Deciphering the structural and functional impact of missense mutations in Egr1-DNA interacting interface: an integrative computational approach

Hemavathy Nagarajan and Umashankar Vetrivel

ABSTRACT

Early growth response-1 (Egr1) is a zinc-finger transcription factor that plays a critical role in controlling cell growth, proliferation, differentiation, angiogenesis, and apoptosis. Egr1 is induced by many growth factors, cytokines, and stress signals and is also known to be involved in several pathological conditions like cancer, neurological and ocular disorders. The DNA binding domain of Egr1 is a highly conserved Cys2His2 (C2H2) zinc finger (ZNF) domain which specifically binds to GC-rich consensus sequence GcG(G/T)GGGCG and activates transcription. As the C2H2 domain specifically recognizes its DNA target, the mutations spanning this region shall perturb DNA recognition and may hinder transcription of target genes. Therefore, in this study, the missense mutations occurring specifically at the DNA binding domain (DBD) of Egr1 were probed by computational approaches involving in silico screening of pathogenic and functional mutants coupled with extensive molecular dynamics simulations, to determine the mutants that affect its structural stability and interactions with DNA. From the pathogenicity analysis of 38 missense mutations spanning Egr1-DBD, 17 were predicted as pathogenic, and 7 amongst these were found to have functional impact on Egr1. On combined analysis of molecular dynamics simulation, residue interaction analysis and Egr1-DNA interaction analysis results, the mutants R371C and R375C showed least impact, whilst, H382R tend to increase the structural stability, whereas R360H, H390R, E393V, and H414Y conferred greater impact by altering the structural stability and DNA interactions. Hence, this study exposes the prospects of considering these 4 deleterious mutations for clinical significance, but needs further experimental validation.

HIGHLIGHTS

- Egr1’s DNA binding domain is a highly conserved Cys2His2 (C2H2) zinc finger domain that specifically recognizes its DNA target.
- Mutations spanning in the DNA binding domain shall perturb DNA recognition and may hinder transcription.
- Among the missense mutations, mutants R360H, H390R, E393V, and H414Y were inferred to have a greater impact on Egr1 by altering the structural stability and DNA interactions.

1. Introduction

Zinc-finger-containing proteins constitute the most abundant protein superfamily and are known to be involved in a variety of cellular activities such as development, differentiation, and tumor suppression (Abbehausen, 2019; Cassandri et al., 2017; Ravasi et al., 2003). These proteins also act as transcriptional regulators due to their association with DNA-binding motifs (Bird et al., 2003; Emerson & Thomas, 2009). Early growth response-1 (Egr1) also known as NGFI-A, Zif268, Krox24, and Tis8, is a serum-inducible zinc finger transcription factor that plays a critical role in controlling cell growth, proliferation, differentiation, angiogenesis, and apoptosis (Gashler & Sukhatme, 1995; O’Donovan et al., 1999). Egr1 is induced by many growth factors, cytokines, and stress signals. Egr1 is also known to be involved in a number of pathological conditions such as cancer, neurological disorders, and ocular disorders (Duclot & Kabbaj, 2017; Pagel & Deindl, 2012). Moreover, Egr1 is known to comprise a well-conserved DNA binding domain composed of three zinc fingers that bind to specific GC-rich consensus sequence GcG(G/T)GGGCG (Hamilton et al., 1998; Russo et al., 1993). It also activates transcription by binding to DNA as a monomer (Carman & Monroe, 1995; Crosby et al., 1991).

The DNA binding domain of Egr1 is a highly conserved Cys2His2 zinc finger domain. In particular, the Cys2His2 zinc finger (C2H2-ZNF) domain is a highly conserved sequence specific DNA-binding region that play a crucial role in gene regulation. The C2H2 domains are also well confirmed on the key role of the amino acids at these positions for the specific binding to DNA (Fedotova et al., 2017; Namuswe & Berg, 2012; Wolfe et al., 2000). The sequence specificities of
C2H2-ZNF proteins are highly variable, with many human C2H2-ZNF proteins having a unique set of DNA-contacting specificity residues (Eom et al., 2016; Garton et al., 2015). C2H2-ZNF proteins also often harbor one or more of a small number of effector domains. An individual C2H2-ZNF domain contains a well-conserved DNA-binding structural interface and specifically recognizes its DNA target via amino acids C2H2-ZNF proteins also often harbor one or more of a small number of effector domains. An individual C2H2-ZNF domain contains a well-conserved DNA-binding structural interface and specifically recognizes its DNA target via amino acids (Kluska et al., 2018; Yella et al., 2018). Therefore, mutations in amino acids spanning this region may interfere with DNA binding and could alter the functionality of the Zinc-finger proteins (ZNFs), which highly aids in regulation of wide range of cellular processes. Thereby, the alterations in ZNFs are known to be involved in cancer progression and metastasis, neurodegeneration, etc (Végran et al., 2013). Hence, it is essential to probe the mutations occurring at the DNA binding domain of Egr1 that affects its structural stability and its interaction with DNA.

In the human genome, variations caused by single base pair changes known as SNPs, amongst these, missense SNPs (in specific, non-synonymous SNPs (nsSNPs)) contribute towards significant functional and structural variations (Agrahari et al., 2018). Thus, missense SNPs might be involved in various genetic disorders as they result in structurally and functionally deformed proteins. Moreover, various earlier studies have provided insights on the impact of the mutations occurring specifically at the DNA binding domain in different proteins (Agrahari et al., 2018; Nixon et al., 2004; Xiao et al., 2020; Yoneda et al., 2012). Hence, this study is aimed to investigate the nsSNPs of Egr1 specifically at the DNA binding domain (DBD), and to probe their effects on its structure and function. Therefore, an extensive in silico analysis of nsSNPs was implemented in prioritizing the mutants based on their pathogenicity, functionality impact, and conservation (Figure 1). Moreover, advanced computational methods, such as essential molecular dynamics and structural impact analysis were also performed to decipher the impact of DNA binding due to Egr1’s (DBD) structural changes upon mutation.

