Identification and In-silico analysis of alternations in DNA binding region of *CEBPA* gene and hematological pattern in *CEBPA* mutant and non-mutant Acute Myeloid Leukemia patients

Muhammad Muntazir Mehdi[1], Sikandar Hayat*[1], Zulqarnain[2], Mazia Shahid Butt[3], Maria Muddassir[1], Munaza Hafeez[4], Tahir Maqbool[1], Faheem Hadi*[2], Muhammad Khawar Abbas[2], Samiullah Malik[2], Afia Muhammad Akram[5], Malik Ihsan Ullah Khan[1], Humaira Sultana[6], Farah Abid[7], Naila Ashraf[7], Qurban Ali[8], Sundus Bukhari[1], Kashif Ali[1]

[1]Institute of Molecular Biology and Biotechnology, The University of Lahore, Pakistan.
[2]Faculty of Medicine and Allied Health Science, The Islamia University of Bahawalpur, Bahawalpur, Pakistan.
[3]General Internal Medicine, LocumLink, Ireland.
[4]National Institute of Food Science and Technology, University of Agriculture, Faisalabad, Pakistan.
[5]Department of Zoology, Division of Science and Technology, University of Education, Township Lahore, Pakistan.
[6]Bahawalpur college of Pharmacy,
[7]Khaldunia college of Pharmacy, Khaldunia Institute of Technology and Applied Sciences, Lahore Pakistan.
[8]Department of Plant Breeding and Genetics, Faculty of Agricultural Sciences, University of the Punjab, Lahore, Pakistan.

***Correspondence:**

Dr. Sikandar Hayat

Institute of Molecular Biology and Biotechnology, The University of Lahore, Pakistan.

Dr. Faheem Hadi

Faculty of Medicine and Allied Health Science, The Islamia University of Bahawalpur, Bahawalpur, Pakistan.

Abstract

The current study has been designed to determine the frequency of *CEBPA* gene mutations in cohort of 30 acute myeloid leukemia (AML) patients, and their role in the development of AML, and their effect on the structure and function of translated protein was also studied. Blood samples were collected from 30 patients from Jinnah and Mayo hospitals, and their hematological characteristics were noted. DNA was isolated from all blood samples, and after amplification, samples were sent for sequencing. Mutations were identified by comparing sequencing results with normal CEBPA gene nucleotide sequences. *CEBPA* gene mutations were observed in 7 (23.3%) 30 patients. These 7 patients had two missense mutations both in coding and non-coding regions. Non-coding region mutation was detected in 4 patients, while coding region mutation was observed in 3 patients. There was no significant difference in hemoglobin level, TLC, platelets count, and blast count percentage between *CEBPA* gene mutant and non-mutant AML patients. While high LDH was observed in *CEBPA* mutant patients. The deleterious effect of S348T mutation was checked using SNAP2, PROVEAN, PolyPhen-2, PhD-SNP, Align GVGD, and PANTHER. Protein's stability was studied by using I-Mutant 3.0. ConSurf 2.0 tools were used to check the conservation status of mutation, while NetsurfP predicted protein's secondary structure. The SWISS model generated a 3D model of both mutant and wild-type residues, which MolProbity, Protparam, ProSA web, and VADAR analysis further analyzed. The mutational effect on RNA's secondary structure was analyzed using the

Vienna package, while the interaction of *CEBPA* gene with other proteins was analyzed by STRING tool.

Keywords: myeloid leukemia, mutations,

Introduction

Acute myeloid leukemia (AML) is hematopoietic malignancy that is caused by different genetic alternations and epigenetic events occurring in progenitor cells that result in uncontrolled production of cells and cells fail to differentiate, and they lack apoptosis and clonal myeloblasts start to accumulate in bone marrow. Due to this, the number of normal hematopoietic progenitor cells has decreased.[1,2,3] The absence of mature hematopoietic cells like erythrocytes, lymphocytes, platelets, or granulocytes characterizes AML. If not treated, the infection becomes severe, bleeding and wasting start, and the results are deadly [1]. It is a heterogeneous disease that is the result of many types of abnormal cytogenetic and molecular changes like gene mutations, RNA splicing, epigenetic alternations, cohesion complexity, and chromosomal abnormalities that influence the transcriptional and epigenetic regulators and singling pathways. [4] Acute myeloid leukemia is found in patients of all ages but most frequent in adults with a median age of 68. [5] Even during infancy and adolescence, AML may also occur. [6]

The researcher has described various risk factors for leukemia. Still, exposure to benzene, ionizing radiations and alkylatic agents, dietary intake, smoking, obesity, cytotoxic chemotherapy, and agents that can damage DNA have been noted as main leukemia-associated risk factors. Therapy-related AML (t-AML) may also cause due to the previous treatment with ionizing radiations and cytotoxic chemotherapy.[7,8,9] Cytotoxins such as chlorambucil, melphalan, and topoisomerase-II inhibitors may involve in t-AML development[10]. Moreover, the link between smoking and AML has also been observed. Similarly, incidence of AML is greater in smokers than non-smokers in males and females. [11,12] The main cause of AML is the genetic mutations in cancerous cells. Mutations in *CEBPA* double alleles, *ASXL1*, *RUNX1, FLT3-ITD*, *KIT*, *NPM1,* and *TP53* are common AML patients. [13] In 50 % of AML cases, these genetic changes are associated with karyotypic mutations, which can easily identified with the help of cytogenetic analysis (such as translocations and deletions). [14] These karyotypic abnormalities are linked with AML treatments; different treatments show unfavorable outcomes and survival in these AML. [15] Cytogenetically, around 40 to 50 % of AML patients are identified as normal. [13,14] *CEBPA* gene is located at chromosome 19q13.1 [16] and translates a transcription factor *CEBPA* protein. Only one exon and no intron are preset in *CEBPA* gene. [17] *CEBPA* is part of bZIP transcription factors family. [18] *CEBPA* protein exists as p42 (42 KDa) or p30 (30 KDa) isoform.[19, 20] Wild-type protein (42 kDa) has two regulatory transactive domains (TAD1 and TAD2) in the N terminus region and basic leucine zipper (bZIP) domain in C-terminus for DNA binding and dimerization while shorter isoform has only TAD2 (Figure 1.1).[21, 22, 23] Many transcription factors contain bZIP domain, which is crucial in protein dimerization. DNA binding to the major groove of DNA molecule is enabled by bZIP domain. [24] *CEBPA* plays a vital role in differentiating and reproducing myeloid progenitor cells. Alternations in *CEBPA* are regarded as the most commonly occurring genetic alternation in AML patients. Alternation in *CEBPA* can take place in whole genes but they commonly occur in N- and C-terminal regions of *CEBPA*.[25] Alternation in *CEBPA* has been identified in 7-15 % of all AML cases and 33.33% of all *CEBPA*sm patients are associated with alternation in bZIP domain.[24] N-terminal associated mutations are found on one allele, while C-terminal linked mutations are found on another allele (monoalllelic *CEBPA* mutations).[23, 26, 27] Double allelic mutations have also been identified.[27] Mutations occurring in N-terminal shorten the *CEBPA*-producing protein by producing premature

stop codon that leaves negative effects such as uncontrolled growth and dysregulation in the cell cycle while in-frame deletions and insertions of C-terminal cause disturbance in the dimerization and binding of DNA to *CEBPA.* [23, 26, 27]

Methodology

Blood samples of 30 AML patients aged between 16-75 years were collected from Jinnah and Mayo hospital, Lahore, and their hematological characteristics were also noted. Collected blood samples were transported to the university laboratory and kept at -20°C in the refrigerator until further processing.

DNA was isolated from collected blood samples using Thermo Scientific Gene JET Genomic DNA isolation Kit (K0722) following the manufacturer protocol. Isolated DNA was quantified with 1% agarose gel, and then DNA binding region of *CEBPA* gene was amplified by PCR using primers given in Table 1. The amplified products were sequenced commercially by Sanger sequencing to identify any nucleotide alterations. After being sequenced, the obtained findings were examined using Geneious prime software (version 2022.0.1), and mutations were identified by comparing the sequence of the bZip domain of *CEBPA* gene with the reference sequences.

