

# Prediction of Deleterious nsSNPs Causing CHARGE Syndrome Associated with the CHD7 Protein using Computational Approaches

**Nithya S. Rathinam**

Sri Ramachandra Institute of Higher Education and Research

**Madhana Priya N**

Sri Ramachandra Institute of Higher Education and Research

**Magesh Ramasamy** (✉ [magesh.r@sriramachandra.edu.in](mailto:magesh.r@sriramachandra.edu.in))

Sri Ramachandra Institute of Higher Education and Research

---

## Research Article

**Keywords:** CHARGE Syndrome, nsSNPs, Mutational changes, Sequence analysis, Structural analysis, Molecular Dynamics Simulation, Principal Component Analysis, Molecular Docking Studies.

**Posted Date:** February 26th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-233256/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The Chromo domain helicase DNA binding protein 7 (CHD7) is also known as ATP-dependent helicase CHD7, in humans, the *CHD7* gene encodes it. Heterozygous mutations in this protein cause aggregation and has been determined to have an adverse role in causing CHARGE syndrome. Non-synonymous single nucleotide polymorphism (nsSNP) analysis tends to be deleterious of the protein yet to be employed with computational methods though being the highlight for novel investigations. Various computational methods were used to categorize the 201 identified nsSNPs in the catalytic domain of the CHD7 protein (the nsSNPs are observed to have a damaging effect in the catalytic domain). Pathogenicity analysis determined 81 nsSNPs to be pathogenic and further narrowed down to 61 nsSNPs by stability analysis. Based on the structure availability, the two nsSNPs (P2683S and R2702C) were selected and were checked in the computational tools for sequence analysis (pathogenicity analysis, stability analysis, physiochemical property analysis, and conservational analysis) and were determined to have a high impact over the protein molecule. The molecular dynamics simulation and principal component analysis (PCA) were performed to determine the conformational stability and flexibility change of the proteins. Subsequently, a molecular dynamic simulation (MDS) for 100ns was performed to understand the impact of the differences between the native and the mutant structures of the CHD7 protein. The simulation plots disclose very minute changes in patterns of stability, residue fluctuation, structure compactness, and flexibility regarding P2683S and R2702C mutation compared to the native structure. Further, Molecular docking was performed for the native and the mutant structures P2683S and R2702C to study the binding efficacy of the drugs Methyltestosterone and Estradiol resulting in a similar score with a very little difference to each other. The Native and mutants P2683S and R2702C have similar interaction of -5.7 kcal/mol, -5.9 kcal/mol and - 5.6 kcal/mol respectively with Methyltestosterone followed by a binding score of -6 kcal/mol, -5.6 kcal/mol and - 5.8 kcal/mol respectively for Estradiol. Detailed study about the disease, effect of nsSNP's and the response of the drug towards the mutation are the key factors in order to launch a new personalized medicine. Therefore, in this study using various computational prediction methods, molecular dynamics simulation and molecular docking studies we have determined the nsSNP's responsible to cause CHARGE syndrome and the drug response with respect to the determined nsSNP mutations. The outcomes acquired from our investigation will provide the data for experimental biologists for the additional procedure for examining the rest of the variations in CDH7 protein.

## Introduction

The first scientific report on CHARGE syndrome as an autosomal dominant case was recorded in 1981.<sup>1</sup> It was named CHARGE syndrome for its seven major clinical features, termed as Coloboma of the eye, Heart defects, Atresia of the nasal choanae, Retardation of growth, Genital/urinary abnormalities, Ear abnormalities, and deafness.<sup>2</sup> It includes the following symptoms: Anosmia (lost smell), Aplasia/Hypoplasia of the earlobes (absence/ small ear lobes), Cryptorchidism (Undescended testes), and Anophthalmia (absence of eyeballs).<sup>3</sup> CHARGE syndrome often results in developmental delays.<sup>4</sup> It is a complex hereditary disorder along with the extensive scope of tissues/frameworks influenced by

different mutations of CHD7 protein.<sup>5</sup> It is an unusual kind of disorder with a predominance of around one in 10,000 in the wide-ranging population. In a nationwide study, its existence was assessed to be 1/8500, in North America with the previous reports of 1/12 500. On the other hand, in Europe, it was evidenced to be 1/110 000 births.<sup>6</sup> The most elevated CHARGE syndrome frequency was evaluated in Canada, in the proportion of 1:8,500 live births.<sup>7</sup> Heart defects are most likely to arise in around 75% of patients with pathogenic CHD7 mutation.<sup>8</sup> Further CHD7 mutations that were found were reported, it was identified that around 72% of the mutations are nonsense mutation, 13% of the mutations are reported to be splice site, and about 10% of the mutations are detected to be missense mutation. These mutations can get uncovered and experience structural misfolding and, in this way, illustrate structural and functional abnormalities. For CHD7 mutations, a precise determination of genotype-phenotype connections has not yet been distinguished, even among CHD7 mutation indistinguishable patients.<sup>3,5,9</sup> In the larger part of the cases, the mutations in CHD7 are anticipated to have a functional loss, most likely prompting to target an unusual mRNA for corruption through mediated decay. Subsequently, it shows that the haploinsufficiency for CHD7 is the major pathogenic component underlining the reason for CHARGE disorder.<sup>10,11</sup> The novel mutation in protein CHD7 triggers to cause CHARGE syndrome, and in some cases, it occurs due to the genomic alterations in chromosome 8.<sup>12</sup>

The *CHD7* gene is positioned at chromosome 8 (8q12) with 61.59 Mb ranging from its p-arm telomere, with a genetic magnitude of 188 kb, it is made up of 38 exons. The heterozygous mutations and deletions may also contribute CHARGE syndrome. Presently, the *CHD7* gene is known to be related to CHARGE disorder.<sup>13</sup> The creation of chromodomain helicase DNA binding protein. The responsible gene provides the guidelines for the creation of chromodomain helicase DNA binding protein.<sup>14</sup> This protein is usually originated before birth in many locales of the body, including the eye, inner ear, and brain.<sup>15</sup> A hefty portion of the changes in the *CHD7* gene leads to abnormal CHD7 protein construction that is fragmented down.<sup>16</sup>

Genetic polymorphisms exist in the human genome. The most general sort of polymorphism to be known is SNP, which arises in recurrence around 1 of each 300 nucleotide base pairs, and more than around 10 million in the general populace. Divergence can appear in coding and non-coding area of the genes and likewise affect the structural organization and function of the protein implied by a specific gene, for the most part in the cases when the polymorphism prompts the replacement of an amino acid in the conserved functional district of the protein.<sup>17-20</sup> The stability loss of the proteins is one of the other significant reasons for the cause of disease. The proteins are imperceptibly stable; in this way, even a small impact on the stability may change protein's thermodynamic harmony to turn the folded state unstable. Brought down by stability in a protein prompts a decrease of protein's effective concentration, which further prompts to reduce its capacity to perform its biochemical function.<sup>21</sup>

In this study, the structural changes between the native protein and the mutated proteins were identified, analysed, and visualized to understand the highly Deleterious nsSNPs that have the highest probability to cause CHARGE syndrome.

# Materials And Methods

## Information Compilation

Human CHD7 protein data information of the nsSNPs from the NCBI, UniProt, and HGMD were gathered.<sup>22-24</sup> From Protein Data Bank (PDB), the 3D structure of UniProt ID Q9P2D1 with the PDB ID 2VOF was acquired.<sup>25</sup>

## Sequence Investigation

PredictSNP and iStable tools were used to study the pathogenicity and stability of the nsSNPs, respectively. Computational tools such as PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, and SNAP were used to determine the pathogenicity of nsSNPs to anticipate pathogenicity of mutations as Deleterious or Neutral individually.<sup>26,27</sup> The stability of the nsSNPs was determined using iStable, which has an enclosed procedure of iStable, I-Mutant, and MUpro.<sup>28</sup> The FASTA sequence retrieved from the public database UniProt and the nsSNPs were submitted as input. The prediction of the missense substitution ranges from enhanced Deleterious to enhance Neutral. Align-GVGD depends on the biophysical properties of amino acids and protein's multiple sequence alignments.<sup>29</sup> When an estimation of GV = 0 represents invariant residue in the alignment, a GV score between 60 and 65 is the conservative upper limit. While the estimation of GD = 0 compares a missense substitution present at the invariant position, a GD score somewhere in the range of 60 and 65 is the most elevated conserved missense substitution, and GD > 100 represents radical replacement.<sup>30</sup>

## Structural Investigation

The Protein Data Bank (PDB; <http://www.rcsb.org/pdb/>) is the solitary collection of structural information and three-dimensional structures of biological macromolecules.<sup>31,32</sup> The mutation analysis was performed with the outcomes acquired, employing computational tools utilizing Swiss-PdbViewer.<sup>33</sup> Dependent on the values of RMSD, the deviance among the native structure and mutant structure was investigated.