2. Materials and methods

2.1. Retrieval of SNP details of Egr1

The sequence information of Egr1 protein was retrieved from UniProt using the ID: P18146 and it has a length of 543 amino acids. The SNPs were retrieved from dbSNP database (https://www.ncbi.nlm.nih.gov/snp/) (Sherry et al., 2001), Ensembl database (https://asia.ensembl.org/index.html) (Cunningham et al., 2019), gnomAD browser (https://gnomad.broadinstitute.org/), Exome Variant Server (https://evs.gs.washington.edu/EVS/). In this study, mutations occurring at the DNA binding domain (DBD) (335–423) of Egr1 that affects the DNA interaction were considered for further analysis.

2.2. Screening of deleterious nsSNPs

The shortlisted Egr1’s DBD nsSNPs were initially examined for their deleterious or pathogenic effect with sequence and evolutionary-based and machine learning-based approaches (Nagarajan et al., 2020). Thereby the servers namely, SIFT, Polyphen2, PANTHER-PSEP, PMut, MutPred2, were used to screen the pathogenic mutants that might have a structural or functional impact on Egr1’s DBD. Among the servers used, SIFT (Sorting Intolerant from Tolerant) prediction is based on a presumption that certain amino acid positions in protein sequences are conserved through evolution. Consequently, a substitution in the conserved regions that has an impact on protein functionality is considered deleterious (Sim et al., 2012). The SIFT output score ranges from 0 to 1 where substitutions ≤ 0.05 indicate a deleterious effect and those variants greater than 0.05 are tolerant. While, PANTHER-PSEP (Position-Specific Evolutionary Preservation) predicts potentially pathogenic or deleterious nsSNPs based on the measure of evolutionary preservation of each amino acid of the target, where the amino acids that are preserved for a long evolutionary timescale were considered deleterious (Tang & Thomas, 2016). PolyPhen-2 (Polymorphism Phenotyping v2) utilizes BLAST to identify homologues of the input missense variant. A matrix is then constructed from the aligned sequences to calculate the PSIC (Position-Specific Independent Count) scores ranging from 0 to 1, from which phylogenetic information is derived (Adzhubei et al., 2013). With sequence alignments, phylogenetic and structural data, the effect of a particular point mutation in the protein is predicted by a naive Bayesian classifier, which classifies the nsSNPs as ‘Possibly Damaging’ and ‘Probably Damaging’ (PSIC > 0.5) or ‘Benign’ (PSIC < 0.5). It also includes structure-based characteristics such as solvent accessibility, changes in accessibility for the buried residues, and B-factor. PMut (Prediction of pathological Mutation on proteins) prediction is based on the neural network to predict the pathogenicity of amino acid substitutions. The scores are accompanied by a reliability index ranging from 0 to 10, where 0 is the most unreliable and 10 is the most reliable (López-Ferrando et al., 2017). MutPred (Mutation Prediction) is a random-forest classifier that analyses the changes caused by amino acid substitutions on a molecular level. It provides a ‘g’ score which signifies the probability for the extent to which a variant is deleterious. MutPred classifies nsSNPs as deleterious (p > 0.75), disease-associated (0.5 < p < 0.75), or neutral (p < 0.5) by taking into account structural and functional aspects, conservation of the protein sequence, and the conformational dynamics of the molecule. Finally, the mutants above the threshold scores across all the servers were considered deleterious and were further probed for their impact on functionality (Li et al., 2009).

2.3. Functionality analysis of deleterious mutants

The nsSNPs have a higher impact on both structural and functional aspects of proteins. Therefore, the resultant deleterious mutants from the earlier screening process were prioritized further based on their functionality impact using the Mutpred2 server. P-score of MutPred (Mutation Prediction) (Li et al., 2009) represents the impact on the structural and/or functional properties such as gain or loss in helical
structures, intrinsic disorder, and phosphorylation, metal-binding disorder, loss and gain of Post-translational modifications, etc. Since the loss of ZNF features has a greater impact on DNA binding which in turn leads to its functionality related disorders. Based on the P-score and the mutants that have a higher impact on the metal-binding site and loss of ZNF helices were considered for their destabilization analysis by MD simulation.

2.4. Structural stability analysis of WT and mutant forms by molecular dynamics simulation

The WT-Egr1 DBD in complex with DNA (PDB ID: 4R2A) (Hashimoto et al., 2014) had missing residues and atoms, moreover, it consists of methylated DNA. Since Egr1 can recognize the target sequence of DNA, despite methylated and non-methylated states. Thereby WT-Egr1 DBD in complex with non-methylated DNA was considered for generating the mutant forms using Maestro suite of Schrodinger (Jayaraj et al., 2019; Kirubakaran et al., 2014; Sastry et al., 2013). Since the protonation states of the histidine and cysteine residues that constitute the Cys2His2 (C2H2) zinc finger domain Zinc coordination by C2H2 histidines occurs with the ND1-H tautomer (only ND1 protonated) (Kemme et al., 2019; Li et al., 2008), whereas cysteine residues coordinates metal ions in their deprotonated form (Godwin et al., 2017; Pace & Weerapana, 2014). Thereby both the WT and mutants were pre-processed and minimized using Prepwizard module of Maestro with appropriate protonation states of Histidines and cysteines.