Table 1: Primers sequence for amplification

Target gene	**Forward primer**	**Reverse primer**	**Product size**
DNA binding region	CAAGGCCAAGAAGTCGGTGGACA	ATCTCGAGGCTTGCCCGAGCCGTG	356

In silico analysis

Harmful effects of deleterious SNPs

PolyPhen-2, Align GVGD, SNAP2, PROVEAN, PhD-SNP, and PANTHER software were used to check the harmful effects of mutations identified in the *CEBPA* gene. SNAP2 and Align GVGD software were used to check whether the mutations identified in *CEBPA* gene had a neutral or harmful effect. FASTA sequence of the protein was input for both these sofwares. SNAP2 displayed a score ranging from -100 (which predicts the mutation is neutral) to +100 (which indicates the mutation has a deleterious effect). Using score value, the effect of mutations on protein's function was predicted. [28] Align GVGD displyed results in a spectrum of different class mutations ((C0, C15, C25, C35, C45, C55, and C65). Class 15 predicts that the mutation is less deleterious, while class 65 expected mutation as the most harmful. [29]

The help of PolyPhen-2 webserver checked the effect of change in the amino acids sequence of a protein. As input, accession number or database ID, protein sequence, and information regarding amino acids change were used. PROVEAN webserver was used to predict the effect of change in amino acid on protein function. As an input query, FASTA sequence of the protein was used. This software displayed results in the form of score. The effects of change in amino acid sequence were predicted using score value. Less than a 2.5 score value predicted that the variation in amino acid sequence had a deleterious effect. Score values greater than 2.5 predicted variation as neutral.[28] The position of amino acid occurrence in evolutionary-related protein family was predicted using PANTHER cSNP. Evolutionary proteins alignment is used as a base, and on the base of it, PANTHER cSNP collected substitution position-specific evolutionary conservation (subPSEC) score and predicted the functional impact of cSNP. PANTHER cSNP predicted results as neutral or

deleterious. Information about amino acid alternation and sequence of protein was used as input.[30]

PhD-SNP webserver was used to check whether amino acids substitution was neutral or deleterious. This webserver was used as input for the sequence of CEBPA protein, position of amino acid alternation, and substituted amino acid. [31]

Analysis of protein stability

I-Mutant Suite predicted the effect of amino acid alternation on protein stability. I-Mutant predicted results using DDG values. [28]

ConSurf webserver for the prediction of conservation profile

The ConSurf was used to check the evolutionary conservation status of amino acids alternation. It is available (https://consurf.tau.ac.il/). FASTA sequence of protein was used as input. ConSurf displayed results in conservation scores (range 1-9), and scores were divided into conserved, variable, and average. [32]

NetsurfP-2.0 for predicting solvent accessibility

Secondary structure, surface accessibility, and amino acids phi/psi dihedral angles present in the protein sequence were predicted by NetsurfP-2.0. As an input query, FASTA format sequence of the protein was used. NetsurfP-2.0. showed the exposed and buried regions present in the structure of *CEBPA* protein. [33]

Analysis of amino acid alternation effect on protein structure

The effect of amino acid alternation on protein structure, chemical and physical characteristics, hydrophobicity, and function was analyzed using HOPE. UniProt ID and protein sequence were used as inquiry input. [32]

Protein modeling

The SWISS model web server created the protein 3D model for both mutant and wild residues. Model template was selected based on recovery value and similarity, and GMQE value, QMEAN Z-score, and sequence identity parameters were noted to assess the model. Global Model Quality Estimate of mutant and wild residues was also analyzed to create a protein model [34]. MolProbity webserver was used to check the validity of both wild and mutant residues created by SWISS model. PDB file obtained from SWISS model was used as input at MolProbity webserver. [35] Chemical and physical properties and changes in the structure of the mutant model were analyzed with the help of ProtParam webserver by considering some parameters i.e. aliphatic index, GRAVY, instability index, and theoretical pI value. The aliphatic index is the measurement of space taken by the aliphatic chains. Theoretical pI specifies is the pH where protein net charge is zero. The instability index gives information about the stability of protein and tells how much protein is stable. Instability index value <40 predicts that a protein is stable; a protein is considered unstable if the value is >40. A positive GRAVY value indicates that the residues are hydrophobic, while a negative value shows that residues are hydrophilic.[36] Protein secondary structure was analyzed by SOPMA.[28]

Analysis of mutational effect on the secondary structure of RNA

Mutational effect on RNA secondary structure was examined by RNAfold of Vienna package. Effect was analyzed based on RNA folding energies, i.e., Minimum Free Energy (MFE).[37]

Analysis of *CEBPA* protein association with other proteins

Link of *CEBPA* protein with other proteins was analyzed by using STRING webserver. Protein names were used as query input.[32]

Results

Isolated DNA

DNA was separated from all collected blood samples. DNA isolation was confirmed on 1% agarose gel at 120V for 45 minutes (Figure 4.5).

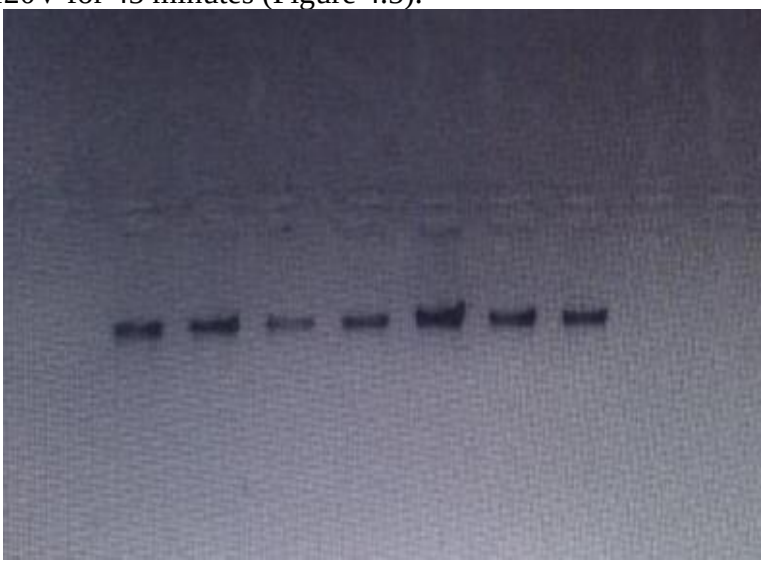

Figure 1: Isolated DNA from blood samples

Amplification of DNA binding region of *CEBPA* gene

Polymerase chain reaction (Figure 2) was used to amplify the DNA binding region of *CEBPA* gene. Annealing temperature for primers of DNA binding region of *CEBPA* gene was 56ºC.

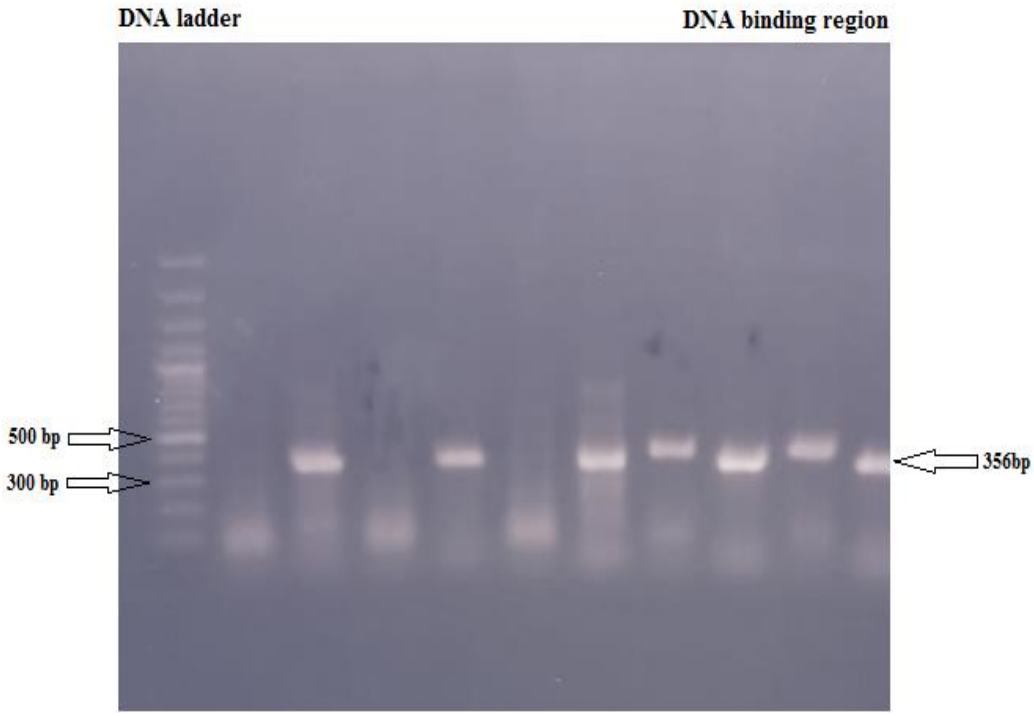

Figure 2: Visualization of amplified DNA binding region of *CEBPA* gene by 1% agarose gel electrophoresis

Nucleotide alterations found in DNA binding region of *CEBPA* gene

Alternations in nucleotide sequence of *CEBPA* gene and subsequent amino acid substitution were identified by comparing the sequencing results with NCBI reference

sequence. Geneious Prime software was used to investigate results. In this study, 2 heterozygous novel missense mutations S348T and C.+G87A (sample 14, 29, 34, 36, 51, 54, 73) were identified in 7 patients, out of which 1 mutation was present in coding region and 1 in non-coding region.