## Evolutionary Investigation

Prediction of the sequence conservation becomes an essential factor for understanding the effect of the mutation. The nsSNPs that fall under the conserved regions of the protein were determined using the ConSurf web server. Any mutation observed in the conserved districts of the ConSurf server may have an increasingly Deleterious outcome on the structural or functional properties.<sup>34</sup> The conservation rankings are determined with the utilization of ConSurf shading codes, ranging from 1–9, where 1 indicated the top speedy developing positions, five demonstrate the places of transitional rates, and 9 indicates the leading conserved positions of the protein.

## Neighbouring Amino Acid Changes Prediction

Neighbouring amino acid changes exploration from the mutational point would evoke the conformational variations in the mutant structures. Visualization of the structures via Discovery Studio was performed within the range of 4 Å from the substituted amino acid, close residue examination for the fluctuated native and mutant structures.<sup>35</sup>

## **Molecular Dynamics Simulations**

The native and mutant proteins were cleaned, and the GROMOS96 43a1 force field was applied to generate topology for the MDS using GROMACS. The topology of the structure was built, and a cubic water box was built around the protein. Ions were added based on the charge to neutralize the entire setup. After the system was set, Energy minimization was carried out for 100ps. Once the minimal energy was achieved, the model was equilibrated with an Isothermal-isochoric ensemble (NVT) and Isothermal-isobaric ensemble (NPT) analysis for 100ps each. The entire model was exposed to MD simulation for 100 ns. After the MDS, the trajectories were analyzed, and each structure was analyzed for its RMSD, RMSF, Gyration, HBond, and SASA by implementing the `gmx rmsd`, `gmx rmsf`, `gmx gyrate`, `gmx hbond`, and `gmx sasa` commands, respectively. Both the mutant and the native structures were simulated with a similar protocol to realize how the mutation can change the protein structure and recognize protein conformation in a dynamic state.<sup>36,37</sup>

## **Principal Component Analysis**

Principal Component Analysis (PCA) was done to understand the conformational stability of the protein structure. PCA is a statistical technique that operates on variance and covariance matrix by extracting the protein MD-trajectory's important motions. PCA involves a mathematical procedure that transforms a several (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The mathematical technique used in PCA is called eigenanalysis: eigenvalues and eigenvectors. Diagonalisation of the covariance matrix is the corresponding eigenvectors and eigenvalues. Eigenvectors are also called the principal or essential modes. PCA is based on the computation of the elements of the positional fluctuation covariance matrix of the protein backbone. The variance and covariance matrix were created using the `gmx covar` module. The changes imposed and most prominent motions in the protein systems were obtained using principal component analysis employing the `gmx ana eig` module in gromacs. The eigenvectors and eigenvalues are given as input files and return RMS fluctuation per atom of selected eigenvectors. The first five, tenth and twentieth projections of the protein trajectory were plotted for cosine content with `xmgrace` against the dynamics time period of 100ns.<sup>31,38</sup>

## **Docking of the Native and Deleterious Mutant Structures**

The structure 2V0F along with P2683S and R2702C mutations were docked with the drugs Methyl testosterone and Estradiol retrieved from the PubChem database. Molecular docking study was performed to identify the functional changes and also the activity of the drug while binding with the native and mutant structures. The structures were energy minimized to make the proteins stable. Polar

hydrogens were added along with Kollman and AD4 type charges to the proteins and the Grid was set at the identified Binding Site. Binding Site was predicted using CastP online tool (<http://sts.bioe.uic.edu/castp/index.html?2011>). The drugs were docked at the binding site for three times consecutively and the average was taken using AutoDock tool. AutoDock is a freely available software used for computational docking and virtual screening of small molecules and macromolecular receptors. The docked Complexes were visualized using PyMol Tool.<sup>39</sup>

## Results

### Dataset Collection

A sum of around 201 SNPs was collected for CHD7 protein; then, their Deleterious natures were analyzed by numerous computational techniques. After removing the overlapping nsSNPs, the number of 201 nsSNPs of CHD7 mutations was retrieved via freely available databases like NCBI, UniProt, and HGMD. (Table 1 and Supplementary Table 1)

Table 1  
Datasets retrieved from various databases for CHD7 protein

Protein Name	UniProt ID	AA Length	NCBI Protein Accession Number	SNP Functional Class Information			
				NCBI	UniProt	HGMD	Total nsSNP's
CHD7	Q9P2D1	2997	NP_060250.2	97	76	28	201

### Pathogenicity Estimation Approaches

The impacts of SNPs tangled in the protein functioning were verified utilizing the computational tools, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, SNAP, anticipating 81 nsSNPs out of 201 nsSNPs to be highly pathogenic. (Table 2, Supplementary Table 2 and Supplementary Fig. 1).

Table 2  
Determination of the Deleterious nsSNP's in pathogenicity prediction tools

PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP	Deleterious nsSNP's present in all tools
75	93	127	127	78	81

### Stability and Physiochemical Property Prediction

The stability prediction and physiochemical property prediction was performed for the highly pathogenic 81 nsSNPs. iStable tool was used to determine the stability of the nsSNPs, which incorporates I-Mutant, MUpro, and iStable. It was found that out of 81 nsSNPs, 61 nsSNPs were determined to be highly

unstable in their protein structures. (Table 3, **Supplementary Table 3 and Supplementary Fig. 2**). The align-GVGD tool was used to perform physicochemical property prediction; from the retrieved results, it was shown that 69 nsSNPs out of the 81 nsSNPs pursue missense substitutions in the protein of interest (**Supplementary Table 4**).

Table 3  
Determination of the unstable nsSNP's after stability prediction

Stability prediction tools	I-Mutant	MUpro	iStable	Unstable nsSNP's present in all tools
No. of nsSNP's with decreasing stability	55	15	61	61

### Structure Retrieval

Based on the 3D structure availability, we have 3 NMR structures available for the CHD7 protein in Protein Data Bank (Table 4). Based on the 3D structure 2VOF with the mutational position 2631–2715, 2 highly Deleterious nsSNPs (P2683S and R2702C) are selected, which was predicted to be deleterious and unstable from pathogenicity analysis, stability analysis, and Align – GVGD prediction (Table 5a-c) The structure contains only one chain namely A chain, with the resolution of 1.80 Å. Therefore, we retrieved the 3D structure of the protein ID 2VOF from Protein Data Bank (PDB) and analyzed it for further studies.

Table 4  
Table representing the structures available for CHD7 protein

PDB ID	Positions
2CKC	2563–2622
2V0E	2561–2614
2V0F	2631–2715

Table 5  
a: Pathogenicity analysis

rsID/Variant	nsSNP's	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
201319489	P2683S	-	D	D	D	D
VAR_068155	R2702C	D	-	D	D	D
(D indicates Deleterious)						

Table 5  
b: Stability analysis

rsID /Variant	nsSNP's	I-Mutant	MUpro	iStable
201319489	P2683S	Inc	Null	Inc
VAR_068155	R2702C	Dec	Null	Dec

Table 5  
c Align – GVGD prediction of selected nsSNP's

rsID /Variant	nsSNP's	GV	GD	Prediction
201319489	P2683S	0.00	73.35	Class C65
VAR_068155	R2702C	0.00	179.53	Class C65

### Protein Conservation Prediction

When considering the two mutant positions, P2683S and R2702C, we witnessed that these two mutant positions are present in the extremely conserved regions by a score of 8 and 9. The changes in the conserved region will lead to a substantial-high impression on structural and functional alternations to cause the disease conditions in the individuals and considered to have a significant characteristic in functional Consequence of the protein. (Fig. 1).

