusion protein.

### Three dimensional model building

Various online tools/server are accessible for homology modeling of proteins and previous studies established that a sequence similarity higher than 25% among two proteins is significative of analogous 3D-structures [35]. Here, in order to execute homology modeling, the amino acid sequence in the query was subjected as input to I-TASSER

server. This server employs the method where the sequence of target is threaded by utilizing an emblematic PDB structure library to explore for the probable folds through numerous prominent alignment algorithms including Needleman-Wunch & Smith-Waterman, PSI-BLAST, Hidden Markov Model (HMM) and Profile-Profile Alignment (PPA). The server robotically carries out BLASTP for every protein sequence to recognize best possible templates for homology modeling and in total ten best alignments was incurred through implementation of versatile threading programs (Neff-PPAS, MUSTER, SPARKS-X, FAS03, SP3, PROSPECT2 etc) (table 1). For every recognized template, the template's lineament has been anticipated from characteristics of

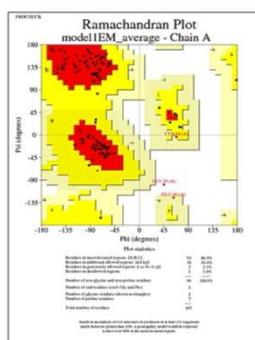
target-template alignment. After extensive sequence & structure alignments, the templates with the utmost value have then been preferred for constructing the molecular model [21]. In this case, PDB ID 2YSM had the most excellent Z-score of 4.96 obtained from all the algorithms and was chosen as the template for homology modeling which is the solution structure of the first & second PHD domain from Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog. I-TASSER anticipated 5 models in total, from which the model with best Conf-Score of 0.32 was selected with estimated accuracy of 0.76 (TM-Score) and 3.5 Å (RMSD). The 3D protein structure so modeled was visualized by Pymol.

**Table 1: Top identified structural analogs in PDB used by I-Tasser to model the protein**

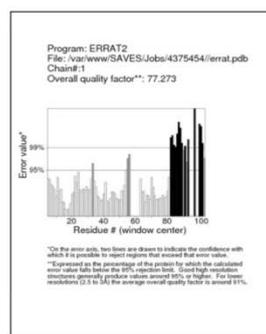
Rank	PDB Hit	TM-score	RMSD A <sup>o</sup>	Identity	coverage
1	2ysmA	0.843	1.64	0.287	0.944
2	2kwjA	0.719	2.67	0.267	0.944
3	2ln0A	0.705	2.76	0.238	0.944
4	4b9yA	0.504	4.09	0.067	0.878
5	2x2hA	0.491	4.21	0.049	0.888
6	2e6sA	0.485	2.98	0.264	0.635
7	1llqB	0.474	4.66	0.049	0.906
8	2aw5B	0.469	4.30	0.021	0.860
9	2e6rA	0.467	2.04	0.333	0.551
10	2k17A	0.462	3.01	0.185	0.589

The modeled protein structure was undertaken for energy minimization by utilizing YASARA plugin [24]. Energy minimization is fundamentally in relation to "reconcile" the model into a reasonably energetically favorable condition. Protein structures (either NMR, modeled, crystallography or molecularly docked) frequently have fault of varied level and energy minimization seem to provide the most diminution in system's energy on the whole by attenuating, non-bonded

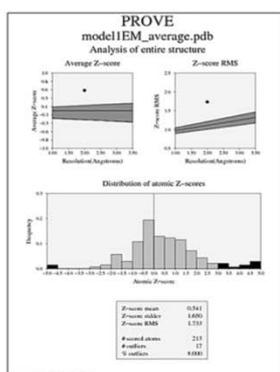
interactions, bond angles, lengths etc. into favorable condition to a greater extent. Energy minimization was executed by AMBER force field implemented in YASARA server to obtain optimized model structure with 6989.3 kJ/mol of initial energy to a final energy of -2366.6 kJ/mol. The energy minimized model of AF9-MLL fusion protein was considered for structural validation studies by various online tools and softwares like PROCHECK, VERIFY 3D, ERRAT and PROVE.