For both WT and mutant forms, the MD simulations were performed using the Gromacs 5.1.4 simulation package with Amber99sb-ParmBSC0 (Pérez et al., 2007) as a force field. Initially, the complexes were centered in a cubic environment of 1 nm and were solvated using the TIP3P water model. Further, the solvated systems were then neutralized using sodium counter ions and energy minimized using the steepest descent algorithm for 50,000 steps. During these steps, the coulombs and Vander Waals interactions were maintained at a cut-off of 0.9 nm and 1.0 nm, respectively. Later, the system was equilibrated under NVT and NPT ensemble using Leapfrog integrator for 100 picoseconds (ps) while simultaneously...
maintaining Berendsen temperature coupling (Berendsen et al., 1984) and pressure coupling at 300 K and 1 bar, respectively. Also, the long-range electrostatics were maintained by Particle-Mesh Ewald (PME) and short-range electrostatics maintained by coulombs and Vander Waals at a cut off 1.0 nm each at a time-step of 2 femtoseconds (fs) (Darden et al., 1993). LINCS algorithm was used to maintain the geometry of molecules (Hess et al., 1997). A well-equilibrated system was then subjected to MD simulations for 100 nanoseconds (ns) each at 300 K without any constraints, where the conformations were sampled at an interval of 2 ps. Finally, the trajectories were analyzed using GROMACS utilities to probe the conformational changes, fluctuations, compactness, and stability of the protein-DNA complexes. Moreover, to estimate the global concerted movement which represents the functional behavior of the protein, Principal Component Analysis (PCA) and free energy landscape (FEL) analysis were performed (Maissuradze et al., 2010; Nagarajan & Vetrievel, 2018, 2020). Further, from the resulting minimum energy cluster of the proteins, representative stable near-native conformations were used for further studies. The schematic diagrams of protein-DNA interactions in the pre-and post-MD simulated complex were deduced using Nucplot (https://www.ebi.ac.uk/thornton-srv/software/NUCPLLOT/) (Luscombe et al., 1997). Further, the pivotal interactions of protein-DNA complexes were also analyzed by PLIP (Protein-Ligand Interaction Profiler) (Salentin et al., 2015). In addition, the electrostatic potential calculations of conformations were carried out using PBEQ Solver module of CHARMM-GUI server (Im et al., 1998; Jo et al., 2008).

2.5. Residue interaction network analysis

Residue Interaction Network (RIN) analysis provide the graphical network of the inter-residue interactions of a protein, where the residues are considered as nodes and their corresponding physicochemical interactions between them such as hydrogen bond, van der Waals, Ionic, π-cation, and π–π stack interactions as edges to determine the connectivity of the residues. In the case of mutations in protein, the topological changes based on RIN infer the loss of covalent and non-covalent bonds that may alter the protein network. In this study, RIN was generated by Residue Interaction Network Generator (RING) v2.0.1 (Piovesan et al., 2016) for the conformations generated for every 200 frames of Egr1’s DBD, while the DNA and the hetero molecules were not considered. The resulting XML files were imported into Cytoscape v3.8.2 (Shannon et al., 2003), the plugin RNalyzer (Doncheva et al., 2011) was used to visualize the residue interaction network. To visualize the three-dimensional structure of proteins, PDB files were imported to UCSF Chimera v1.13 (Pettersen et al., 2004).

3. Results and discussion

3.1. Retrieval of SNPs

Among 838 SNPs of Egr1, 532 were found to be missense mutations. However, only 38 missense mutations out of 532 were found to span the DNA binding domain (DBD) of Egr1. Therefore, these 38 mutations of Egr1’s DBD region were considered for further analysis to probe the deleterious effects and their intensity.

3.2. Screening of pathogenic mutants

Amongst the 38 mutations occurring at Egr1’s DBD region, SIFT reported only 28 mutations to be highly deleterious out of 38 mutations (73.7%). While Polyphen showed 25 deleterious mutations (65.8%) and Mutpred predicted 26 deleterious mutations (68.4%) among the 38 missense SNPs. The other tools namely, Panther_subspec and pMut showed 24 mutations (63.1%) and 21 mutations (55.3%), respectively. On overall consensus analysis of deleterious mutations among all the servers, only 17 mutations (Table S1) were found to be deleterious or pathogenic and were further considered for functionality analysis.

3.3. Functionality impact analysis of pathogenic mutants

On analyzing the 17 shortlisted pathogenic mutations (Table 1), only 7 mutations were found to be involved in altered metal binding (zinc-binding), while the remaining 10 were found to lead to altered disorder interface which doesn’t have a major impact on the DNA binding. Therefore, the 7 mutations which showed a higher probability of altering the metal binding at the ZNF regions were considered, as the distortion of metal coordination may lead to the improper conformation of the ZNF residues, thus leading to the lack of proper recognition of DNA.

3.4. Deciphering the effect on conformational changes and compactness of native and mutants of Egr1

The Wild type (Figure 2) and the resultant 7 mutations (Figure 3) that were found to be pathogenic and has major impact on Egr1’s functionality were considered for molecular dynamics simulation studies for 100 ns in order to infer its structural destabilization effect.

From the RMSD analysis (Figure 4a and Table S2), WT was observed to have least deviations of about ~0.45 nm till 80 ns yet has attained major deviations (~ 0.6 nm) in the last 20 ns. While amongst the mutants, the mutants R360H, H390R, R375C have shown deviations of about ~0.55~0.6 nm than the WT and have tried to maintain stability after 40 ns till the end of MD production run. In the case mutant E393V, unstable deviations were observed within 40~80 ns, whereas mutants, H382R and H414Y were observed to have the least deviations than that of the WT. Whereas, the RMSF plot (Figures 4b and S1) infers that except for the N-terminal and C-terminal residues, E343 (~0.6 nm), R371 (~0.4 nm), were found to be highly fluctuating in mutants H390R, H382R, R375C, respectively. While D399 (~0.4 nm) was observed to show higher fluctuation in both WT and all mutants except for the mutants (H382R, R371C). From the radius of the gyration plot (Rg) (Figure 4c and Table S2), it was inferred that...
except for mutant R375C, all other mutants and the WT were found to have the least compactness.