### Structural Investigation

The structural examination was completed for the two extremely Deleterious nsSNPs (P2683S and R2702C). The native structure (PDB ID: 2V0F) was taken as the template and mutated through the support of the Swiss-PdbViewer. The energy was minimized for native, and two mutant assemblies by the same Swiss-PdbViewer and their values were determined to be 2738.306 KJ/mol for native structure 2V0F, -2528.754 KJ/mol for the mutant structure P2683S, and – 2193.724 KJ/mol for the mutant structure R2702C. From the result, it was shown that all the mutant structures obtained an increased energy value when compared with native structures. Since the mutant structures gained more energy, it will show a Deleterious impression on the structure and function of the mutant proteins.

The RMSD values were calculated for the native structure, and both the mutant structures; the resulting RMSD values for native structure 2V0F were 0nm, for the mutant structure P2683S was 0.224nm, and for R2702C was 0.218nm. All the mutant structures had relatively identical RMSD standards. The higher the RMSD standards, the more determined will be a deviation amongst the native and mutant kind proteins. Hence in RMSD analysis, the substituted amino acids in both mutant structures were highly deviated from the native amino acid and having the RMSD value > 0 Nanometer. These results suggest that the substituted amino acids will disturb the structure and function of the mutant proteins.

### Superimposed structure of nsSNPs

The superimposed structural investigation was completed for a native and mutant amino acid with PyMOL software.<sup>39</sup> The superimposed structures of P2683S and R2702C nsSNPs. In both the images,

green color stick model represents the native amino acid, and the blue color represents the substituted Deleterious amino acid. From the image, it was observed that the substituted amino acids of the nsSNP P2683S were slightly deviated from the native structure, whereas the nsSNP R2702C shows a high deviation from the native structure. From this analysis, it was observed that R2702C have much impression on protein structure and function. (Fig. 2)

### Neighbouring Amino acid Changes Prediction

P2683 possessed 12 neighbouring amino acids in its native structure and number of a surrounding amino acid are reduced to 11 in its mutant structure P2683S. Native R2702 possessed 14 surrounding amino acids, and the R2702C mutant structure has reduced to 12 amino acids shown in Table 6 (Supplementary Fig. 3). The gain and loss/damage of conventional hydrogen bonds were also observed in the case of P2683S and R2702C mutants, respectively.

In P2683S, there is an addition of 2 Hydrogen bonds in the mutant structure suggests that this mutation has less impact on the stability of the protein even though they lost one surrounding amino acid. Whereas in R2702C, both the surrounding amino acids and hydrogen bonds were lost accordingly 2 and 1, which indicates that this mutation has a very high Deleterious impact on the protein functioning. (Table 6).

Table 6

Surrounding amino acid change with presence and absence of hydrogen bond around the substituted position.

nsSNP's	No. of surrounding amino acids	No. of conventional hydrogen Bond
P2683S	VAL, ALA, VAL, LEU, TRP, THR, ASP, GLY, ASP, ARG, ALA, ASP	THR, ASP
	VAL, ALA, VAL, LEU, TRP, THR, ASP, GLY, ASP, ALA, ASP	THR, ASP, ASP, GLY
R2702C	GLU, PRO, ASN, LEU, LEU, ALA, VAL, LEU, PHE, ASP, PHE, MET, ASP, LEU	LEU, VAL, MET
	LEU, LEU, ALA, VAL, LEU, PHE, ASP, PHE, MET, ASP, LYS, LEU	VAL, MET

### Amino acid property change analysis

Changes in the amino acid property may lead to a structural change. A protein with a changed structure may directly correlate with the stability of the protein.<sup>40</sup> In the circumstance of P2683S, the changes observed are Hydrophilic to Neutral, which indicates no effect over the protein due to mutation, the side

chain flexibility changes from Restricted to low, which indicates a slight impact over the protein structure. In interaction modes the H-bonds are added in the mutant structure leading to the increase of the conventional H-bonds which in turn helps in the sustenance of the stability of the protein molecule. Whereas in the case of R2702C, the changes observed are Hydrophilic to Hydrophobic, the side chain flexibility changes from High to Low, which has a vast impression on the protein construction once after the mutation. In the interaction modes, the mutant structure lacks H-bonds when related to the native structure, directly affecting the stability of the molecule (Table 7).

Table 7  
Native and Mutant Amino acid property analysis

Variant ID/ rsID	Missense Mutation	Property changes	Side chain flexibility	Interaction modes
201319489	P2683S	Hydrophilic to Neutral	Restricted to Low	van der Waals to H-bonds
VAR_068155	R2702C	Hydrophilic to Hydrophobic	High to Low	Ionic, H-bonds, van der Waals to Covalent disulfide bonds, van der Waals

## Molecular Dynamics

The Root Mean Square Deviation (RMSD) was run for the time period of 100ns. Based on the outcomes obtained, it is concluded that the native shows a higher deviation than the mutant structures. Therefore, it results that, native has more significant structural changes than that of the mutants. The native oscillates high throughout the simulation time and was not stable with a high RMSD of ~ 0.56nm, whereas mutant P2683S was stable and also had the least RMSD of ~ 0.35nm, followed by the R2702C, which was stable till 100ns and oscillated slightly and again became stable with higher RMSD of ~ 0.45nm after the 45th ns. This higher structural change observed in the P2683S compared to the native structure makes it more pathogenic and stable (Fig. 3), which correlates with our pathogenicity, stability, and structural analysis results.

The Radius of Gyration is used to determine the compactness changes (Structural flexibility) of the protein due to mutation. Lower the radius of gyration value indicates the more compact (best) protein. The native structure was found to have the highest radius of gyration, which indicates that the native structure is less compact. The nsSNP R2702C, which showed a lower Rg value, also shows compactness (Fig. 4).

The determination of the deviation in hydrogen bond is used to measure the changes in the stability due to mutation. The native structure showed the least hydrogen bond changes, and mutant nsSNPs R2702C and P26823 showed more hydrogen bond changes throughout the time period of 100ns (Fig. 5).

The Solvent Accessible Surface Area (SASA) is used to determine the changes in the protein compactness (Structural flexibility) due to the change. The native structure was found to have a high SASA value, which indicates that the native structure has the least compactness. The mutant structure P2682S showed a higher SASA value, which signifies the least compactness compared to the R2702C mutant structure and native structure (Fig. 6).

Root Mean Square Fluctuation (RMSF) values indicate greater flexibility during the MD simulation. Based on the results, we conclude that the nsSNP P2683S shows higher fluctuation than the native structure compared to R2702C. Therefore, from the result, it is found that the nsSNP P2683S has the flexibility compared to the native structure and the nsSNP R2702C. This higher flexibility change observed in the P2683S compared to the native structure makes it more pathogenic and reduced stability (Fig. 7) and this correlates with our pathogenicity, stability, and structural analysis results.

### **Principal Component Analysis**

Principal component analysis (PCA) was done to understand the structural and conformational variations of the native and mutant proteins. The first step involved the generation of the covariance matrix of the protein backbone. The top eigenvectors are responsible for approximately 90% of the internal motions of the protein. Therefore, the first, fifth, tenth and the twentieth projections from the trajectory were retrieved and projected onto the eigenvectors against time during the 100ns simulation period (Fig. 8a-c).

The mutant system trajectories were projected over the apo form to identify the influence of the mutations over the protein. The results indicate the existence of multiple conformations of the proteins in the solution, and an equilibrium shift mechanism is induced upon mutation (Fig. 9). The mutant proteins were found to have a vast conformational space followed by the native. This conformational space directly signifies the flexibility of proteins, i.e., the mutants are more flexible than native.

### **Interaction of Drug with the native and deleterious mutants**

Methyltestosterone and Estradiol are most widely used in the treatment of hypogonadism in men and women respectively.<sup>41,42</sup> Therefore, Methyltestosterone and Estradiol were used to identify their binding affinity with the native and the deleterious & unstable nsSNPs (P2683S & R2702C) structures with the aid of molecular docking studies. The drug fits into the binding site of the CHD7 protein. The binding site was identified using CastP tool.<sup>43</sup> The details of the identified Binding site are given in Table 8.