Three out of seven patients had coding region S348T, while four had non-coding region C.+G87A mutation. In S348T mutation, normal nucleotide base guanine at 1043 position was changed to cytosine, due to which normal base pair codon AGC, which encodes serine amino acid was altered to ACC as a result of which threonine amino acid substituted serine at position 348 in protein. While in C.+G87A, nucleotide base guanine was altered by adenine at position 87 (Figure 3 & Figure 4).

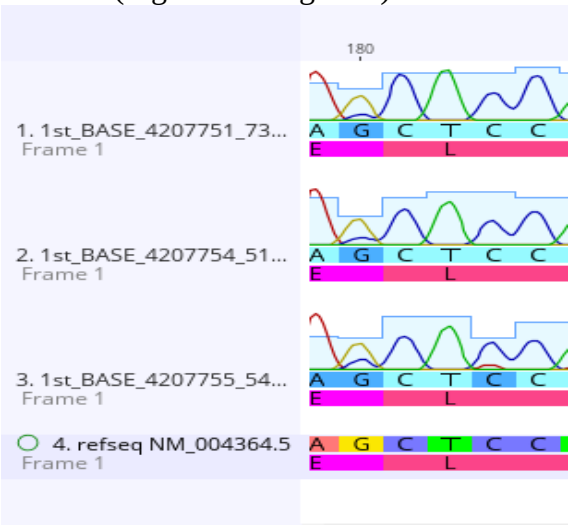

Figure 3: Heterozygous mutation G1043C resulting in an amino acid substitution S348T in AML patients identified through Geneious prime software (Version 2022.0.1)

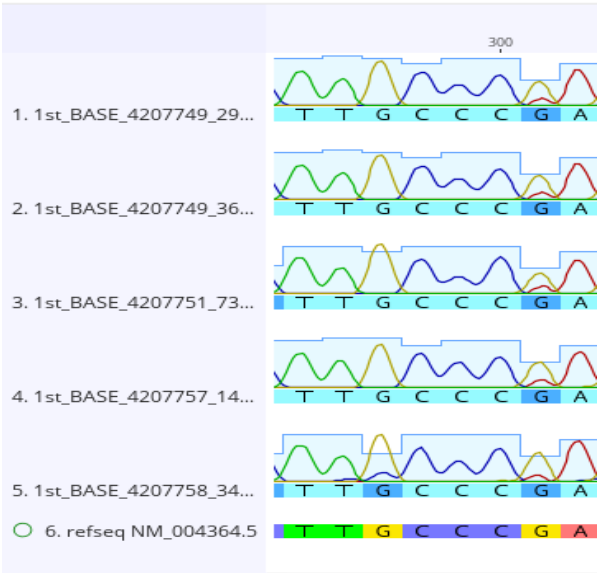

Figure 4: Mutation C.+G87A in non-coding region of CEBPA gene in AML patients identified through Geneious prime software (version 2022.0.1)

Comparison between the hematological characteristics of *CEBPA* mutant and non-mutant patients

Hematological characteristics of non-mutant and *CEBPA* mutant were compared by using one-way ANOVA. There was no significant difference between total leucocyte count

(TLC), platelets count, and haemoglobin level of *CEBPA* mutant and non-mutant AML patients (Table 2, Figure 5).

Table 2: Comparison of hematological characteristics of *CEBPA* mutant and non-mutant patients.

Parameters	*CEBPA* non-mutant patients	*CEBPA* mutant patients	P-value
No. of patients (%)	23 (77%)	7 (23%)	—
Mean age, years (range)	40 (20-73)	77 (52-79)	0.019*
Hemoglobin level (g/dl), mean (range)	8.2 (4.1-12.6)	8.8 (4.8-10.9)	0.990
Platelets ($\times 10^3$/µl), mean (range)	45 (5-720)	39 (2-941)	0.898
Total leucocyte count ($\times 10^3$/µl), mean (range)	31.9 (1.4 to 430)	27.3 (0.9 to 350)	0.165
Lactate Dehydrogenase (U/L), mean value (range)	456 (150-860)	650 (120-2587)	0.006*
Blast cells (%age)	65.1%	70.6%	0.157

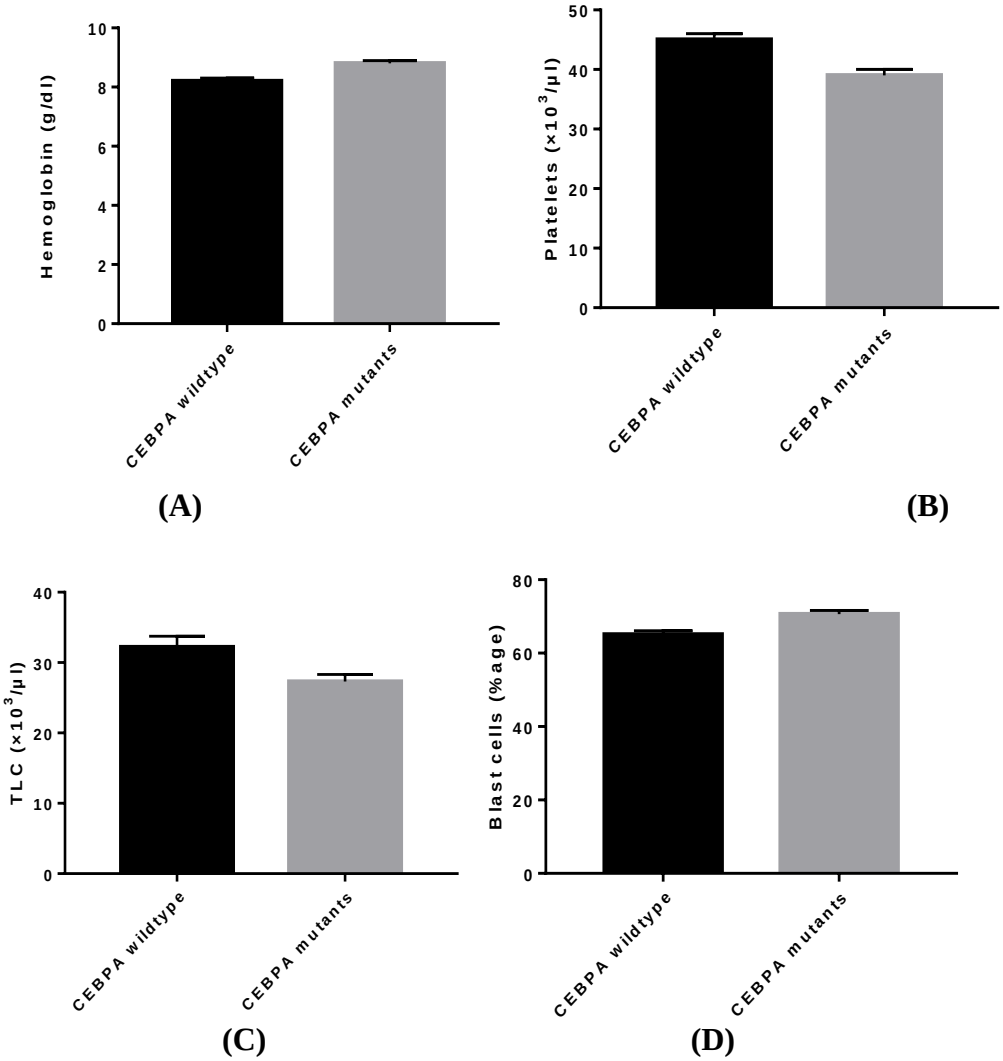

Figure 5: (A) Hemoglobin, (B) platelets count, (C) TLC, (D) Blast cells level graphical representation.