3.5. Analyzing the distortion of zinc coordinates

On analysing the distance of zinc coordinating residues (Figures 5, S3 and S4) along each Zinc finger regions, it was observed that among the Zinc fingers (ZNFs), Zinc finger3 (ZNF3) [Phe396-His418] was inferred to have higher deviations and deteriorations of zinc coordinates in all the mutants than the WT. This drastic deviation at ZNF3 was due to the loss of $\alpha$ helix of ZNF3 which was transformed as loops. In Zinc finger1 (ZNF1) [Tyr338-His362], it was observed that only His362 was shown to have higher deviations than the WT in almost all mutants except for R371C. In the case of Zinc finger2 (ZNF2) [Phe368-His390], the residue His390 was found to be highly deviating in all mutants, while Cys373 in H382R, H390R, and E393V, respectively. Among the ZNF3 Zn$^{2+}$ coordinating residues, His414 and His418 were found to be completely deteriorated in all the mutants, while the Cys398 and Cys401 showed drastic deviations in all mutants except for mutants R371C and R375C than WT.

3.6. Exploring the secondary structural modifications of Egr1

The secondary structural analysis (Figure 6) showed that the $\beta$ – turn – $\beta$ (Gly364-Arg379) in the mutants R360H, R371C, R375C, and $\alpha$ helix (Ser408-Lys416) to be completely lost in all mutants. These loops transformed secondary structures are known to play a crucial role in maintaining the stability of the zinc coordinating residues of the ZNFs. Thereby the loss of the secondary structures, on being transformed as loops attains the higher flexibility, which in turn contributes to the disorientation of the zinc coordinating residues leading to the deteriorations of ZNF.

3.7. Structural rearrangement triggered by the mutation

The interaction between the Arg360 and Gln369 aids in close contact of ZNF1 and ZNF2 and also maintains its intra stability and hydrogen bond interactions between Cys370 and Arg375 which contributes to the stabilization of $\beta$ sheets that were observed to be lost in mutants R360H and R375C, respectively. Thereby loss of these interactions had a greater impact on the secondary structural loss and also the displacement of the ZNFs as shown in Figures 6 and 7. From the distance plots of ZNFs (Figure 7 and Table S2), it was inferred that the distance of ZNF1 with ZNF2 is found to be deviating by about 0.4 nm in mutant R371C and R360H while the distance of ZNF3 with ZNF1 and ZNF2 was observed to have deviations in almost all mutants specifically in H390R, H414Y mutants. The distal placement of the ZNF3 may be due to the secondary structure loss and higher flexibility of the $\alpha$-helix-3 (H3) that has been transformed as loops. Besides, it was noted that the conserved hydrophobic residues, namely Cys370, His386 which interacts with each other, and Phe377, Leu383, Cys373 on interacting with His390 to form a strong hydrophobic core that maintains the intramolecular stability and orients the zinc coordinating residues
Figure 3. Position of 7 mutations on Egr1 DBD-DNA complex.

Figure 4. Comparative MD analysis of Egr1's-DBD-DNA complex both wild (WT) and the mutants (a) RMSD, (b) RMSF, (c) Radius of gyration.
were found to be lost in H390R mutant, thereby distorting the hydrophobic environment. Moreover, except for the mutants H414Y and R371C, the other mutants were observed to have lesser intra hydrogen bonds than the WT, yet have maintained about \( \frac{1}{2} \) hydrogen bonds throughout the simulation (Figure S2b and Table S2). Thus, this infers that these mutations alter the hydrophobic core and indirectly affects the structural stability by increasing the flexibility and decreasing the intramolecular interactions, leading to the varying conformational states of the mutants than the WT.

Figure 5. Distance plot of ZNF1’s Zn\(^{2+}\) coordinating residues with Zn\(^{2+}\) 1 (Zn1) (a) Cys340, (b) Cys345, (c) His358, (d) His362; Zn\(^{2+}\) 2 (Zn2) (e) Cys370, (f) Cys373, (g) His386, (h) His390; Zn\(^{2+}\) 3 (Zn1) (i) Cys398, (j) Cys401, (k) His414, (l) His418.

Figure 6. Secondary structure analysis plot obtained from 100 ns of MD production run (a) WT, (b) R360H, (c) R371C, (d) R375C, (e) H382R, (f) H390R, (g) E393V, (h) H414Y.
From the Residue interaction network analysis (Figures 8, 9, and S5–S12; Table S3), it was inferred that all mutants to have fewer intra hydrogen bonds, whereas the van der Waals interactions were found to be highly increased in the mutants than the WT. It was also inferred that only one mutant R371C has attained the $\pi$-cation interaction, while ionic interactions have increased in mutant H382R whereas the other mutants except for R371C and R375C has lacked to form ionic interactions when compared to the WT. Apart from that, it was also observed that $\pi-\pi$ stacking interactions to be maintained only in mutants R371C, R375C as that of WT, while the other mutants have decreased $\pi-\pi$ stacking interactions. Overall from the residue-based network interaction analysis, it was inferred that apart from R371C and R375C, other mutants have a major impact in decreasing the structural stability. Moreover, on comparative structural analysis of the initial and near-native complexes (Figure S13) of wild and mutants (Figures S14 and S15), it was observed that the major conformational changes be in ZNF2 and ZNF3 in all the mutants except for the WT. In the case of ZNF2, the loss of $\beta$ sheets was found to alter the orientation of the zinc coordinating residues namely His390, Cys373, while in ZNF3 the flexibility of the loop transformed $\alpha$ helix and the sparsely deteriorated $\beta$ sheets have led to the drastic orientation changes of the Zn$^{2+}$ coordinating residues that may impact the stability of the ZNFs.