Table 8  
Binding Site of CHD7 Protein

Volume of the binding site	Surface Area	Residues
1850.926 SA	3576.052 SA	GLY2629, SER2630, ARG2631, ASN2632, PRO2633, ASN2634, LEU2636, ASP2637, ILE2638, ASN2639, THR2640, LEU2641, THR2642, GLY2643, GLU2644, GLU2645, ARG2646, VAL2647, PRO2648, VAL2649, VAL2650, ASN2651, LYS2652, ARG2653, ASN2654, GLY2655, LYS2656, LYS2657, MET2658, GLY2659, GLY2660, ALA2661, MET2662, ALA2663, PRO2664, PRO2665, MET2666, LYS2667, ASP2668, LEU2669, PRO2670, ARG2671, TRP2672, LEU2673, GLU2674, GLU2675, ASN2676, PRO2677, GLU2678, PHE2679, ALA2680, VAL2681, ALA2682, PRO2683, ASP2684, TRP2685, THR2686, ASP2687, ILE2688, VAL2689, LYS2690, GLN2691, SER2692, GLY2693, PHE2694, VAL2695, PRO2696, GLU2697, SER2698, MET2699, PHE2700, ASP2701, ARG2702, LEU2703, LEU2704, THR2705, GLY2706, PRO2707, VAL2708, VAL2709, ARG2710, GLY2711, GLU2712

The Native interacts with an average binding score of -5.7 kcal/mol, trailed by 2 mutant structures with the average score - 5.9kcal/mol and - 5.6 kcal/mol for P2683S and R2702C respectively with Methyltestosterone (**Fig. 10a-c**). The native, P2683S and R2702C structures showed a binding score of -6 kcal/mol, -5.6 kcal/mol and - 5.8 kcal/mol respectively (**Fig. 11a-c**) for Estradiol. The interaction of the three structures with both the drug showed similar score with a very little difference to each other.

## Discussion

CHARGE syndrome is a rare genetic condition characterized by autosomal dominant clinical features including Coloboma, Heart malformations, Atresia of the choanae, Retardation of growth or development, Genital anomalies, and Ear malformations in an individual. Pathogenic and unstable nsSNPs initiated in the CHD7 protein are the most crucial cause of CHARGE syndrome, which have been acknowledged in about 70–90% of CHARGE syndrome patients. CHD7 gene was discovered to be the leading cause of CHARGE syndrome since 2004. The CHD7 gene encodes the Chromodomain Helicase DNA binding protein; mutation in this protein is responsible for causing CHARGE syndrome. Additionally, the determination of the CHARGE syndrome-related nsSNPs is a vital criterion to identify the primary cause of developing the disorder.<sup>44</sup> The SNP classifications from the various datasets have proved to be an effective method to understand the pathological impact over the rare disorders.<sup>28</sup> Determining the nsSNPs responsible for causing several diseases is more accurate by employing computational tools (in silico tools, sequence analysis, structural analysis, molecular dynamic simulation, and principal component analysis).<sup>45–48</sup>

Single Nucleotide Polymorphism (SNPs) are the single base substitution found in specific positions affecting the coding regions of the human genome, which might cause an alteration in the structural properties of the protein leading to the functional changes responsible for causing certain diseases.<sup>49</sup>

SNPs have a vital contribution to the progression of the disease. The genetic changes caused due to the SNPs exist in > 1% of the overall population.<sup>50</sup> The evolution of molecular sequencing and number of SNPs have been stored consistently in the public databases.<sup>51</sup> Utilization of various computational tools could dependably expand the predictions of nsSNP pathogenicity. Analysing these SNPs through experimental methods might be challenging and time-consuming; nevertheless, the computational analysis of the nsSNPs is proved to improve the functional characterization of the nsSNPs that would provide massive support in the development of personalized medicine.<sup>52</sup> The classification and the diagnosis of genetic diseases are expected to be enhanced using computational tools. In this current study, the main objective is to determine the most Deleterious nsSNPs (P2683S and R2702C) that have a high possibility to cause CHARGE syndrome by employing various existing computational tools.

Bartels et al. reported an investigation report regarding the mutations in CHD7 detected by a clinical laboratory. CHD7 mutations were recognized in 203 of 642 (~32%) cases alluded to GeneDx for clinical testing. A sum of variants determined from cases that were known to be novel or eluded organization. In furthermost cases, the missense mutations are hard to categorize, the inheritance of a novel missense mutation is difficult to detect if the parent is mosaic for a pathogenic transformation or benign polymorphism. The nsSNPs P2683S and R2702C were determined in the clinical assessment of the parent. The protein variation of P2683S situated in the BRK domain from CHD7 was identified in the conserved region of a parent, and R2702C was determined to be identified in the conserved region of another parent.<sup>40,53</sup> The presence of a mutation in the conserved region indicates the high possibility to cause CHARGE syndrome.

The number of 201 nsSNPs of CHD7 protein was collected from the freely available databases (NCBI, UniProt, and HGMD) and were exposed to various computational tools to analyze the pathogenicity and stability of the nsSNPs.<sup>27</sup> In the sequence-based analysis, the nsSNPs were analyzed to predict their functional effect over the protein molecule. Out of 201 nsSNPs, 81 nsSNPs were highly Deleterious by all the 5 tools used (PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, SNAP). Then 81 nsSNPs were subjected to stability prediction tools (iStable, I-Mutant, and MUpro) to determine the unstable nsSNPs, and it was shown that 61 nsSNPs have a decrease in the stability. Based on sequence analysis, 61 nsSNPs were predicted to be highly Deleterious, unstable, and certainly selected for the structural analysis.

In the structure-based analysis, crystallographic 3D structure (PDB ID: 2VOF) was available for the two nsSNPs P2683S and R2702C, and they had pathogenicity impact in all the 5 tools and stability impact ranged between 0 and 2 out of 3 tools used. The amino acid present in the conserved region of the protein is significantly important both functionally and biologically. If there is a nsSNP change in these amino acids it might lead to a high-risk and contribute to the structural and functional changes in the protein molecule. These two nsSNPs, P2683S, and R2702C positions were analyzed for their conservation analysis using ConSurf server, and we found that both the mutations fall in highly conserved regions 8 and 9, respectively, of the protein sequence, it shows that both the positions have a functional effect over the protein. Further structural differences between the native and mutant protein

structure done through Swiss-PdbViewer<sup>54</sup> and was analyzed using PyMOL software, which predicts that R2702C nsSNP has highly deviated from the native structure when compared to the nsSNP P2683S. Changes in surrounding amino acids and amino acid properties between the native and the mutants were determined to detect any change in the stability/function.<sup>40,55</sup> From this analysis, we predict that nsSNP R2702C shows the difference in surrounding amino acids (lack of hydrogen bond) and amino acid properties (hydrophilic to hydrophobic) when compared to the native structure (R2702), which directly affects the stability of the protein structure and This indicates that the nsSNP R2702C has a vast impression on protein structure and function.

Moreover, the molecular dynamics simulation was performed for 100ns to recognize the structural and functional changes at the secondary and tertiary level (dynamic conformational changes, stability changes, flexibility changes, and the difference of the structural compactness) between the native and the mutant (P2683S and R2702C) structures.<sup>36</sup>

In contrast to all the above results, the Molecular Dynamics Simulation results state that nsSNP P2683S is highly Deleterious, highly stable, and possessed high functional effects over the native protein. The native protein deviated in all analysis showing very low stability and compactness compared to the nsSNP structures. Molecular dynamics demonstrates to be an improved procedure for the investigation of massive protein with high variation data collection coordinating with the computational analysis. Overall, in the molecular dynamics' simulation study, the RMSD, radius of gyration, hydrogen bond changes, SASA, and RMSF analysis suggest that the nsSNPs P2683S has a lower deviation and high compactness than that of the native structure and the nsSNP R2702C.

The principal component analysis is determined to be a significant tool to understand the conformational changes occurred in the biological system of the protein molecule during the process of molecular dynamics simulation. PCA also confirms that the conformational stability and flexibility of the structures are high in mutants and less in native.

Finally, molecular docking study was performed for the native and the mutant protein structure with Methyl testosterone and Estradiol to determine the ability of the drugs over the mutations. The docking results concluded that the interaction and effect remain the same and no much changes are observed except the position at which the drugs fit in the binding site, stating that the drugs bind effectively with the native and Mutants.