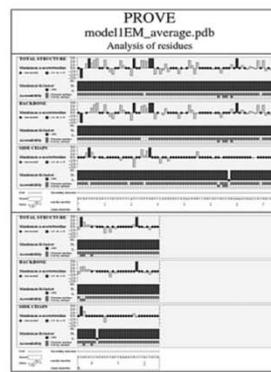
a



b



c



d

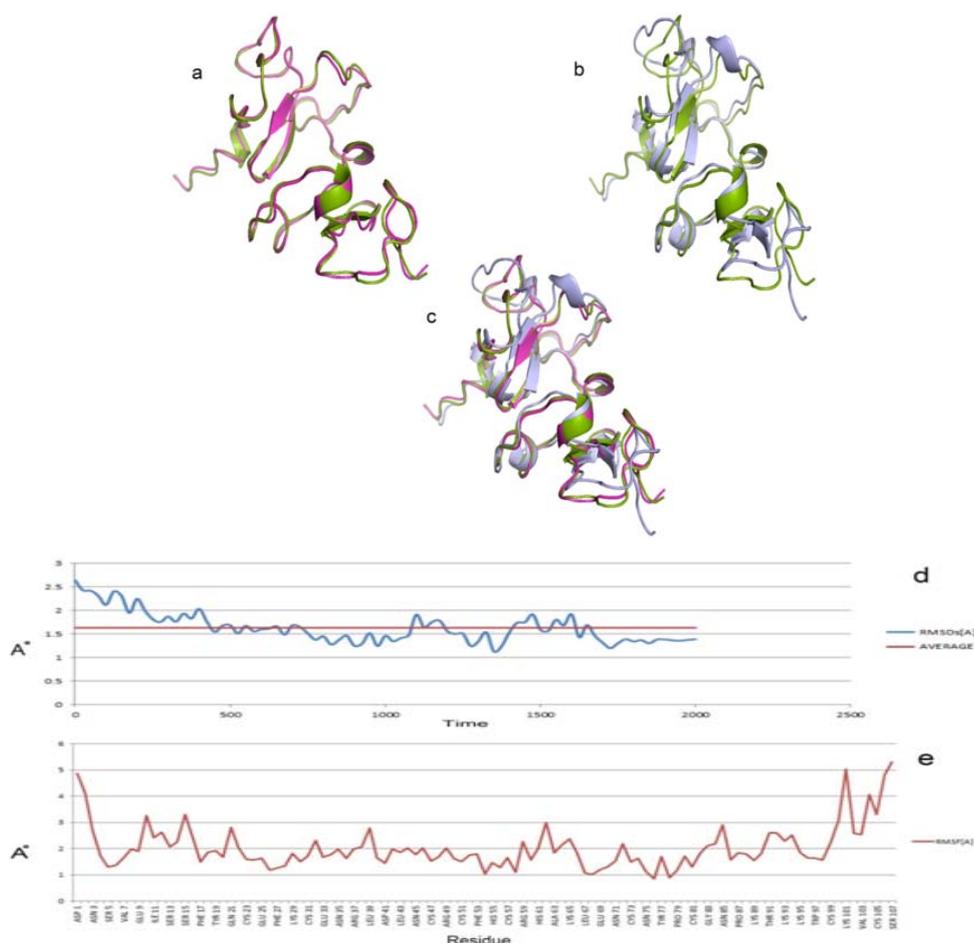
**Fig. 2: (a) Ramachandran plot values showing number of residues in favored, allowed and outlier region. (b) Errat plot where Black bars show the misfolded region, gray bars demonstrate the error region between 95% and 99%, and white bars indicate the region having less error rate for protein folding. (c) PROVE shows Z-score (c) PROVE Analysis of residues**