3.8. Egr1-DNA interaction analysis

On analysing the intermolecular hydrogen bond interactions (Figure S2a) between the Egr1’s DBD and DNA, it was observed that the WT showed a lesser H-bonds ($\sim$27) initially, yet has tried to increase the stability of the DNA interaction by gradually increasing the H-bonds to about $\sim$41 after 30 ns, however, the WT has maintained an average of $\sim$35 H bonds throughout the MD simulation run. On comparative analysis of the intermolecular hydrogen bonds (Figure S2a and Table S2), it was inferred that the mutants R371C and R375C to have maintained $\sim$30 H-bonds, while other mutants to have maintained lesser H bonds than that of WT of nearly $\sim$25–26 throughout the simulation. This indicates that the mutants R371C and R375C doesn’t have any major impact in altering the DNA binding and interactions with Egr1’s DBD. From the 2D interaction analysis plot of Egr1’s DBD and DNA complexes (Figures S16–S19 and Table 2), it was observed that the residues that are known to form interactions with DNA base pairs namely, ZNF1: R351, D353, R357; ZNF2: R379, D381, H382; ZNF3: R407, D409, R413, were found to be interacting in the WT and mutant complexes, except for mutant H382R. Moreover, the interactions of R413 with the base pair were found to be observed in the near-native complexes than the pre-simulated complexes. The higher incidence of the $\pi$-cation interactions was observed only in the pre-simulated complexes, contributed highly by the H382 (ZNF2), R407 (ZNF3), whereas, in the case of mutant H382R, the $\pi$-cation interaction of H382 is completely lost. In post simulated complexes, the $\pi$-cation interactions were completely lost, however, the stability of the complexes was noted to be maintained by the increased salt bridge interactions than that of the pre simulated complexes. Among these residues, R357 (ZNF1), R413 (ZNF3), and H382 (ZNF2) that are known to form stable interactions with the guanine moiety were also observed to have stable interactions in both the pre-and post-simulated complexes, whilst...
in the case of R413 the interactions were observed only in the post simulated complexes. In addition, even the electrostatic interaction calculations of Egr1-DNA (Table S4) were found to corroborate well with the interaction analysis, except for the mutants R371C and R375C, all the other mutants were found to have drastic changes in the electrostatic interaction energies when compared to the WT.

4. Conclusion

Early growth response-1 (Egr1) is a zinc-finger transcription factor that plays a critical role in many pathological conditions such as cancer, neurological disorders, and ocular disorders. As mutations occurring at the DNA binding domain of the Egr1 region deteriorates DNA recognition and may
hinder transcription, this study was implemented to determine the pathogenic and functionality affecting mutants that may potentially alter the structural stability and interactions with DNA. Overall, from the 38 missense SNPs, the resulting 7 SNPs were known to have an impact on structural stability and functionality loss. It was inferred from the MD analysis and the residue interaction network analysis only the mutant R371C and R375C have the least impact in altering the

![Figure 9](https://example.com/figure9.png)

**Figure 9.** Three-dimensional structure representation of the residue interaction network of the Egr1’s DBD domain (a) H382R, (b) H390R, (c) E393V, (d) H414Y; where the π-π stacking (Red) and Ionic interactions (Blue) as sticks, while the hydrogen bonds (Cyan) and Van der Waals (brown) as lines.
| Interaction analysis of Egr1 with DNA. | Pre-simulated complex | Hydrophobic interactions | Hydrogen bonds | π-Cation interactions | Salt bridges |
|--------------------------------------|-----------------------|--------------------------|----------------|---------------------|-------------|
| WT                                  |                       | 382His                   | 351Arg, 357Arg, 379Arg, 382His, 407Arg, 409Asp, 351Arg, 353Asp, 357Arg, 360Arg, 380Ser, 381Asp, 407Arg, 408Ser, 411Arg | 382His, 407Arg | 347Arg, 358His, 386His, 403Arg |
| R360H                               | 382His                | 351Arg, 357Arg, 379Arg, 382His, 407Arg, 409Asp, 353Asp, 380Ser, 381Asp, 408Ser, 411Arg | 382His, 407Arg | 347Arg, 358His, 386His, 403Arg |
| R371C                               | 382His                | 351Arg, 357Arg, 379Arg, 382His, 407Arg, 409Asp, 353Asp, 360Arg, 380Ser, 381Asp, 408Ser, 411Arg | 382His, 407Arg | 347Arg, 358His, 386His, 403Arg |
| R375C                               | 382His                | 351Arg, 357Arg, 379Arg, 382His, 407Arg, 409Asp, 353Asp, 360Arg, 380Ser, 381Asp, 408Ser, 411Arg | 382His, 407Arg | 347Arg, 358His, 386His, 403Arg |
| H382R                               | 382Arg                | 351Arg, 357Arg, 379Arg, 382His, 407Arg, 409Asp, 353Asp, 360Arg, 380Ser, 381Asp, 408Ser, 411Arg | 407Arg | 347Arg, 358His, 386His, 403Arg |
| H390R                               | 382His                | 351Arg, 357Arg, 379Arg, 382His, 407Arg, 409Asp, 353Asp, 360Arg, 380Ser, 381Asp, 408Ser, 411Arg | 382His, 407Arg | 347Arg, 358His, 386His, 403Arg |
| H390R                               | 382His                | 351Arg, 357Arg, 379Arg, 382His, 407Arg, 409Asp, 353Asp, 360Arg, 380Ser, 381Asp, 408Ser, 411Arg | 382His, 407Arg | 347Arg, 358His, 386His, 403Arg |
| E393V                               | 382His                | 351Arg, 357Arg, 379Arg, 382His, 407Arg, 409Asp, 353Asp, 360Arg, 380Ser, 381Asp, 408Ser, 411Arg | 382His, 407Arg | 347Arg, 358His, 386His, 403Arg |
| H414Y                               | 382His                | 351Arg, 357Arg, 379Arg, 382His, 407Arg, 409Asp, 353Asp, 360Arg, 380Ser, 381Asp, 408Ser, 411Arg | 382His, 407Arg | 347Arg, 358His, 386His, 403Arg |