Therefore, based on the sequence and structure analysis both the nsSNPs P2683S and R2702C are determined to have a huge impact to cause CHARGE syndrome in individuals harbouring this mutation and the drugs Methyltestosterone and Estradiol both works effective with the mutations retrieved from the current study. High-end computational approaches such as molecular dynamics and molecular docking studies deliver an understanding of the mutations and gather a heap of data for the experimental analysis. The outcomes acquired from our investigation will give a magnificent piece of information to

experimental biologists for the additional procedure and might go about as a stage for examining the rest of the variations in CDH7 protein.<sup>56</sup>

## **Conclusion**

This remains the first study that comprises the entire mutational investigation of the CHD7 protein; various computational approaches have been utilized to identify the most deleterious nsSNPs responsible for causing CHARGE syndrome. In sequence analysis, it was evident that the nsSNPs P2683S and R2702C to be more significant out of 201 nsSNPs. Based on the Molecular Dynamics Simulation, the nsSNPs P2683S and R2702C were more stable than those of the native structure. The sequence and structural analysis conclude that both the nsSNPs were found to have a high potential to cause this syndrome. The results from this study will aid in computational and therapeutic approaches against CHARGE Syndrome for the development of novel drugs and treatments that are potent to inhibit the disease.

## **Declarations**

### **Acknowledgment**

The authors like to acknowledge Sri Ramachandra Institute of Higher Education.

### **Conflict of Interest**

The authors declare no conflicts for the work.

### **Author Contribution Statement:**

The authors NR and MP were involved in work design and in drafting the manuscript. NR was involved in data collection and mutational analysis. MP carried out Molecular Dynamics and Molecular Docking related to the study. RM was involved in making the study design, supervising the work and critically examined the manuscript for submission. The authors approve the manuscript at its correct form.

### **Source of Funding:**

No funding was received.

### **Compliance with ethical standards:**

This article does not contain any studies with human or animal subjects performed by the any of the authors.

### **Data availability statement:**

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

## References

1. Jatana, S. K., Venkatnarayan, K. & Nair, M. N. G. CHARGE Syndrome. *Medical Journal Armed Forces India*. 261–263 [https://doi.org/10.1016/s0377-1237\(03\)80027-8](https://doi.org/10.1016/s0377-1237(03)80027-8) (2003).
2. Blake, K. D. *et al.* CHARGE Association: An Update and Review for the Primary Pediatrician. *Clinical Pediatrics*, pp159–173, (1998). <https://doi.org/10.1177/000992289803700302>.
3. Lalani, S. R. *et al.* Spectrum of CHD7 Mutations in 110 Individuals with CHARGE Syndrome and Genotype-Phenotype Correlation. *The American Journal of Human Genetics*. 303–314 <https://doi.org/10.1086/500273> (2006).
4. Dammeyer, J. Development and Characteristics of Children with Usher Syndrome and CHARGE Syndrome. *International Journal of Pediatric Otorhinolaryngology*. 1292–1296 <https://doi.org/10.1016/j.ijporl.2012.05.021> (2012).
5. Hsu, P. *et al.* CHARGE Syndrome: A Review. *Journal of Paediatrics and Child Health*. 504–511 <https://doi.org/10.1111/jpc.12497> (2014).
6. Sanlaville, D., Verloes, A. C. H. A. R. G. E. & Syndrome An Update. *European Journal of Human Genetics*. 389–399 <https://doi.org/10.1038/sj.ejhg.5201778> (2007).
7. earle, L. C., Graham, J. M., Prasad, J. & Blake, C. K. D. CHARGE Syndrome from Birth to Adulthood: An Individual Reported on from 0 to 33 Years. *Am. J. Med. Genet. A*. **133A** (3), 344–349 (2005).
8. Corsten-Janssen, N., van Ravenswaaij-Arts, C. M. A. & Kapusta, L. Congenital Arch Vessel Anomalies in CHARGE Syndrome: A Frequent Feature with Risk for Co-Morbidity. *IJC Heart & Vasculature*. 21–25 <https://doi.org/10.1016/j.ijcha.2016.05.015> (2016).
9. ongmans, M. C. J. *et al.* L. H. CHD7 Mutations in Patients Initially Diagnosed with Kallmann Syndrome—the Clinical Overlap with CHARGE Syndrome. *Clin. Genet*. **75** (1), 65–71 (2009).
10. Bosman, E. A. *et al.* Multiple Mutations in Mouse Chd7 Provide Models for CHARGE Syndrome. *Hum. Mol. Genet*. 3463–3476 <https://doi.org/10.1093/hmg/ddi375> (2005).
11. Hurd, E. A. *et al.* Loss of Chd7 Function in Gene-Trapped Reporter Mice Is Embryonic Lethal and Associated with Severe Defects in Multiple Developing Tissues. *Mammalian Genome*, pp94–104, (2007). <https://doi.org/10.1007/s00335-006-0107-6>.
12. Gennery, A. R. *et al.* Mutations inCHD7in Patients with CHARGE Syndrome Cause TBnatural Killer Cellsevere Combined Immune Deficiency and May Cause Omenn-like Syndrome. *Clinical & Experimental Immunology*. 75–80 <https://doi.org/10.1111/j.1365-2249.2008.03681.x> (2008).
13. Lalani, S. R. *et al.* W. SNP Genotyping to Screen for a Common Deletion in CHARGE Syndrome. *BMC Medical Genetics*. <https://doi.org/10.1186/1471-2350-6-8> (2005).
14. Balasubramanian, R. *et al.* Functionally compromisedCHD7alleles in Patients with Isolated GnRH Deficiency. *Proceedings of the National Academy of Sciences*, pp 17953–17958(2014).

- <https://doi.org/10.1073/pnas.1417438111>.
15. Bergman, J. E. H. *et al.* W. A Novel Classification System to Predict the Pathogenic Effects of CHD7 Missense Variants in CHARGE Syndrome. *Hum. Mutat.* 1251–1260 <https://doi.org/10.1002/humu.22106> (2012).
  16. Jongmans, M. C. J. *et al.* M. A. CHARGE Syndrome: The Phenotypic Spectrum of Mutations in the CHD7 Gene. *J. Med. Genet.* **43** (4), 306–314 (2006).
  17. Pemble, S. *et al.* Human Glutathione S-Transferase Theta (GSTT1): cDNA Cloning and the Characterization of a Genetic Polymorphism. *Biochem. J.* **300** (Pt 1), 271–276 (1994).
  18. Ali-Osman, F., Akande, O., Antoun, G., Mao, J. X. & Buolamwini, J. Molecular Cloning, Characterization, and Expression in *Escherichia Coliof* Full-Length cDNAs of Three Human Glutathione S-Transferase Pi Gene Variants. *Journal of Biological Chemistry.* 10004–10012 <https://doi.org/10.1074/jbc.272.15.10004> (1997).
  19. Hadi, M. Z. Functional Characterization of Ape1 Variants Identified in the Human Population. *Nucleic Acids Res.* 3871–3879 <https://doi.org/10.1093/nar/28.20.3871> (2000).
  20. Matullo, G., XRCC1, X. R. C. C. & Gene Polymorphisms, X. P. D. Smoking and 32P-DNA Adducts in a Sample of Healthy Subjects. *Carcinogenesis.* 1437–1445 <https://doi.org/10.1093/carcin/22.9.1437> (2001).
  21. Pakula, A. A., Young, V. B. & Sauer, R. T. Bacteriophage Lambda Cro Mutations: Effects on Activity and Intracellular Degradation. *Proc. Natl. Acad. Sci. U. S. A.*, 83 (23), 8829–8833(1986).
  22. Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD®): 2003 Update. *Human Mutation.*, pp577–581, (2003). <https://doi.org/10.1002/humu.10212>.
  23. Bairoch, A., Apweiler, R., The, S. W. I. S. S. P. R. O. T. & Protein Sequence Data Bank and Its Supplement TrEMBL. *Nucleic Acids Research.*, pp31–36, (1997). <https://doi.org/10.1093/nar/25.1.31>.
  24. Sayers, E. W. *et al.* Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **48** (D1), D9–D16 (2020).
  25. Kouranov, A. *et al.* The RCSB PDB Information Portal for Structural Genomics. *Nucleic Acids Res.* **34** (Database issue), D302–D305 (2006).
  26. Bendl, J. *et al.* Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Computational Biology.* 1003440 <https://doi.org/10.1371/journal.pcbi.1003440> (2014).
  27. Doss, G. P., Nagasundaram, C., Chakraborty, N., Chen, C. & Zhu, L. H. Extrapolating the Effect of Deleterious nsSNPs in the Binding Adaptability of Flavopiridol with CDK7 Protein: A Molecular Dynamics Approach. *Hum. Genomics*, 7, 10, (2013).
  28. Thirumal Kumar, D. *et al.* H. Computational Approach to Unravel the Impact of Missense Mutations of Proteins (D2HGDH and IDH2) Causing D-2-Hydroxyglutaric Aciduria 2. *Metab. Brain Dis.* **33** (5), 1699–1710 (2018).