### Model refinement

Homology models are erroneous as structure emerges by a course of amino acid insertions, substitution and deletions [36-38]. Imprecision in model comprises of deformation in secondary structure elements, side chain packaging error and inadequately delineated loop conformations which necessitates that all predicted structures are mandatory for further refinement. Model enhancement is basically two step procedure where first, the local structural error are identified & eliminated through energy minimization and second, global (backbone) structural amendment for improving overall folds through MD simulation which is proficient sampling system to exactly recognize nearest native conformation [39-43]. Herein, extent of refinement was measured in terms of root mean square deviation (RMSD), by deviation of the resultant best fit structure against the initial structure in the course of simulation as a function of time. RMSD was calculated for the backbone and residues to verify the constancy of the trajectories. Moreover, the root mean square fluctuation (RMSF) was evaluated for each amino acid to analyze the flexibility of the trajectories. The predicted model fusion protein attained state of stabilization after

1.7 ns and average RMSD all atoms and backbone congregated to 1.63 Å (fig. 3e). The RMSF of individual residues is shown (fig. 3f) where the residues Asn 10 & 85, Ser 15, Gln 21 & 62 and Lys 101-Ser 107 demonstrated elevated peaks suggesting higher fluctuation of those amino acids. Amongst these, Gln 21 and end chain fusion protein residues showing higher flexibility possess cysteine residue in their neighbourhood indicative of certainty that there is maximum destabilization found around CXXC Domain of AF9-MLL fusion protein which hence can be directly correlated with conformational activation demonstrated by in-vitro studies [44, 45].

Moreover, Leu 39 found in disallowed region of the Ramachandran plot was also contributing towards overall destabilization of protein as seen from trajectory. Optimized model demonstrated presence of 86.5% of residues in favored zone which is higher than (46.9%) that of the raw modeled structure, signifying better steadiness of the refined structure. On the contrary, lower quality factor/scores were obtained from ERRAT & Verify3D for refined model as compared to higher values that of raw model, which is may be because of the exception that this protein is the fusion product and not a solitary protein.



**Fig. 3: Superimposed view of all three structures: (a) Predicted (Pink color) with energy minimized structure (Green color). (b) Refined Structure after simulation (Grey color) with Energy Minimized (Green color). (c) All Three structure final view. (d) RMSD Trajectory Graph. (e) RMSF trajectory of predicted protein structure for 2 nano second**

### Model validation, Quality assessment, and visualization

Each homology model integrates errors and the error counts for a given system primarily rely upon two standards. First, the proportion sequence similarity among the target & template and second is the total counts of erroneous in template [46]. Consequently, authentication of the model is an indispensable step in the procedure of homology modeling. Validation studies were performed for model which includes analysis of geometric

properties of backbone conformations by utilizing numerous structure evaluation tools and the results displayed in tabular form (table 2) verifies the superior model quality.

The PROCHECK examination on basis of Ramachandran plot endows with an interpretation about the stereo-chemical characteristic of the protein model. It focuses on protein regions that seemed to possess atypical geometry and allows for structural assessment on the whole [25].

Table 2: Comparative values of Procheck, Errat, Verify\_3D, Prove in different stages of refinement used in I-TASSER software

Validation	Predicted model	Model energy minimized	Model_refined
<b>Procheck</b>	<b>Regions of ramachandran plot</b>		
	<i>Favoured</i>	46.9%	67.2%
	<i>Additionally allowed</i>	43.8%	27.1%
	<i>Generously allowed</i>	6.2%	2.1%
	<i>Disallowed</i>	3.1%	3.1%
<b>ERRAT</b>		86.869	95.918
<b>VERIFY_3D</b>		82.24	90.65
<b>PROVE Z score</b>	Error	0.767	0.541

The Ramachandran plot in fig. 2a designated the area of probable angle formations by psi and phi angles. The traditional term correspond to the torsional angles on both side of  $\alpha$ -carbon in peptides. Thus, statistical investigation through PROCHECK provides the legitimate statistical factor that 67.2% of protein residues appeared in favorable region, 29.2% of residues falls in an allowed region and 3.1% of residues i.e. only one residue (Leu39) is there in disallowed region, speculating some steric obstruction as a

consequence of poor templates. For a superior models, the amino acids positioned in the favored and allowed regions is supposed to be greater than 90% which is holding true for the model existing here (that is, 67.2%+29.2% = 96.4%).