In silico analysis of S348T and C.+G87A mutations found in DNA binding region of *CEBPA* gene

Six in silico SNP prediction tools (PolyPhen-2, Align GVGD, PROVEAN PANTHER, SNAP2, and PhD-SNP) were used to analyze the mutational effects on protein structure and function. PolyPhen-2 predicted mutation as possibly damaging. SNAP2 indicated that the mutation was neutral and had no impact. Align GVGD results showed that the SNP was most likely affected. PROVEAN and PhD-SNP showed that the alternation of amino acid was neutral, while PANTHER predicted that the SNP was possibly damaging (Table 3). A decrease in protein stability was expected by I-mutant (Table 4).

Table 3: Effect of identified mutation on protein integrity

Mutations	**Polyphen-2**	**SNAP2**	**Align GVGD**	**PROVEAN**	**PhD-SNP**	**PANTHER**
S348T	Possibly Damaging	Neutral	C55	Neutral	Neutral	Possibly damaging

Table 4: Mutational effect on the stability of protein

Mutations	**I-Mutant**	
	DDG value (kcal/mol)	**Stability (SVM2)**
S348T	-0.40	Decrease

Conservation status of deleterious SNP in DNA binding region of *CEBPA* gene

ConSurf results indicated that the variant S348T was an exposed and average residue. The conservation score of S348T variant and its role are given in table 5 and figure 5.

Table 5: Analysis of mutation evolutionary conservation profile by ConSurf

Mutations	**Conservation score**	**Conservation scale status**	**Exposed/ Buried**	**Structural/ Functional Role**
S348T	7	Conserved	Exposed	functional

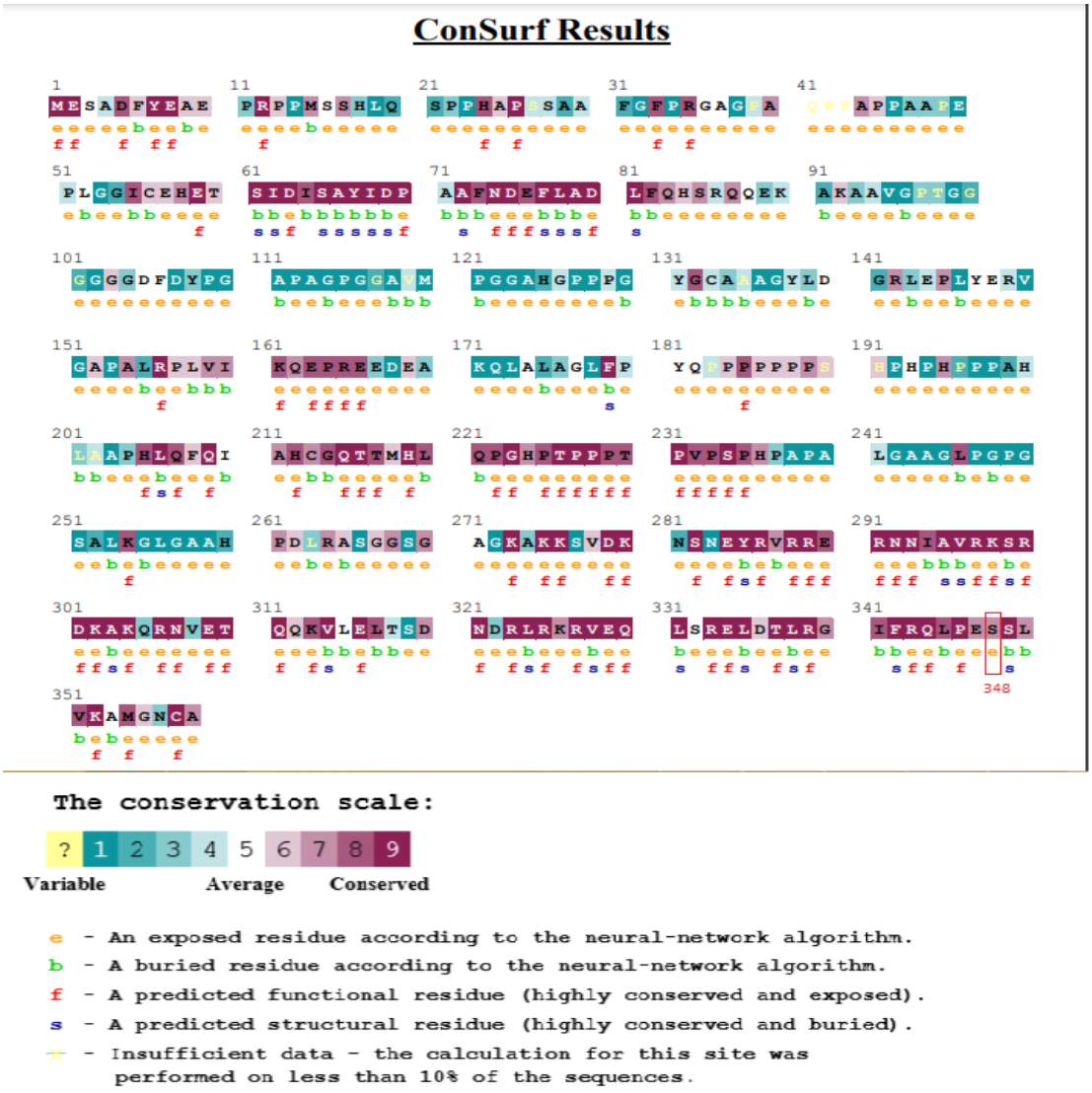

Figure 6: Conservation status of deleterious SNP

NetSurfP-2.0 results indicated that the variant was exposed (Figure 6). Absolute and relative surface accessibility of S348T variant is given in Table 6.

Table 6: NetSurfP-2.0 evolutionary conservation profile of mutation.

Mutations	Class alignment	Relative surface accessibility (RSA)	Absolute surface accessibility (ASA)
S348T	Exposed	0.512	59.976

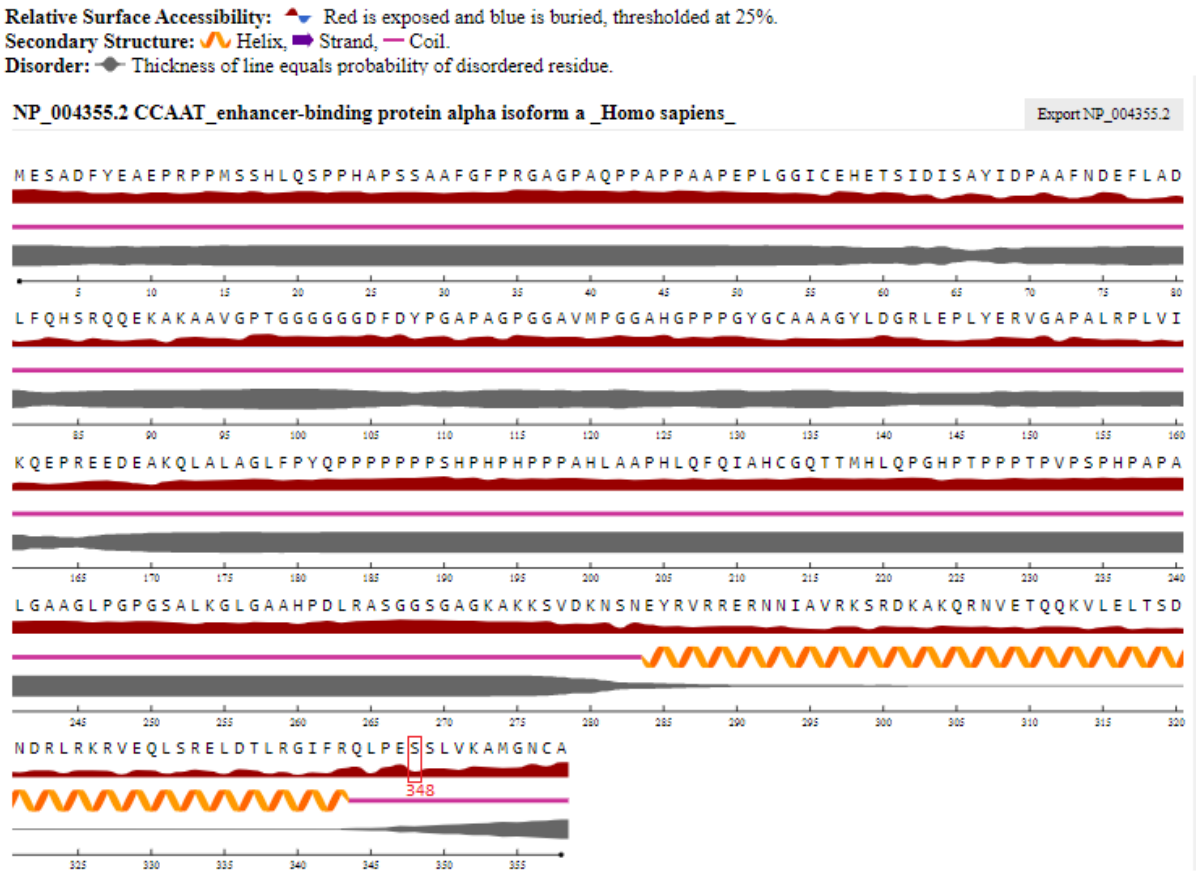

Figure 7: NetSurfP2-0 conservation profile

SWISS model

Swiss model predicted no difference in the protein model of both wild and mutant residues. Sequence identity, GMQE, and QMEAN Z-score of both wild and mutant are given in table 7.