| Post-simulated complex | Hydrophobic interactions | Hydrogen bonds | π-Cation interactions | Salt bridges |
|------------------------|--------------------------|----------------|---------------------|-------------|
|                       | 381Asp, 409Asp           | 347Arg, 351Arg, 357Arg, 378Ser, 390His, 394Lys, 403Arg, 360Arg, 412Lys, 413Arg, 416Lys, 418His, 420Arg |
|                       | 351Arg, 354Glu           | 347Arg, 407Arg, 413Arg, 421Gln, 353Asp, 380Ser, 381Asp, 409Asp |
|                       | 331Thr, 351Arg, 353Asp, 357Arg, 379Arg, 407Arg, 409Asp, 411Arg, 412Lys |
|                       | 350Ser, 351Arg, 354Glu, 357Arg, 379Arg, 407Arg, 413Arg, 421Gln, 330Lys, 353Asp, 380Ser, 381Asp, 409Asp |
|                       | 330lys, 336Arg, 347Arg, 347Arg, 366Lys, 420Arg, 360Arg, 388Arg, 411Arg, 412Lys |
|                       | 336Arg, 347Arg, 357Arg, 340Arg, 403Arg, 360Arg, 388Arg, 411Arg, 412Lys |
|                       | 336Arg, 347Arg, 379Arg, 382Arg, 403Arg, 360Arg, 388Arg, 411Arg, 412Lys |
|                       | 336Arg, 347Arg, 379Arg, 382Arg, 403Arg, 360Arg, 388Arg, 411Arg, 412Lys |
|                       | 336Arg, 347Arg, 379Arg, 382Arg, 403Arg, 360Arg, 388Arg, 411Arg, 412Lys |
|                       | 336Arg, 347Arg, 379Arg, 382Arg, 403Arg, 360Arg, 388Arg, 411Arg, 412Lys |
|                       | 336Arg, 347Arg, 379Arg, 382Arg, 403Arg, 360Arg, 388Arg, 411Arg, 412Lys |
structural stability than the other mutants compared to WT. Moreover, it was also inferred that the mutants R360H, H382R, H390R, E393V, and H414Y to have a greater impact in altering the structural stability and DNA interactions. Since these mutations were observed to cause the secondary structure loss, they are transformed as loops with increased flexibility and may lead to the dis-orientation of Zinc coordinating residues and also deterioration of the crucial interactions that stabilize the intra and inter stability of Egr1-DNA complexes. However, the insights from this study needs to be confirmed by further validation using biophysical methods like NMR and Cryo-EM techniques in order to confirm the applicability.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by Indian Council of Medical Research (ICMR). Ref: No. 5/4/6/08/OPH/2015-NCD-II.

References

Abbehausen, C. (2019). Zinc finger domains as therapeutic targets for metal-based compounds - an update. Metallokins: Integrated Biometalscience, 7(11), 15–28. https://doi.org/10.1039/cdm00262b

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics, 7, Unit7.20. https://doi.org/10.1002/9780471142905.hg072os76

Agrahari, A. K., Kumar, A., Shiva, R., Zayed, H., & Doss, C. G. P. (2018). Substitution impact of highly conserved arginine residue at position 75 in GJB1 gene in association with X-linked Charcot-Marie tooth disease: A computational study. Journal of Theoretical Biology, 437, 305–317. https://doi.org/10.1016/j.jtib.2017.10.028

Agrahari, A. K., Muskan, M., George Priya Doss, C., Siva, R., & Zayed, H. (2018). Computational insights of K1444N substitution in GAP-related domain of NF1 gene associated with neurofibromatosis type 1 disease: A molecular modeling and dynamics approach. Metabolic Brain Disease, 33(5), 1443–1457. https://doi.org/10.1007/s11011-018-0251-1

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics, 81(8), 3684–3690. https://doi.org/10.1063/1.448118

Bird, A. J., McCall, K., Kramer, M., Blankman, E., Winge, D. R., & Eide, D. J. (2003). Zinc fingers can act as Zn2+ sensors to regulate transcription activation domain function + sensors to regulate transcription activation domain function. The EMBO Journal, 22(19), 5137–5146. https://doi.org/10.1093/emboj/cdg484

Carman, J. A., & Monroe, J. G. (1995). The EGR1 protein contains a discrete transcriptional regulatory domain whose deletion results in a truncated protein that blocks EGR1-induced transcription. DNA and Cell Biology, 14(7), 581–589. https://doi.org/10.1080/105435895.15.581

Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G., & Raschella, G. (2017). Zinc-finger proteins in health and disease. Cell Death Discovery, 3, 17071. https://doi.org/10.1038/cddiscovery.2017.71