This is suggestive of fact that the model so constructed is of the superior kind. The RMSD value of the predicted structure with energy minimized and refined structure is shown in table 3.

Table 3: RMSD between modeled protein, Energy minized and simulated model of AF9-MLL fusion protein

	Predicted model and energy minimized	Energy minimized and simulated model	Predicted model and simulated model
RMSD[Å]	0.56	2.29	2.37

Consistency of the generated model was further computed by ERRAT which is a sophisticated methodology that calculates statistical organization of the particular kind of atom with respect to each other and is an exclusive approach for spotting erroneously folded regions in preliminary models. ERRAT hence gives *overall quality factor* for non-bonded interactions and the resultant higher score (with least accepted range of 50) is directly proportional to good model quality [27]. For the current model, overall estimated quality factor of ERRAT was 95.92 which is evocative of the fact that the structure is of good quality having high resolution with insignificant error standards of individual amino acid residues in modeled fusion protein (fig. 2b). The Verify 3D technique determines protein structures by means of three-dimensional visibilities. This tool evaluates the compatibility of 3D molecular model with its own (1D) amino acid sequence where the score ranges from -1 (not acceptable) to +1 (acceptable) [26]. As designates by the Verify3D server, the outcome demonstrated that 90.65% of residues possessed mean 3D to 1D score  $\geq 0.2$  which is symptomatically signifying that these structures were well-matched and reasonably of high quality. Another model authentication tool, PROVE evaluates statistical Z-score deviation for the modeled protein by determining the volumes of atoms in macromolecules utilizing an algorithm which considers the atoms as solid spheres [28]. PROVE analysis demonstrated the average statistical Z-score value of 0.76 (fig. 2 c & d).

From the entire results of structural validation program, it is deduced that the homology modeled protein is trustworthy for conducting further computational analysis on the oncogenic fusion protein including docking algorithms, molecular dynamic simulation in order to investigate protein-ligand interaction studies, moreover it aids in the recognition of potent ligands for particular therapeutic indications.

#### Submission of the protein structure in protein model database (PMDB)

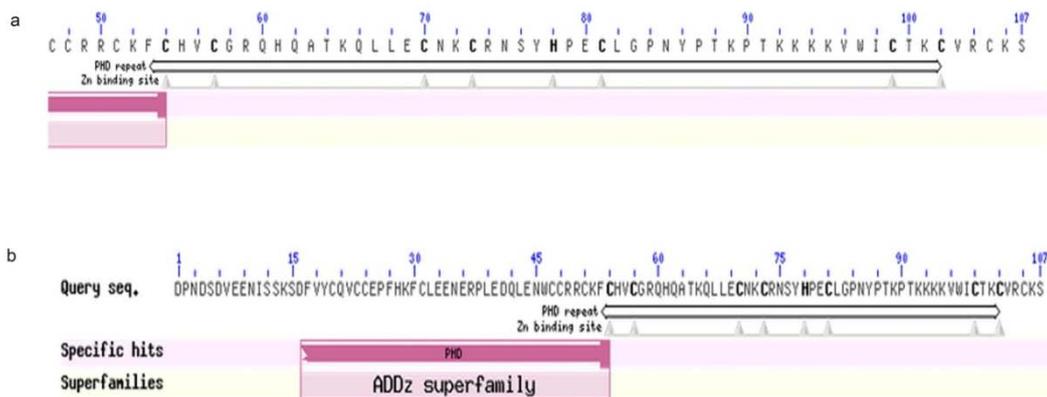
The final authenticated modeled structure of AF9-MLL oncogenic fusion protein was successfully submitted in Protein Model Database (PMDB) after effectively surpassing PMDB stereo-chemical quality tests and it is accessible under PMDB ID: PM0080061. This *insillico* constructed proteins structures database is open for public use from where users can freely obtain the model by its accession number and these structures may be further utilized for experimental characterization of the protein.