29. Mathe, E. Computational Approaches for Predicting the Biological Effect of p53 Missense Mutations: A Comparison of Three Sequence Analysis Based Methods. *Nucleic Acids Res.* 1317–1325 <https://doi.org/10.1093/nar/gkj518> (2006).
30. P, S., D, T. K., C, G. P. D., R, S. & Zayed, H. Determining the Role of Missense Mutations in the POU Domain of HNF1A That Reduce the DNA-Binding Affinity: A Computational Approach. *PLoS One*, 12 (4), e0174953(2017).
31. Berman, H., Henrick, K. & Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nature Structural & Molecular Biology.* 980–980 <https://doi.org/10.1038/nsb1203-980> (2003).
32. Sussman, J. L. *et al.* Protein Data Bank (PDB): Database of Three-Dimensional Structural Information of Biological Macromolecules. *Acta Crystallographica Section D Biological Crystallography.* 1078–1084 <https://doi.org/10.1107/s0907444998009378> (1998).
33. Johansson, M. U., Zoete, V., Michielin, O. & Guex, N. Defining and Searching for Structural Motifs Using DeepView/Swiss-PdbViewer. *BMC Bioinformatics.* 173 <https://doi.org/10.1186/1471-2105-13-173> (2012).
34. Glaser, F. *et al.* Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics.* 163–164 <https://doi.org/10.1093/bioinformatics/19.1.163> (2003).
35. Accelrys Discovery Studio Visualizer 2.0, (2008).
36. Rampogu, S. *et al.* Discovery of Lonafarnib-Like Compounds: Pharmacophore Modeling and Molecular Dynamics Studies. *ACS Omega.* **5** (4), 1773–1781 (2020).
37. Doss, C. G. P. *et al.* Genetic Epidemiology of Glucose-6-Phosphate Dehydrogenase Deficiency in the Arab World. *Sci. Rep.* <https://doi.org/10.1038/srep37284> (2016).
38. Sundarrajan, S., Nandakumar, M. P., Prabhu, D., Jeyaraman, J. & Arumugam, M. Conformational Insights into the Inhibitory Mechanism of Phyto-Compounds against Src Kinase Family Members Implicated in Psoriasis. *Journal of Biomolecular Structure and Dynamics.* 1398–1414 <https://doi.org/10.1080/07391102.2019.1605934> (2020).
39. Yadav, P., Sharma, B., Sharma, C., Singh, P. & Awasthi, S. K. Interaction between the Antimalarial Drug Dispiro-Tetraoxanes and Human Serum Albumin: A Combined Study with Spectroscopic Methods and Computational Studies. *ACS Omega.* **5** (12), 6472–6480 (2020).
40. Aydogdu, A. & Swerdloff, R. S. Emerging Medication for the Treatment of Male Hypogonadism. *Expert Opin. Emerg. Drugs.* **21** (3), 255–266 (2016).
41. Quinn, S. M. & Rubin, D. J. Treatment Options for Females with Hypogonadotropic Hypogonadism and Thromboembolism. *Endocrinology Studies.* **1** (2), e15–e15 (2011).
42. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res.* **46** (W1), W363–W367 (2018).
43. Bartels, C. F., Scacheri, C., White, L., Scacheri, P. C. & Bale, S. Mutations in the CHD7 Gene: The Experience of a Commercial Laboratory. *Genetic Testing and Molecular Biomarkers.*, pp881–891, (2010). <https://doi.org/10.1089/gtmb.2010.0101>.

44. Hale, C. L., Niederriter, A. N., Green, G. E. & Martin, D. M. Response to Correspondence to Hale et Al. Atypical Phenotypes Associated with pathogenicCHD7variants and a Proposal for Broadening CHARGE Syndrome Clinical Diagnostic Criteria. *American Journal of Medical Genetics Part A*. 3367–3368 <https://doi.org/10.1002/ajmg.a.37629> (2016).
45. Rajasekaran, R., Sudandiradoss, C., Doss, C. G. P. & Sethumadhavan, R. Identification and in Silico Analysis of Functional SNPs of the BRCA1 Gene. *Genomics*. **90** (4), 447–452 (2007).
46. Zaki, O. K., Priya Doss C, G., Ali, S. A., Murad, G. G., Elashi, S. A., Ebnou, M.S. A., ... Zayed, H.. Genotype-phenotype correlation in patients with isovaleric acidaemia:comparative structural modelling and computational analysis of novel variants. *Human molecular genetics*, *26*(16), 3105–3115, (2017).
47. Doss, G. P., Zayed, H. & C., & Comparative computational assessment of the pathogenicity of mutations in the Aspartoacylase enzyme. *Metabolic brain disease*. **32** (6), 2105–2118 (2017).
48. Shastry, B. S. SNPs: impact on gene function and phenotype. *Methods in molecular biology*. **578**, 3–22 (2009).
49. Jayakanthan, S., McCann, C. & Lutsenko, S. Biochemical and Cellular Properties of ATP7B Variants. *Wilson Disease*. <https://doi.org/10.1016/b978-0-12-811077-5.00004-9> (2019). )
50. Riva, A. & Kohane, I. S. SNPper: retrieval and analysis of human SNPs. *Bioinformatics*, (2002). <https://doi.org/10.1093/bioinformatics/18.12.1681>
51. B, R. & C, G. P. D. Path to facilitate the prediction of functional amino acid substitutions in red blood cell disorders—a computational approach. *PloS one*. **6** (9), e24607 (2011).
52. Carlson, B. SNPs-A Shortcut to Personalized Medicine | GEN. <https://www.genengnews.com/gen-articles/snpsa-shortcut-topersonalized-medicine/2507.>, (2008).
53. Zlotogora, J. Germ line mosaicism. *Human Genetics*, (1998). <https://doi.org/10.1007/s004390050708>
54. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis*, (1997). <https://doi.org/10.1002/elps.1150181505>
55. Tanwar, H., Thirumal Kumar, D., Doss, G. P., Zayed, H. & C., & Bioinformatics classification of mutations in patients with Mucopolysaccharidosis IIIA. *Metab. Brain Dis*. <https://doi.org/10.1007/s11011-019-00465-62019>
56. Zentner, G. E., Layman, W. S., Martin, D. M. & Scacheri, P. C. Molecular and phenotypic aspects ofCHD7mutation in CHARGE syndrome. *American Journal of Medical Genetics Part A*. <https://doi.org/10.1002/ajmg.a.333232010>

## Unsectioned Paragraphs

### Author Contribution Statement:

The authors NR and MP were involved in work design and in drafting the manuscript. NR was involved in data collection and mutational analysis. MP carried out Molecular Dynamics and Molecular Docking

related to the study. RM was involved in making the study design, supervising the work and critically examined the manuscript for submission. The authors approve the manuscript at its correct form.

### Source of Funding:

No funding was received.