Table 7: Values of sequence identity, GMQE, and QMEAN Z-score predicted by Swiss model for mutant and wild residues. Total number of amino acids was 358, while a range of amino acids in model was 281-340 with template query no. of 1nwq.1.C.

Mutation	Sequence Identity (%)	GMQE	QMEAN Z-Score
Wild type	100.00	0.10	1.11
S348T	100.00	0.10	1.11

Structural analysis by project HOPE

HOPE tool was used to check the effects of amino acid change on the physical and chemical characteristics and the function, hydrophobicity, and spatial structure of protein. The results predicted that mutant residue was bigger than wild type, and it resulted in producing pumps. The difference in size and hydrophobicity of mutant and wild-type residue disturbed protein structure by weakening H-bonding interactions with other residues (Table 8, Figure 7).

Table 8: Results predicted by HOPE

Mutation	Size	Change of Charge
S348T	W< M	---

Figure 8: Amino acid change in *CEBPA* protein due to nucleotide alterations shown by HOPE. (S348T)

ProtParam

From the result displayed by the ProtParam, we noted the instability index, theoretical pI, aliphatic index, molecular weight, and GRAVY value, as shown in table 9.

Table 9: Representing the difference between wild-type and mutant variations based on chemical and physical factors

Mutations	Molecular weight	Theoretical Pi	Instability Index	Aliphatic Index	GRAVY
Wild-type	37561.24	7.27	58.52	62.07	-0.633
S348T	37575.27	7.27	57.45	62.07	-0.633

ProSA-Web

Errors in the 3D model of mutant residue generated by the SWISS model were identified using ProSA web. The Z-score value was noted from the results displayed by ProSA web to determine protein's authenticity. The range of amino acids in 3D model of protein was 281-340. The Z-score value changed from -6.58 to -1.44, indicating that the mutant and wild type lay within the natural protein's typical range, allowing minor structural flaws (Table 10).

Table 10: Z-score values of *CEBPA* protein for wild and mutants calculated by ProSA Web. Total number of amino acids was 358, and range of amino acids in all models was 281-340.

Mutations	Z-Score	Chain
Wild-Type	---	---
S348T	-1.44	C

MolProbity

MolProbity tool was used for the qualitative structural analysis of protein. The results were displayed as a Ramachandran plot, which described the location of amino acids in the allowed and favored regions based on their torsional angels. According to Ramachandran plot, 100% of amino acids residues of wild and mutant (S348T) were found in the allowed region, while 99.1% of amino acid residues of wild and mutant (S348T) were found in the favored region, as shown in table 11 and figure 8.

Table 11: Percentage of favored and allowed regions amino acids shown by Ramachandran plot

Mutations	Allowed region amino acid (%)	Favored region amino acid (%)
Wild-Type	100.0% (116/116)	99.1% (115/116)
S348T	100.0% (116/116)	99.1% (115/116)

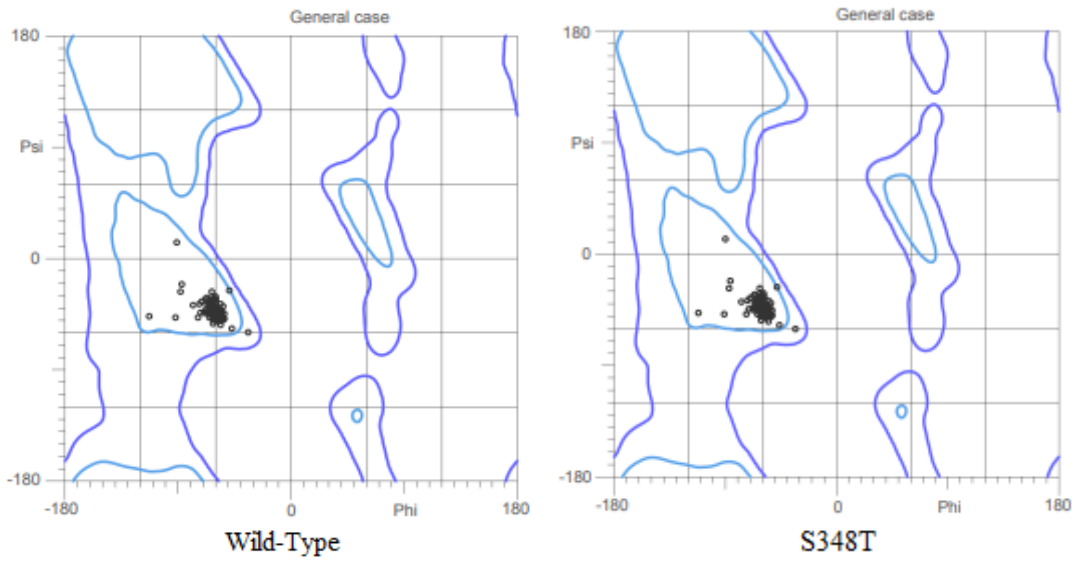

Figure 9: Ramachandran plots showing amino acids in the favored and allowed regions. (A) Wild-type (B) S348T

VADAR analysis

VADAR analysis was used for the evaluation of quantitative analysis of protein structure. Coils, helix, turns, and beta chain parameters were analyzed by VADAR analysis tool. No difference was observed in helix and coils of wild and mutant S348T and beta coils and turns of wild and mutant (Table 12).

Table 12: Results predicted by VADAR

Mutations	Helix	Beta	Coils	Turns
Wild-Type	112 (93%)	0 (0%)	8 (6%)	4 (3%)
S348T	112 (93%)	0 (0%)	8 (6%)	4 (3%)

SOPMA

According to SOPMA server statistics, the blue H (alpha helix) contributed around 30.17 percent, the extended strand (red E) contributed around 8.38 percent, beta-turn (green T) contributed around 6.42 percent, and yellow C (random coil) contributed around 55.03 percent of the total (Figure 9).

```
         10         20         30         40         50         60         70
          |          |          |          |          |          |          |
MESADFYEAEPRPPMSSHLQSPPHAPSSAAFGFPRGAGPAQPPAPPAAPEPLGGICEHETSIDISAYIDP
hcttceeccccccccccccccccccccccceecccccccccccccccchhhhhhtccccceeeecct
AAFNDEFLADLFQHSRQQEKAKAAVGPTGGGGGGDFDYPGAPAGPGGAVMPGGAHGPPPGYGCAAAGYLD
tcccchhhhhhhhhhccccchhhhhhhccttccccccccccccccccccccccccccccccccthhhhhhhh
GRLEPLYERVGAPALRPLVIKQEPREEDEAKQLALAGLFPYQPPPPPPPSHPHPHPPPAHLAAPHLQFQI
tcccchhhhhccccttccceeeeeccccchhhhhhhhhhhccccccccccccccccccccccccccchhhhhh
AHCGQTTMHLQPGHPTPPPTPVPSPHPAPALGAAGLPGPGSALKGLGAAHPDLRASGGSGAGKAKKSVDK
httccceeeecttcccccccccccccccccccccccccccchhhhcccccccchettcccccccccheecct
NSNEYRVRRERNNIAVRKSRDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSRELDTLRGIFRQLPESSL
tccheeeettttceeeeccccchhhccchhhhhhheehhccchhhhhhhhhhhhhhhhhhhhhhhhhccccchh
VKAMGNCA                                                            348
hhhhhhhh
```

```
Sequence length :    358
```

```
SOPMA :
    Alpha helix      (Hh) :    108 is   30.17%
    3₁₀ helix        (Gg) :      0 is    0.00%
    Pi helix         (Ii) :      0 is    0.00%
    Beta bridge      (Bb) :      0 is    0.00%
    Extended strand  (Ee) :     30 is    8.38%
    Beta turn        (Tt) :     23 is    6.42%
    Bend region      (Ss) :      0 is    0.00%
    Random coil      (Cc) :    197 is   55.03%
    Ambiguous states (?)  :      0 is    0.00%
    Other states          :      0 is    0.00%
```

Figure 10: Secondary structure of protein predicted by SOPMA

STRING database

STRING database revealed that *CEBPA* was functionally linked to 10 other proteins, which are *JUN*, *CDK2*, *CDK4*, *SPI1*, *FOXO1*, *PPARG*, *HDAC1*, *CDK2*, *EP300*, *E2F4*, and *FOS* as shown in figure 10. Structural alternation in *CEBPA* protein can disturb the functions of other associated proteins.