Crosby, S. D., Puetz, J. J., Simburger, K. S., Fahrner, T. J., & Milbrandt, J. (1991). The early response gene NGFI-C encodes a zinc finger transcriptional activator and is a member of the GCGG/GGCGG/C (GSG) element-binding protein family. Molecular and Cellular Biology, 11(8), 3835–3841. https://doi.org/10.1128/mcb.11.8.3835-3841.1991

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amede, M. R., Armean, I. M., Bennett, R., Bhai, J., Billis, K., Boddu, S., Cummins, C., Davidson, C., Dodiya, K. J., Gall, A., Girion, C. G., Gil, L., Grego, T., Haggerty, L., Haskell, E., … Flicek, P. (2019). Ensembl 2019. Nucleic Acids Research, 47(D1), D745–D751. https://doi.org/10.1093/nar/gky113

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N method for Ewald sums in large systems. The Journal of Chemical Physics, 98(12), 10089–10092. https://doi.org/10.1063/1.464397

Doncheva, N. T., Klein, K., Domingues, F. S., & Albrecht, M. (2011). Analyzing and visualizing residue networks of protein structures. Trends in Biochemical Sciences, 36(4), 179–182. https://doi.org/10.1016/j.tibs.2011.01.002

Duclot, F., & Kabaj, M. (2017). The role of early growth response 1 (EGR1) in brain plasticity and neuropsychiatric disorders. Frontiers in Behavioral Neuroscience, 11, 35. https://doi.org/10.3389/fnbeh.2017.00035

Emerson, R. O., & Thomas, J. H. (2009). Adaptive evolution in zinc finger transcription factors. PLoS Genetics, 5(1), e1000325. https://doi.org/10.1371/journal.pgen.1000325

Eom, K. S., Cheong, J. S., & Lee, S. J. (2016). Structural analyses of zinc finger domains for specific interactions with DNA. Journal of Microbiology and Biotechnology, 26(12), 2019–2029. https://doi.org/10.4014/jmb.1609.09021

Fedotova, A. A., Bonchuk, A. N., Mogila, V. A., & Georgiev, P. G. (2017). C2H2 zinc finger proteins: The largest but poorly explored family of higher eukaryotic transcription factors. Acta Naturae, 9(2), 47–58. https://doi.org/10.23667/20758251-2017-9-2-47-58

Garton, M., Najafabadi, H. S., Schmitges, F. W., Rodovani, E., Hughes, T. R., & Kim, P. M. (2015). A structural approach reveals how neighbouring C2H2 zinc fingers influence DNA binding specificity. Nucleic Acids Research, 43(19), 9147–9157. https://doi.org/10.1093/nar/gkv919

Gaslher, A., & Sukhatme, V. P. (1995). Early growth response protein 1 (Egr-1): Prototype of a zinc-finger family of transcription factors. Progress in Nucelic Acids Research and Molecular Biology, 50, 191–224. https://doi.org/10.1016/S0079-6603(08)60815-6

Godwin, R. C., Melvin, R. L., Gmeiner, W. H., & Salsbury, F. R. (2017). Binding site configurations probe the structure and dynamics of the zinc finger of NEMO (NF-kB essential modulator). Biochemistry, 56(4), 623–633. https://doi.org/10.1021/acs.biochem.6b00755

Hamilton, T. B., Borel, F., & Romaniuk, P. J. (1998). Comparison of the DNA binding characteristics of the related zinc finger proteins WT1 and EGR1. Biochemistry, 37(7), 2051–2058. https://doi.org/10.1021/bi9717993

Hashimoto, H., Olanrewaju, Y. O., Zheng, Y., Wilson, G. G., Zhang, X., & Cheng, X. (2014). Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. Genes & Development, 28(20), 2304–2313. https://doi.org/10.1101/gad.250746.114

Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry, 18(12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(19970918)19:12<1463::AID-JCC4>3.0.CO;2-H

Im, W., Beglov, D., & Roux, B. (1998). Continuum solvation model: A computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Computer Physics Communications, 11(1–3), 59–75. https://doi.org/10.1016/0010-4655(98)00016-2

Jayaraj, J. M., Krishnasamy, G., Lee, J.-K., & Muthusamy, K. (2019). In silico identification and screening of CYP24A1 inhibitors: 3D QSAR pharmacophore mapping and molecular dynamics analysis. Journal of Biomolecular Structure & Dynamics, 37(7), 1700–1714. https://doi.org/10.1080/07391102.2018.1464958

Jo, S., Vargyas, M., Vasko-Szidar, J., Roux, B., & Im, W. (2008). BPEQ-solver for online visualization of electrostatic potential of biomolecules. Nucleic Acids Research, 36(Web Server issue), W270–W275. https://doi.org/10.1093/nar/gkn314

Kemme, C. A., Loo, R. H., Chen, C., Pietka, C. C., Pettitt, B. M., & Iwahara, J. (2019). Mobility of histidine side chains analyzed with 15N NMR relaxation and cross-correlation data: Insight into zinc-finger-DNA
interactions. The Journal of Physical Chemistry B, 123(17), 3706–3710.
https://doi.org/10.1021/acs.jpcb.9b03132

Kirubakaran, P., Kothandan, G., Cho, S. J., & Muthusamy, K. (2014). Molecular insights on TNKS1/TNKS2 and inhibitor-IWR1 interactions. Molecular bioSystems, 10(2), 281–293. https://doi.org/10.1039/C3mb70355f

Klusk, K., Adamczyk, J., & Krézel, A. (2018). Metal binding properties, stability and reactivity of zinc fingers. Coordination Chemistry Reviews, 367, 18–64. https://doi.org/10.1016/j.ccr.2018.04.009