## Figures

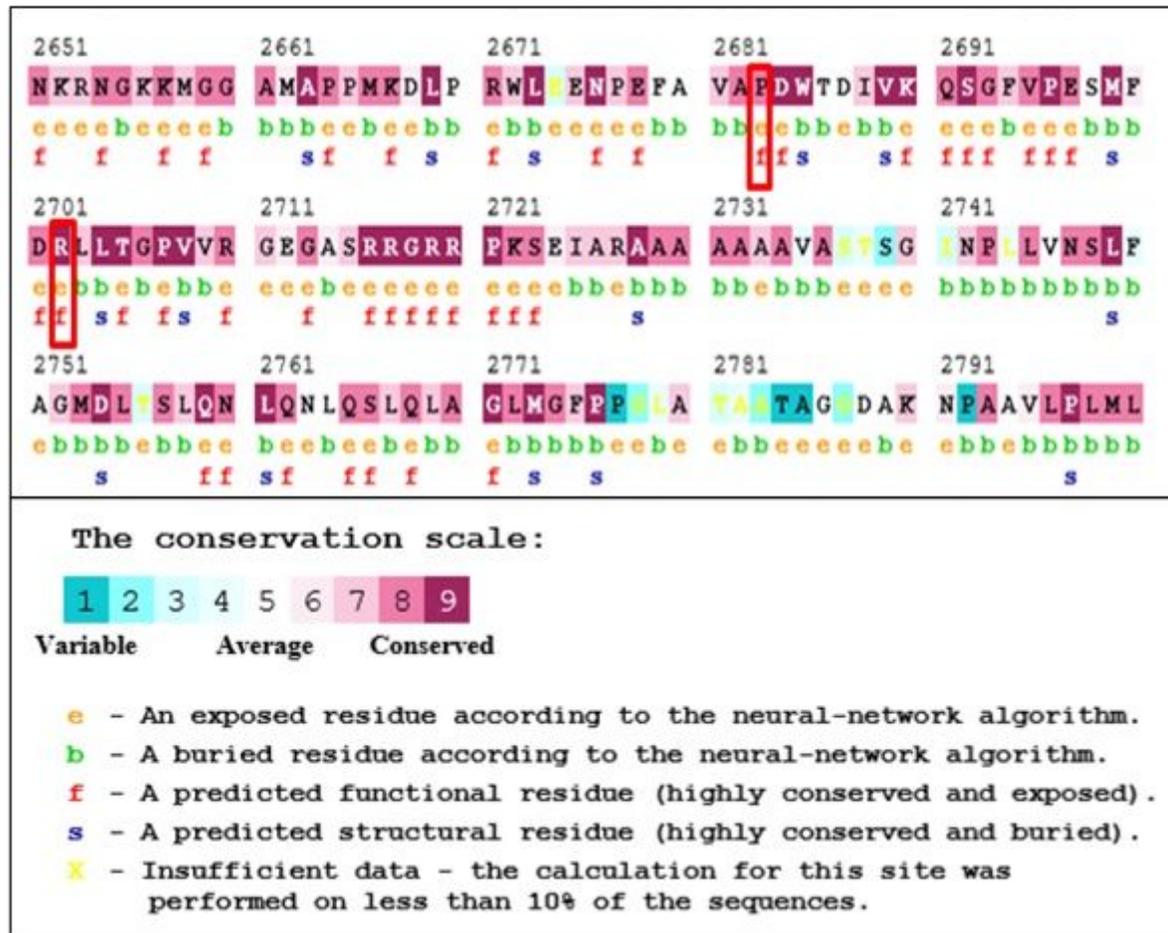
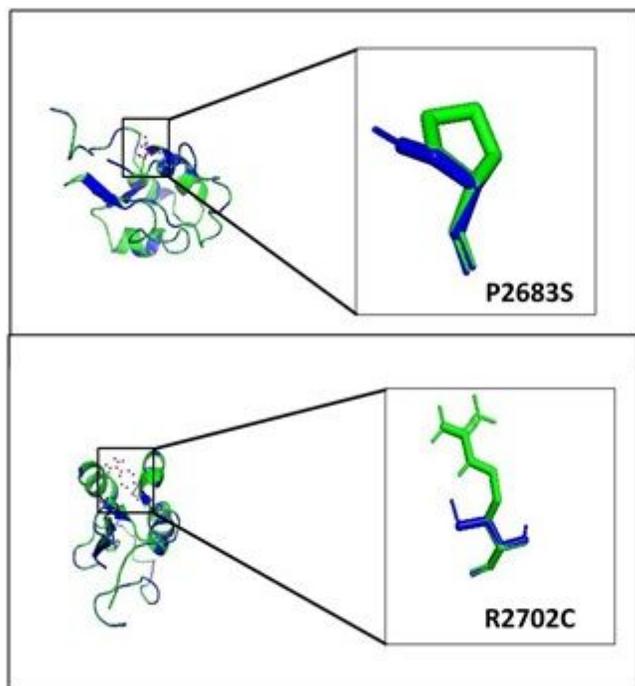


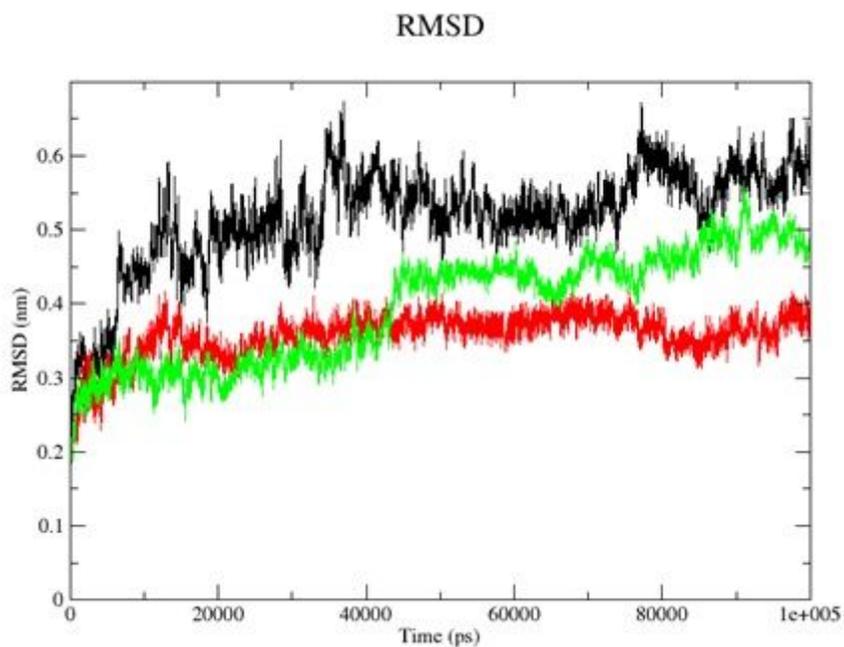
Figure 1

ConSurf results for the selected nsSNP's P2683S and R2702C



**Figure 2**

Superimposed structures of native and mutant models (A: P2683S and B: R2702C) of CHD7 protein respectively. (Green: Native Structure and Blue: mutated region)



**Figure 3**

The above figure represents the Root Mean Square Deviation (RMSD) during the time period of 100ns. Black: Native, Red: P2683S and Green: R2702C. (X axis = time and Y axis = RMSD values)

### Radius of gyration

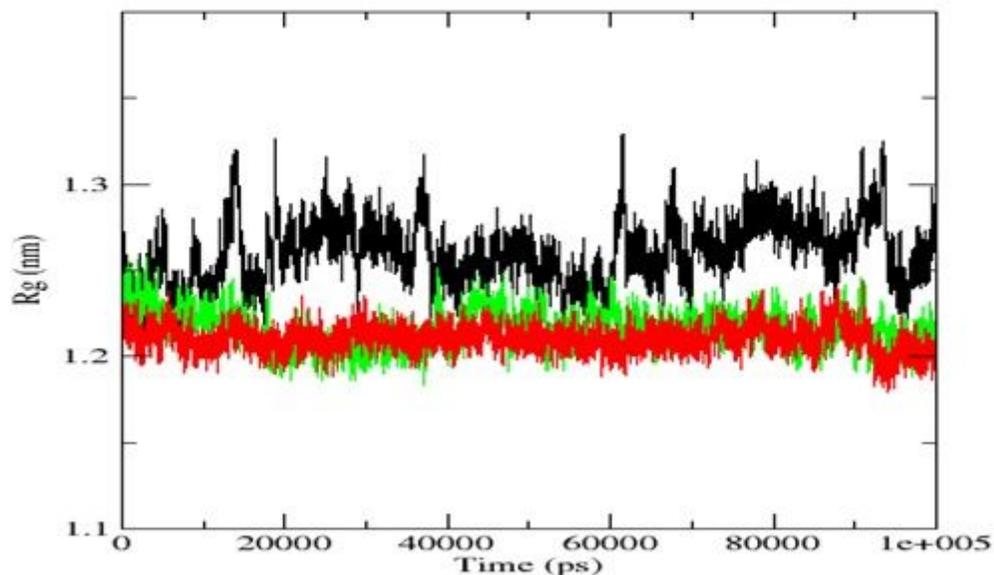


Figure 4

The above figure represents the Structural Flexibility of the protein due to mutation, Black: Native, Red: P2683S and Green: R2702C. (X axis = time and Y axis = radius of gyration)