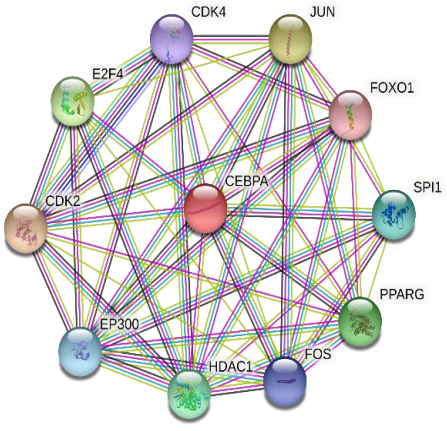

Figure 11: Interaction of *CEBPA* with other proteins predicted by STRING database

Effect of mutation in DNA binding region of *CEBPA* gene on RNA secondary structure

The effect of these identified mutations in the coding region on RNA folding (RNA secondary structure) was checked using the Vienna package's RNA fold function. Both these mutations were found to cause the abnormal folding of RNA (Figure 4.16). We noted

the RNA folding energy (△G), i.e. Minimum Free Energy (MFE) from Vienna package for the examination of the stability and secondary structure of these variations (Table 13). If MFE value is lower, RNA secondary structure will be more stable.

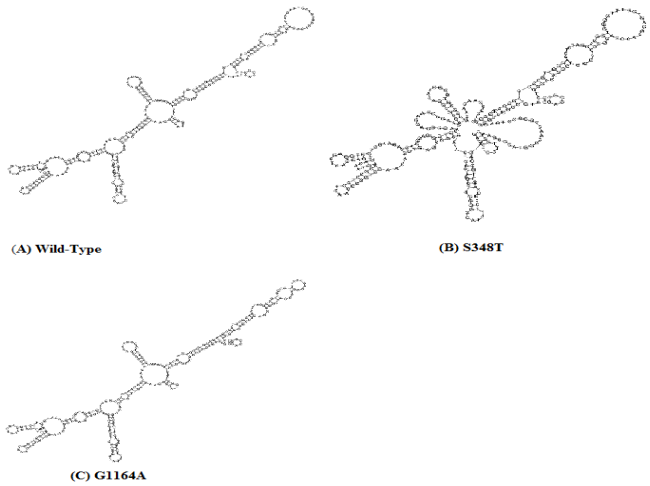

Figure 12: Changes in RNA secondary structure due to alternations in DNA binding region of *CEBPA* gene predicted by RNA fold function of Vienna Package

Table 13: Minimum free energy of wild-type and mutant residues predicted by Vienna package

Mutations	Minimum Free Energy (△G)
Wild-Type	-121.50
S348T	-102.80
C.+G87A	-121.10

Discussion

Acute myeloid leukemia patients were examined in the present study to identify *CEBPA* gene mutations, their role in AML development, and their effect on protein structure and function. Alteration of *CEBPA* were identified in 7 patients (23.33 %). This frequency is analogues to a previous study. [38] A previous study also reported a mutation frequency of 24.3 %.[39] Another study also reported the *CEBPA* mutation frequency of 18 % [40], which is lower than the frequency reported in our study. This decrease in frequency may be due to the difference in lifestyle, geographical, and genetic between these populations. No significant difference was noted in blast cells percentage, hemoglobin level, TLC, and platelets count between non-mutant and *CEBPA* mutant patients, consistent with previous reports.[41] High LDH was observed in *CEBPA* mutant AML patients than in non-mutant patients. Previous studies reported lower LDH in *CEBPA* mutant patients than in non-mutant patients.[41, 42] This difference may be due to differences in age, FAB type, and size of the study group.

Missense mutations that occur in coding region disturb the function and stability of protein and results in many disease conditions, while mutations in non-coding region don't affect the protein but sometimes can affect the phenotype i.e. may affect the splicing process or speed up or lower the protein synthesis process.[43] In the present study, two novel missense mutations, S348T and C.+G87A were observed in AML patients, which had never been reported previously. In a previous study, Q311P missense mutations were reported in DNA binding region of *CEBPA* gene, which showed that mutant Q311P residue had reduced DNA binding ability than the wild residue, and it resulted in the loss of function allele. [44] Another previous study in pediatric AML patients reported G5800T, G690T, C5683T, C573T, C5444T, and C336T nucleotide change missense mutations in *CEBPA* gene that did not affect encoded amino acids, but C5599A and C489A nucleotide alternations resulted in the substitution of Glu163>Asp. All these mutations were benign [45]. L315P (*CEBPA*sm) and R297P, S299R, and R300G (*CEBPA*dm) amino acid substitutions were also reported in DNA binding region in a previous report.[46] Another previous study reported R297K, R297A, R297Q amino acid substitution mutations.[22] In this study, harmful effects of identified S348T alternation were predicted by using different online softwares. Variation was observed in the results of this software. Polyphen-2 and Panther predicted S348T mutation as damaging, while SNAP2 and PEOVEAN predicted that the S348T mutation was neutral and had no pathogenic effects. Align GVGD indicated that the mutation was likely affected. Stability in the structure of a protein is very important for the normal proper functioning of that protein, and any change in the stability of the protein may disturb its function.[47] Effect of detected S348T mutation on protein stability was predicted using Mutant-I software. Mutant-I showed that the mutation resulted in a decrease in protein's stability. This decrease in protein structure stability may increase protein misfolding, aggregation, and degradation.[48]

Change of amino acid greatly affects the chemical and physical parameters, integrity, hydrophobicity, and protein structure. To check the effect of identified S348T mutation on these parameters, HOPE tool was used. In this study, it was observed that mutant residue was greater in size than wild-type residue. This increase in size may cause bumps in protein structure.[48] The evolutionary conservation profile of protein predicts the determination of mutation's severity. Mutations of highly conserved regions are more harmful than those found in the variable areas.[49] Netsurf and conserve tools were used to check the evolutionary conservation status of S348T mutation to predict possible effects. Consurf indicated that the S348 mutation was present in the conserved region with conservation score of 7 and was exposed. Netsurf also predicted that the mutant residue was exposed. Previous studies reported that the mutation in the conserved areas affects protein function.[32]

NetsurfP and SOPMA analyzed protein secondary structure. The NetsurfP tool indicated that the coil structure was dominant over other secondary structures, and mutation lay in the coiled

region. SOPMA also suggested that the mutation lied in a random coiled region, and coil structure was dominant ((55.03 %) followed by alpha helix (30.17 %). Previous studies reported that the mutations in coiled regions are deleterious and polymorphic.[32] The difference in protein structure was analyzed by using SWISS model to generate the 3D models of both mutant and wild-type residues. Results showed no difference in both wild-type and mutant residue. ProtParam, ProSA Web, VADAR Analysis, and Molprobity further investigated these models. ProtoParam analyzed physiochemical properties. In this study, the molecular weight of S348T was greater than wild-type residue because S348T has a greater length than the wildtype. There was no difference in the theoretical pI of mutant and wild residues. Theoretical pI was greater than 7, indicating that both residues were basic. Aliphatic Index showed that mutant protein was less thermostable than wildtype protein. Instability Index indicated that the mutant reside was less stable than the wild type reside. Both mutant and wild-type residue have negative GRAVY values, indicating that both residues are hydrophobic. ProsaWeb revealed minor structural changes. Structure of both wild and mutant reside was comparatively analyzed by VADAR analysis. Results showed no difference in the secondary structures of both resides. Ramachandran plot predicted the validity of protein model. The presence of greater than 90 % residue lies in the allowed protein region, and the structure is said to be of good quality.[50] In this study, 100% amino acids of both wild and mutant residues lay in allowed, showing good model quality.