#### Functional annotation

The predicted protein was analyzed for further functional annotation. Three web tools were used to search the conserved domains and potential function of AF9-MLL Fusion Protein. Based on consensus predictions made by NCBI-CDD, Pro Know and Inter Pro Scan it is confirmed that AF9-MLL belongs ADDZ superfamily and possesses PHD like Zinc finger Domain. NCBI-CDD recognizes (cl17040) ADDZ superfamily, e-value = 9.06e-04 with PHD Domain within residue range from 16-54 (fig. 4b) with Pfam database Accession Pfam 00628. NCBI-CDD further predicted that residues range from 54-102 the 42 amino acid have PHD repeating Zn binding sites which show Conserved feature residue pattern: C CCC H C C [HC] on [47] residue number 54,57,70,72,78,81,99 and 102 (fig. 4a). The Inter Pro Scan could recognize that PHD-Type Zinc finger domain is present on the predicted model which is further confirmed by Pro Know Meta Server.

The canonical PHD finger is identified as Tri-Thorax Consensus (TTC) domain or leukemia-associated protein (LAP) motif which is distinguished as Cys4HisCys3 and seemed to be present in wide range of proteins concerned with transcriptional regulation and chromatin dynamics. Especially, putting an emphasis on their spectacular regulatory potentials, these molecules can identify and interact with a huge repertory of proteins specifically including the modified/unmodified histone tail (H3) and non-histone proteins. In particular, they are specific molecular scaffolds that serve as reader of epigenome governing the genetic expression via molecular mobilization of numerous transcriptional molecules and chromatin regulatory factors constituting multi-protein complexes. The CXXC domain found in numerous chromatin-associated proteins is epitomized by two CGXCXC repeats and it interacts with non-methylated CpG di-nucleotides.

Moreover, this domain encompasses eight conserved cysteine residues that bind to two zinc ions and its DNA binding interface has been recognized by NMR analysis. The RecQ helicase enzyme though possess single repeat that binds to zinc, is exception to be incorporated in family of this domain [48]. Results from PANTHER Family confirm that the GO molecular function shows activity of Methyltransferase and DNA binding activity of AF9-MLL protein. The function of this protein is to interact selectively and non-covalently with Zinc (Zn) ions based on a KEGG search AF9-MLL fusion protein was not found to be essentially involved in any of the bio metabolic pathways till the date.



**Fig. 4: Result obtained from NCBI conserved domain database. (a) Conserved feature residue pattern C C C H H C C [H C]: The Zn binding site. (b) PHD-type Zinc finger Domain of ADDz protein Family on predicted structure**

## CONCLUSION

This fusion protein was found to be critically associated with underlying leukemogenesis pathway and consequently would be a potential drug target in treatment of pediatric AML harboring t(9;11)(p22;q23). Lack of structural information till the date about this fusion protein obstructs the comprehensive characterization of its biological functions and its relevance in structure based design. For this reason, it was indispensable to develop the model of AF9-MLL fusion protein. The current study was conducted to construct the first 3 dimensional structure & to suggest potential functions of the AF9-MLL fusion protein. The model was created by homology modeling technique as well as optimized by MD simulation is near in-vitro environment. After refinement of the model, validation of structure was carried out with appropriate online tools. Outcome of these verification tools and low RMSD score signifies that the ultimate protein product was reasonably of superior quality. This modeled structure can be accessed at protein model database PMDB (ID: PM0080061). Additionally, better-quality modeled structure of AF9-MLL oncogenic protein may be further utilized in molecular docking & simulation studies and drug discovery.

## ACKNOWLEDGMENT

Amongst all the authors, M. D would like to acknowledge Prof. S. P Bhatnagar, Head, Department of Physics M. K Bhavnagar University for providing basic computational facilities and R. R to Department of Botany, Gujarat University for providing Facility of YASARA plugin.