Li, B., Krishnan, V. G., Mort, M. E., Xin, F., Kamati, K. K., Cooper, D. N., Mooney, S. D., & Radijojac, P. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics (Oxford, England), 25(21), 2744–2750. https://doi.org/10.1093/bioinformatics/btp528

Mooney, S. D., & Radivojac, P. (2009). Automated inference of molecular and systems neuroscience. Journal of Computer and Biophysics, 13(10), 13104–13117. https://doi.org/10.1038/jnms.131013104

Pagel, J.-I., & Deindl, E. (2012). Disease progression mediated by EGR-1 and other metal-containing proteins and interfere with immunoglobulin transactivation. Journal of Inorganic Biochemistry, 116, 90–99. https://doi.org/10.1016/j.jinbio.2011.10.018

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. Journal of Computer-Aided Molecular Design, 27(3), 221–234. https://doi.org/10.1007/s10822-013-4644-8

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. Nucleic Acids Research, 29(1), 308–311. https://doi.org/10.1093/nar/29.1.308

Xiao, Z., Guo, L., Zhang, Y., Cui, L., Dai, Y., Lan, Z., Zhang, Q., Wang, S., Liu, W. (2020). Structural analysis of missense mutations occurring in the DNA-binding domain of HSF4 associated with congenital cataract. Biophysical Journal, 129(5), 52465–52472. https://doi.org/10.1017/jbc.M403028200

O’Donovan, K. J., Tourtellotte, W. G., Millbradt, J., & Baraban, J. M. (1999). The EGF family of transcription-regulatory factors: Progress at the interface of molecular and systems neuroscience. Trends in Neurosciences, 22(4), 167–173. https://doi.org/10.1016/S0166-2236(98)01343-5

Pace, N. J., & Weerapan, E. (2014). Zinc-binding cytoxines: Diverse functions and structural motifs. Biomolecules, 4(2), 419–434. https://doi.org/10.3390/biom4020419

Pagel, J.-I., & Deindl, E. (2012). Disease progression mediated by EGR-1 associated signaling in response to oxidative stress. International Journal of Molecular Sciences, 13(10), 13104–13117. https://doi.org/10.3390/ijms131013104

Pérez, A., Marchán, L., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. Biophysical Journal, 92(11), 3817–3829. https://doi.org/10.1529/biophys.106.097782

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera-A visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Piovesan, D., Minervini, G., & Tosatto, S. C. E. (2016). The RNG 2.0 web server for high quality residue interaction networks. Nucleic Acids Research, 44(W1), W367–W374. https://doi.org/10.1093/nar/gkw315

Ravasi, T., Huber, T., Zavolan, M., Forrest, A. Gaasterland, T., Grimmond, S., & Hume, D. A. (2003). Systematic characterization of the zinc-finger-containing proteins in the mouse transcriptome. Genome Research, 13(68), 1430–1442. https://doi.org/10.1101/gr.949803

Russo, M. W., Matheny, C., & Milbrandt, J. (1993). Transcriptional activity of the zinc finger protein NGFI-A is influenced by its interaction with a cellular factor. Molecular and Cellular Biology, 13(11), 6858–6865. https://doi.org/10.1128/mcb.13.11.6858-6865.1993

Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F., & Schroeder, M. (2015). PLIP: Fully automated protein-ligand interaction profiler. Nucleic Acids Research, 43(W1), W443–W447. https://doi.org/10.1093/nar/gkw315

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. Nucleic Acids Research, 29(1), 308–311. https://doi.org/10.1093/nar/29.1.308

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. Nucleic Acids Research, 40(Web Server issue), W452–W457. https://doi.org/10.1093/nar/gks539

Tang, H., & Thomas, P. D. (2016). PANTHER-PSEP: Predicting disease-causing genetic variants using position-specific evolutionary preservation. Bioinformatics (Oxford, England), 32(14), 2230–2232. https://doi.org/10.1093/bioinformatics/btw222

Végrán, F., Reubci, M., Chevrier, S., Cadout, M., Boidot, R., & Lizard-Nicol, S. (2013). Only missense mutations affecting the DNA binding domain of p53 influence outcomes in patients with breast carcinoma. PloS One, 8(1), e55103. https://doi.org/10.1371/journal.pone.0055103

Wolfe, S. A., Nekludova, L., & Pabo, C. O. (2000). DNA recognition by Cys2His2 zinc finger proteins. Annual Review of Biophysics and Biomolecular Structure, 29, 183–212. https://doi.org/10.1146/annurev.biophys.29.1.183

Xiao, Z., Guo, L., Zhang, Y., Cui, L., Dai, Y., Lan, Z., Zhang, Q., Wang, S., Liu, W. (2020). Structural analysis of missense mutations occurring in the DNA-binding domain of HS4 associated with congenital cataracts. Journal of Structural Biology, 194, 100015.

Yella, V. R., Bhimsaria, D., Ghoshdastidar, D., Rodriguez-Martinéz, J. A., Ansari, A. Z., & Bansal, M. (2018). Flexibility and structure of flanking DNA impact transcription factor affinity for its core motif. Nucleic Acids Research, 46(22), 11883–11897. https://doi.org/10.1093/nar/gky1057

Yoneda, Y., Saito, H., Toyama, M., Makita, Y., Miyamoto, A., Hamada, K., Kurotaki, N., Tomita, H., Nishiyama, K., Tsurusaki, Y., Doi, H., Miyake, N., Ogata, K., Naritomi, K., & Matsumoto, N. (2012). Missense mutations in the DNA-binding/dimerization domain of NFIX cause Sotos-like features. Journal of Human Genetics, 57(3), 207–211. https://doi.org/10.3838/jhge.2012.7