Various interactions and associations are present between different proteins due to the presence of non-covalent contacts.[51] In this study, The STRING database predicted that *CEBPA* gene is functionally associated with *JUN, CDK2, CDK4, SPI1, FOXO1, PPARG, HDAC1, CDK2, EP300, E2F4,* and *FOS*. Any abnormal change in *CEBPA* gene will affect the interaction of *CEBPA* with these others proteins. As a result, it will disturb the function of *CEBPA* and other proteins, as a result of which many important processes will be disturbed and mutations may also affect the secondary structure of RNA and may result in disease development due to abnormal changes in RNA secondary structure.[52] In this study, it is observed that the S348T and C.+G87A mutant residues had higher minimum free energy than the wild-type residue. This increase in minimum free energy value indicated a decrease in the stability of RNA secondary structure that resulted in RNA folding and also affected RNA localization.

Conclusion

This study suggests that the SNP S348T being reported for the first time in AML patients can have some effects on the DNA binding ability of *CEBPA* gene. In silico tools showed variation in predicting pathogenicity of S348T mutation, showing it moderately affects the gene function.

References

(1). Port, M.; Boettcher, M.; Thol, F.; Ganser, A.; Schlenk, R.; Wasem, J.; . . . Pouryamout, L. Prognostic significance of FLT3 internal tandem duplication, nucleophosmin 1, and CEBPA gene mutations for acute myeloid leukemia patients with normal karyotype and younger than 60 years: a systematic review and meta-analysis. *Annals of Hematology*. 2014, 93(8), 1279-1286.

(2). Omer, S.-H.; Kordofani, A.; Osman, I.; Musa, O.; Altayb, H.; Elamin, B. Prevalence of the different FAB sub type of acute myeloid leukemia related to hematological parameters in Sudanese. *Journal of Hematology and Blood Disorders*. 2017, 3(1), 1-5.

(3).Stahl, M.; Goldberg, A.-D. Immune checkpoint inhibitors in acute myeloid leukemia: novel combinations and therapeutic targets. *Current Oncology Reports*. 2019, 21(4), 1-10.

(4). Niparuck, P.; Limsuwanachot, N.; Pukiat, S.; Chantrathammachart, P.; Rerkamnuaychoke, B.; Magmuang, S.; . . . Angchaisuksiri, P. Cytogenetics and FLT3-ITD mutation predict clinical outcomes in non transplant patients with acute myeloid leukemia. *Experimental Hematology & Oncology*. 2019, 8(1), 1-14.

(5). Norsworthy, K.-J.; By, K.; Subramaniam, S.; Zhuang, L.; Del Valle, P.-L.; Przepiorka, D.; . . . Leong, R. FDA approval summary: glasdegib for newly diagnosed acute myeloid leukemia. *Clinical Cancer Researc.* 2019, 25(20), 6021-6025.

(6). Lagunas-Rangel, F.-A.; Chávez-Valencia, V.; Gómez-Guijosa, M.-Á.; Cortes-Penagos, C. Acute myeloid leukemia—genetic alterations and their clinical prognosis. *International Journal of Hematology-Oncology and Stem Cell Research.* 2017, 11(4), 328.

(7). Ilhan, G.; Karakus, S.; Andic, N. Risk factors and primary prevention of acute leukemia. *Asian Pacific Journal of Cancer Prevention*. 2006, 7(4), 515-517.

(8). Polychronakis, I.; Dounias, G.; Makropoulos, V.; Riza, E.; Linos, A. Work-related leukemia: a systematic review. *Journal of Occupational Medicine and Toxicology*. 2013, 8(1), 1-16.

(9). Zhang, L.; Wang, S.-A. A focused review of hematopoietic neoplasms occurring in the therapy-related setting. *International Journal of Clinical and Experimental Pathology*. 2014 7(7), 3512-3523.

(10). Giuseppe, L.; Livio, P.; Dina, B.-Y.; Maria Teresa, V. Therapy-related leukemia and myelodysplasia: susceptibility and incidence. *Haematologica.* 2007, 92(10), 1389-1398. doi: 10.3324/haematol.11034

(11). Bispo, J.-A.-B.; Pinheiro, P.-S.; Kobetz, E.-K. Epidemiology and etiology of leukemia and lymphoma. *Cold Spring Harbor Perspectives in Medicine.* 2020, 10(6), a034819.

(12). Ugai, T.; Matsuo, K.; Oze, I.; Ito, H.; Wakai, K.; Wada, K.; . . . Kitamura, Y. Smoking and subsequent risk of acute myeloid leukaemia: a pooled analysis of 9 cohort studies in Japan. *Hematological Oncology*. 2018, 36(1), 262-268.

(13). Wang, M.; Wang, R.; Wang, H.; Chen, C.; Qin, J. Difference in gene mutation profile in patients with refractory/relapsed versus newly diagnosed acute myeloid leukemia based on targeted next-generation sequencing. *Leukemia & Lymphoma*. 2021, 62(10), 2416-2427. doi: 10.1080/10428194.2021.1919661

(14). Medinger, M.; Passweg, J.-R. Acute myeloid leukaemia genomics. *British Journal of Haematology*. 2017, 179(4), 530-542.

(15). Döhner, H.; Weisdorf, D.-J.; Bloomfield, C.-D. Acute myeloid leukemia. *New England Journal of Medicine*. 2015, 373(12), 1136-1152.

(16). Rázga, F.; Dvořáková, D.; Jurŕček, T.; Jeziskova, I.; Křístková, Z.; Mayer, J. CEBPA Gene Mutational Status. *Molecular Diagnosis & Therapy*. 2012, 13, 195-200.

(17). Renneville, A.; Roumier, C.; Biggio, V.; Nibourel, O.; Boissel, N.; Fenaux, P.; Preudhomme, C. Cooperating gene mutations in acute myeloid leukemia: a review of the literature. *Leukemia*. 2008, 22(5), 915-931. doi: 10.1038/leu.2008.19

(18). Su, L.; Tan, Y.; Lin, H.; Liu, X.; Yu, L.; Yang, Y.; . . . Jin, F. Mutational spectrum of acute myeloid leukemia patients with double CEBPA mutations based on next-generation sequencing and its prognostic significance. *Oncotarget.* 2018, 9(38), 24970.

(19). Quintana-Bustamante, O.; Smith, S.; Griessinger, E.; Reyal, Y.; Vargaftig, J.; Lister, T. A.; . . . Bonnet, D. (2012). Overexpression of wild-type or mutants forms of CEBPA alter normal human hematopoiesis. *Leukemia*. 2012, 26(7), 1537-1546.

(20). Schmidt, L.; Heyes, E.; Grebien, F. Gain-of-function effects of N-terminal CEBPA mutations in acute myeloid leukemia. *BioEssays*. 2020, 42(2), 1900178.

(21). Hollink, I.-H.; Van Den Heuvel-Eibrink, M.-M.; Arentsen-Peters, S.-T.; Zimmermann, M.; Peeters, J.-K.; Valk, P.-J.; . . . De Bont, E.-S.-J.-h. Characterization of CEBPA mutations and promoter hypermethylation in pediatric acute myeloid leukemia. *Haematologica*. 2011, 96(3): 384-392.

(22). Kowenz-Leutz, E.; Schuetz, A.; Liu, Q.; Knoblich, M.; Heinemann, U.; Leutz, A. Functional interaction of CCAAT/enhancer-binding-protein-α basic region mutants with E2F transcription factors and DNA. *Biochimica et Biophysica Acta*. 2016, 1859(7), 841-847. doi: 10.1016/j.bbagrm.2016.04.008

(23). Su, L.; Shi, Y.-Y.; Liu, Z.-Y.; Gao, S.-J. Acute Myeloid Leukemia With CEBPA Mutations: Current Progress and Future Directions. *Frontiers in Oncology*. 2022, 12. doi: 10.3389/fonc.2022.806137

(24). Wakita, S.; Sakaguchi, M.; Oh, I.; Kako, S.; Toya, T.; Najima, Y.; . . . Yamaguchi, H. Prognostic impact of CEBPA bZIP domain mutation in acute myeloid leukemia. *Blood Advances*. 2022, 6(1), 238-247. doi: 10.1182/bloodadvances.2021004292

(25). Chareonsirisuthigul, T.; Magmuang, S.; Chuncharunee, S.; Rerkamnuaychoke, B. Detectioon of CEBPA mutation gene in acute myeloid leukemia patients. Ramathibodi Medical Journal. 2017, 40, 15-24.