## CONFLICTS OF INTERESTS

All authors have none to declare

## REFERENCES

1. TH Rabbitts. Chromosomal translocations in human cancer. *Nature* 1994;372:143-9.
2. Bernard O, Berger R. Molecular basis of 11q23 rearrangements in hematopoietic malignant proliferations. *Genes Chromosomes Cancer* 1995;13:75-85.
3. Djabali M, Selleri L, Parry P, Bowe M, Young BD, Evans GA. A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nat Genet* 1992;2:113-8.
4. Gu Y, Nakamura T, Alder H, Prasad R, Canaani O, Cimino G. The t(4,11) chromosome translocation of human acute leukemias fuses the *ALL-1* gene related to *Drosophila trithorax* to the *AF-4* gene. *Cell* 1992;71:701-8.
5. Tkachuk DC, Kohler S, Cleary ML. Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell* 1992;1:691-700.
6. Domer PH, Fakharzadeh SS, Chen CS, Jockel J, Jo hansen L, Silverman GA, Kersey JH, et al. Acute mixed-lineage t(4,11)(q21,q23) generates an MLL-AF4 fusion product. *Proc Natl Acad Sci USA* 1993;90:884-8.
7. Corral J, Forster A, Thompson S, Lampert F, Kaneko Y. Acute leukemias of different lineages have similar *MLL1* gene fusions encoding related chimeric proteins resulting from chromosomal translocations. *Proc Natl Acad Sci USA* 1993;90:8538-42.
8. Lo-Coco F, Mandelli F, Breccia M, Annino L, Guglielmi C, Petti MC, et al. Southern blot analysis of *ALL-1* rearrangements at chromosome 11q23 in acute leukemia. *Cancer Res* 1993;53:3800-3.
9. Iida S, Seto M, Yamamoto K, Komatsu H. *MLLT3* gene on 9p22 in t(9,11) leukemia encodes a serine/proline rich protein homologous to *MLLT1* on 19p13. *Oncogene* 1993;8:3085-92.
10. Nakamura T, Alder H, Gu Y, Prasad R, Canaani O, Kamada N, et al. Genes on chromosome 4,9 and 19 involved in 11q23 abnormalities in acute leukemia share homology and/or common motifs. *Proc Natl Acad Sci USA* 1993;90:4631-35.
11. Yamamoto K, Seto M, Iida S, Komatsu H, Kamada N, Kojima S, et al. A reverse transcriptase-polymerase chain reaction detects heterogeneous chimeric mRNAs in leukemias with 11q23 abnormalities. *Blood* 1994;83:2912-21.
12. Corral J, Lavenir I, Impey H, Warren AJ, Forster A, Larson TA, et al. An MLL-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. *Cell* 1996;85:853-61.
13. Joh T, Hosokawa Y, Suzuki R, Takahashi T, Seto M. Establishment of an inducible expression system of chimeric MLL-LTG9 protein and inhibition of Hoxa7/Hoxb7 and Hoxc9 expression by MLL-LTG9 in 32Dcl3 cells. *Oncogene* 1999;8:1125-30.
14. Dobson CL, Warren AJ, Pannell R, Forster A, Lavenir I, Corral J, et al. The MLL-AF9 gene fusion in mice controls myeloproliferation and specifies acute myeloid leukaemogenesis. *EMBO J* 1999;18:3564-74.
15. Pina C, May G, Soneji S, Hong D, Enver T. *MLLT3* regulates early human erythroid and megakaryocytic cell fate. *Cell Stem Cell* 2008;2:264-73.
16. Gan T, Jude CD, Zaffuto K, Ernst P. Developmentally induced Mll1 loss reveals defects in postnatal haematopoiesis. *Leukemia* 2010;24:1732-41.
17. Collins EC, Appert A, Ariza-McNaughton L, Pannell R, Yamada Y, Rabbitts TH. Mouse Af9 is a controller of embryo patterning like Mll whose human homologue fuses with Af9 after chromosomal translocation in leukemia. *Mol Cell Biol* 2002;22:7313-24.
18. Slany RK. The molecular biology of mixed lineage leukemia. *Haematologica* 2009;4:984-93.
19. Mohan M, Lin C, Guest E, Shilatifard A. Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nat Rev Cancer* 2010;10:721-8.
20. Whitmarsh RJ, Saginario C, Zhuo Y, Hilgenfeld E, Rappaport EF, Megonigal MD, et al. Reciprocal DNA topoisomerase II cleavage events at 5'-TATTA-3' sequences in MLL and AF-9 create homologous single-stranded overhangs that anneal to form der