(26). Braun, T.-P.; Coblentz, C.; Smith, B.-M.; Coleman, D.-J.; Schonrock, Z.; Carratt, S.-A.; . . . Maxson, J.-E. Combined inhibition of JAK/STAT pathway and lysine-specific demethylase 1 as a therapeutic strategy in CSF3R/CEBPA mutant acute myeloid leukemia. *Proceedings of the National Academy of Sciences*. 2020, 117(24), 13670-13679.

(27). Zhang, Y.; Wang, F.; Chen, X.; Zhang, Y.; Wang, M.; Liu, H.; . . . Liu, H. Companion gene mutations and their clinical significance in AML with double mutant CEBPA. *Cancer Gene Therapy*. 2020, 27(7), 599-606. doi: 10.1038/s41417-019-0133-7

(28). Ernst, C.; Hahnen, E.; Engel, C.; Nothnagel, M.; Weber, J.; Schmutzler, R.-K.; Hauke, J. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Medical Genomics*. 2018, 11(1), 35. doi: 10.1186/s12920-018-0353-y

(29). Mustafa, M.-I.; Murshed, N.-S.; Abdelmoneim, A.-H.; Makhawi, A.-M. Extensive In Silico Analysis of the Functional and Structural Consequences of SNPs in Human ARX Gene. *BioRxiv*. 2020, 100447

(30). Ou, L.; Przybilla, M.-J.; Whitley, C.-B. Phenotype prediction for mucopolysaccharidosis type I by in silico analysis. *Orphanet Journal of Rare Diseases*. 2017, 12(1), 125. doi: 10.1186/s13023-017-0678-1

(31). Arshad, M.; Bhatti, A.; John, P. Identification and in silico analysis of functional SNPs of human TAGAP protein: A comprehensive study. *Plos One*. 2018, 13(1), e0188143.

(32). Mustafa, M.-I.; Murshed, N.-S.; Abdelmoneim, A.-H.; Makhawi, A.-M. In silico analysis of the functional and structural consequences of SNPs in human ARX gene associated with EIEE1. *Informatics in Medicine Unlocked*. 2020, 21, 100447.

(33). Fathollahi, M.; Fathollahi, A.; Motamedi, H.; Moradi, J.; Alvandi, A.; Abiri, R. In silico vaccine design and epitope mapping of New Delhi metallo-beta-lactamase (NDM): an immunoinformatics approach. *BMC Bioinformatics*. 2021, 22(1), 1-24.

(34). Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; . . . Bordoli, L. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research.* 2018, 46(W1): W296-W303.

(35). Hasanzadeh, S.; Habibi, M.; Shokrgozar, M.-A.; Ahangari Cohan, R.; Ahmadi, K.; Asadi Karam, M.-R.; Bouzari, S. In silico analysis and in vivo assessment of a novel epitope-based vaccine candidate against uropathogenic Escherichia coli. *Scientific Reports*. 2020, 10(1), 16258. doi: 10.1038/s41598-020-73179-w

(36). Pa'ee, K.-F.; Razali, N.; Sarbini, S.-R.; Ramonaran Nair, S.-N.; Yong Tau Len, K.; Abd-Talib, N. The production of collagen type I hydrolyzate derived from tilapia (Oreochromis sp.) skin using thermoase PC10F and its in silico analysis. *Food Biotechnology*. 2021, 35(1), 1-21. doi: 10.1080/08905436.2020.1869040

(37). Sato, T.; Vries, R.-G.; Snippert, H.-J.; van de Wetering, M.; Barker, N.; Stange, D.-E.; . . . Clevers, H. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature*. 2009, 459(7244), 262-265. doi: 10.1038/nature07935

(38). Su, L.; Li, X.; Gao, S.-J.; Yu, P.; Liu, X.-L.; Tan, Y.-H.; Liu, Y.-M. Cytogenetic and genetic mutation features of de novo acute myeloid leukemia in elderly Chinese patients. *Asian Pacific Journal of Cancer Prevention*. 2014, 15(2), 895-898. doi: 10.7314/apjcp.2014.15.2.895

(39). Leecharendkeat, A.; Tocharoentanaphol, C.; Auewarakul, C.-U. CCAAT/enhancer binding protein-alpha polymorphisms occur more frequently than mutations in acute myeloid leukemia and exist across all cytogenetic risk groups and leukemia subtypes. *International Journal of Cancer*. 2018, 123(10), 2321-2326. doi: 10.1002/ijc.23796

(40). Sarojam, S.; Raveendran, S.; Vijay, S.; Sreedharan, J.; Narayanan, G.; Sreedharan, H. Characterization of CEBPA Mutations and Polymorphisms and their Prognostic Relevance in De Novo Acute Myeloid Leukemia Patients. *Asian Pacific Journal of Cancer Prevention*. 2015, 16(9), 3785-3792. doi: 10.7314/apjcp.2015.16.9.3785

(41). Awad, M.-M.; Aladle, D.-A.; Abousamra, N.-K.; Elghannam, D.-M.; Fawzy, I.-M. CEBPA gene mutations in Egyptian acute myeloid leukemia patients: impact on prognosis. *Hematology*. 2013, 18(2), 61-68. doi: 10.1179/1607845412y.0000000032

(42). Su, L.; Gao, S.; Liu, X.; Tan, Y.; Wang, L.; Li, W. CEBPA mutations in patients with de novo acute myeloid leukemia: data analysis in a Chinese population. *OncoTargets and Therapy*. 2016, 9, 3399-3403. doi: 10.2147/ott.s94975

(43). Tan, K.-P.; Kanitkar, T.-R.; Kwoh, C.-K.; Madhusudhan, M.-S. Packpred: Predicting the Functional Effect of Missense Mutations. Frontiers in Molecular Biosciences. 2021, 8. doi: 10.3389/fmolb.2021.646288

(44). Pathak, A.; Seipel, K.; Pemov, A.; Dewan, R.; Brown, C.; Ravichandran, S.; . . . Stewart, D.-R. Whole exome sequencing reveals a C-terminal germline variant in CEBPA-associated acute myeloid leukemia: 45-year follow up of a large family. *Haematologica*. 2016, 101(7), 846-852. doi: 10.3324/haematol.2015.130799

(45). Akin, D.-F.; Oner, D.-A.; Kurekci, E.; Akar, N. Determination of CEBPA mutations by next generation sequencing in pediatric acute leukemia. *Bratislava Medical Journal*. 2018, 119(6), 366-372. doi: 10.4149/bll_2018_068

(46). Kang, Y.; Chen, X. The clinical characteristics and prognosis of cytogenetically normal AML with single mutations of CEBPA. *International Journal of Laboratory Hematology*. 2021, 43(6), 1424-1431. doi: 10.1111/ijlh.13612

(47). Hossain, M.-S.; Roy, A.-S.; Islam, M.-S. In silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. Scientific Reports. 2020, 10(1), 14542. doi: 10.1038/s41598-020-71457-1

(48). Soltani, I.; Bahia, W.; Radhouani, A.; Mahdhi, A.; Ferchichi, S.; Almawi, W.-Y. Comprehensive in-silico analysis of damage associated SNPs in hOCT1 affecting Imatinib response in chronic myeloid leukemia. *Genomics*. 2021, 113(1), 755-766. doi: https://doi.org/10.1016/j.ygeno.2020.10.007

(49). Rozario, L.-T.; Sharker, T.; Nila, T.-A. In silico analysis of deleterious SNPs of human MTUS1 gene and their impacts on subsequent protein structure and function. *Plos one*. 2021, 16(6), e0252932.

(50). Al-Kindi, M.-N.; Al-Khabouri, M.-J.; Al-Lamki, K.-A.; Palombo, F.; Pippucci, T.; Romeo, G.; Al-Wardy, N.-M. In silico analysis of a novel causative mutation in Cadherin23 gene identified in an Omani family with hearing loss. *Journal, Genetic Engineering & Biotechnology*. 2020, 18(1), 8. doi: 10.1186/s43141-020-0021-4

(51). Rao, V.-S.; Srinivas, K.; Sujini, G.-N.; Kumar, G.-N. (2014). Protein-protein interaction detection: methods and analysis. *International Journal of Proteomics*. 2014: 147648. doi: 10.1155/2014/147648

(52). Shatoff, E.; Bundschuh, R. Single nucleotide polymorphisms affect RNA-protein interactions at a distance through modulation of RNA secondary structures. *PLoS Computational Biology*. 2020, 16(5): e1007852.