- and der genomic breakpoint junctions in treatment-related AML without further processing. *Oncogene* 2003;2:8448-59.
21. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinf* 2008;9:40.
  22. Zhao F, Peng J, DeBartolo J. A probabilistic and continuous model of protein conformational space for template-free modeling. *J Comput Biol* 2010;7:783-98.
  23. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, *et al.* Improving physical realism stereochemistry and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* 2009;7:114-22.
  24. Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* 2004;57:678-83.
  25. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;6:283-91.
  26. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles methods. *Enzymology* 1997;77:396-404.
  27. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 1993;9:1511-9.
  28. Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol* 1996;264:121-36.
  29. Madej T, Address KJ, Fong JH, Geer LY, Geer RC, Lanczycki CJ, *et al.* MMDB: 3D structures and macromolecular interactions. *Nucleic Acids Res* 2012;0:D461-4.
  30. Pal D, Eisenberg D. Inference of protein function from protein structure. *Structure* 2005;3:121-30.
  31. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, *et al.* Inter Pro Scan: protein domains identifier. *Nucleic Acids Res* 2005;3:W116-20.
  32. Barcellos GB, Pauli I, Caceres RA, Timmers LF, Dias R, de Azevedo WF. Molecular modeling as a tool for drug discovery. *Curr Drug Targets* 2008;9:1084-91.
  33. Takeda SM, Takaya D, Chiba C, Tanaka H, Umeyama H. Protein structure prediction in structure based drug design. *Curr Med Chem* 2004;1:551-8.
  34. Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. *Drug Discovery Today* 2009;4:676-83.
  35. Vyas VK, Ukawala RD, Ghate M, Chintha C. Homology modeling a fast tool for drug discovery. *Curr Perspectives Indian J Pharm Sci* 2012;4:1-17.
  36. Abagyan R, Batalov S, Cardozo T, Totrov M, Webber J, Zhou Y. Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search. *Proteins* 1997;1:29-37.
  37. Baker D, Sali A. Protein structure prediction and structural genomics. *Science* 2001;294:93-6.
  38. Bradley P, Misura KM, Baker D. Toward high-resolution de novo structure prediction for small proteins. *Science* 2005;309:1868-71.
  39. Joo K, Lee J, Lee S, Seo JH, Lee SJ. High accuracy template based modeling by global optimization proteins. *Proteins: Struct Funct Bioinf* 2007;69:83-9.
  40. Misura KMS, Chivian D, Rohl CA, Kim DE, Baker D. Physically realistic homology models built with ROSETTA can be more accurate than their templates. *Proc Natl Acad Sci India* 2006;103:5361-6.
  41. Wang Q, Canutescu AA, Dunbrack RL. SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. *Nat Protoc* 2008;3:1832-47.
  42. Krivov GG, Shapovalov MV, Dunbrack RL. Improved prediction of protein side-chain conformations with SCWRL4 proteins. *Proteins: Struct Funct Bioinf* 2009;77:778-81.
  43. Levitt M. Accurate modeling of protein conformation by automatic segment matching. *J Mol Biol* 1992;226:507-33.
  44. Tomasz C, Laurie E, Jolanta G, Stephen ML, Relja P, Monika O, *et al.* Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. *Nat Struct Mol Biol* 2010;17:62-8.
  45. Hiroshi O, Marie K, Akinori K, Hirotaka M, Takeshi K, Toshiya I, *et al.* MLL fusion proteins link transcriptional coactivators to previously active CpG-rich promoters. *Nucleic Acids Res* 2014;42:4241-56.
  46. Sánchez R, Pieper U, Melo F, Eswar N, Martí-Renom MA, Madhusudhan MS, *et al.* Protein structure modeling for structural genomics. *Nat Struct Biol* 2000;7:986-90.
  47. Qiu Y, Liu L, Zhao C, Han C, Li F, Zhang J, *et al.* Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription. *Genes Dev* 2012;26:1376-91.
  48. Cross SH, Meehan RR, Nan X, Bird A. A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat Genet* 1997;16:256-9.