### Hydrogen Bonds

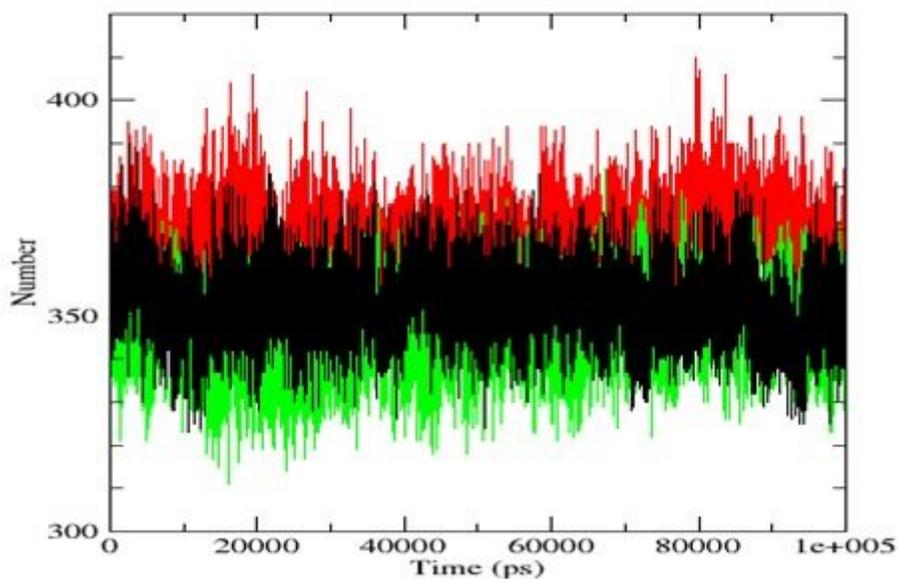


Figure 5

The above figure represents the fluctuations in the number of hydrogen bonds in the protein due to mutation, Black: Native, Red: P2683S and Green: R2702C. (X axis = time and Y axis = no. of hydrogen bond)

### Solvent Accessible Surface

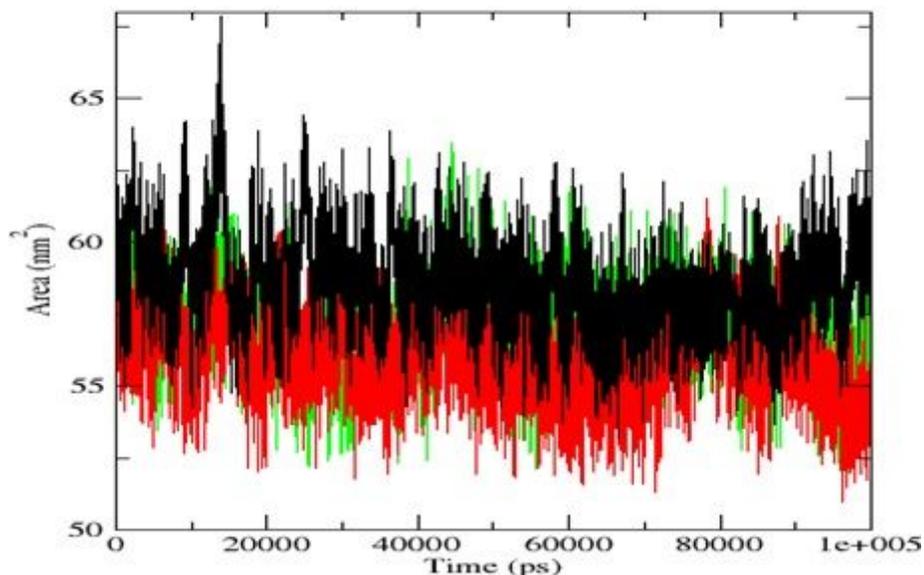


Figure 6

The above figure represents the Structural Flexibility of the protein due to mutation, Black: Native, Red: P2683S and Green: R2702C. (X axis = time and Y axis = radius of gyration)

### RMS fluctuation

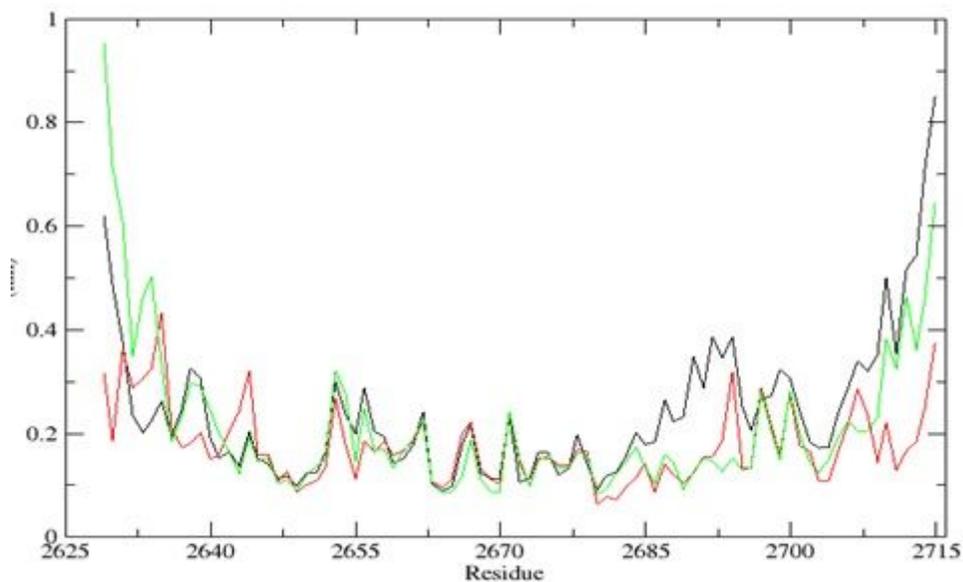
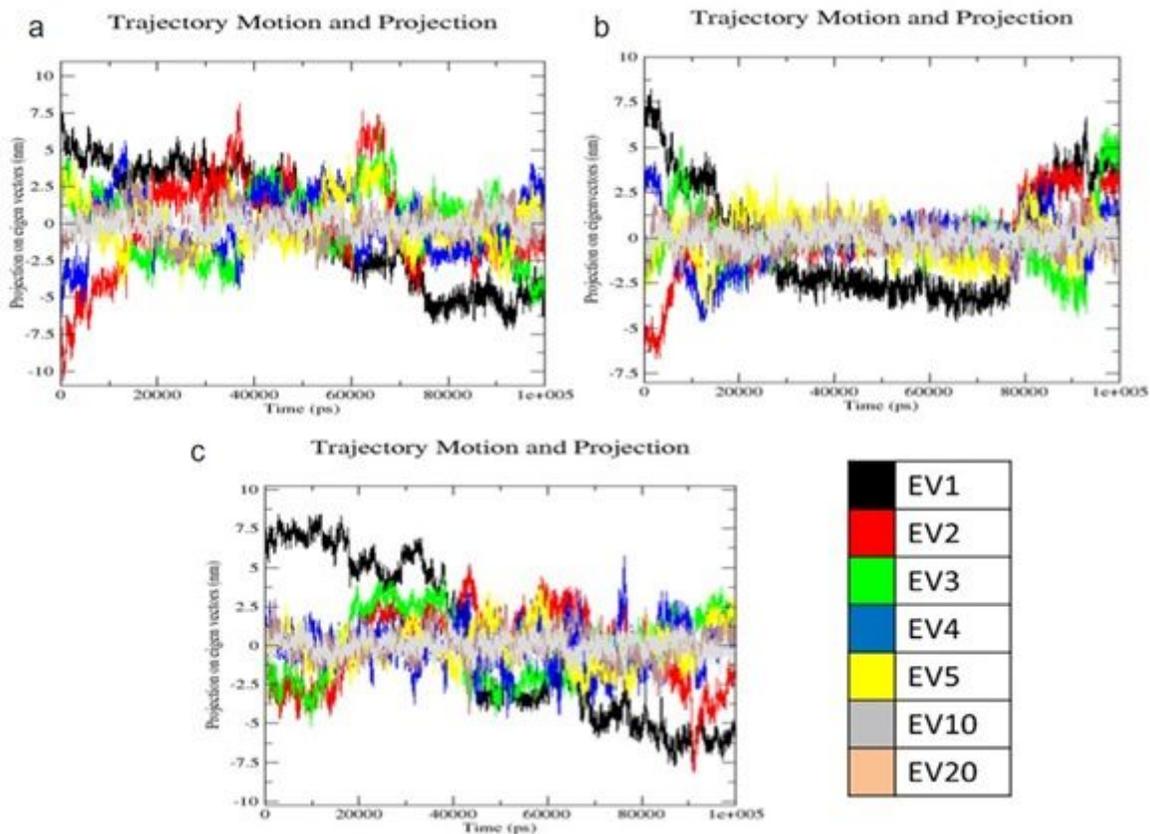


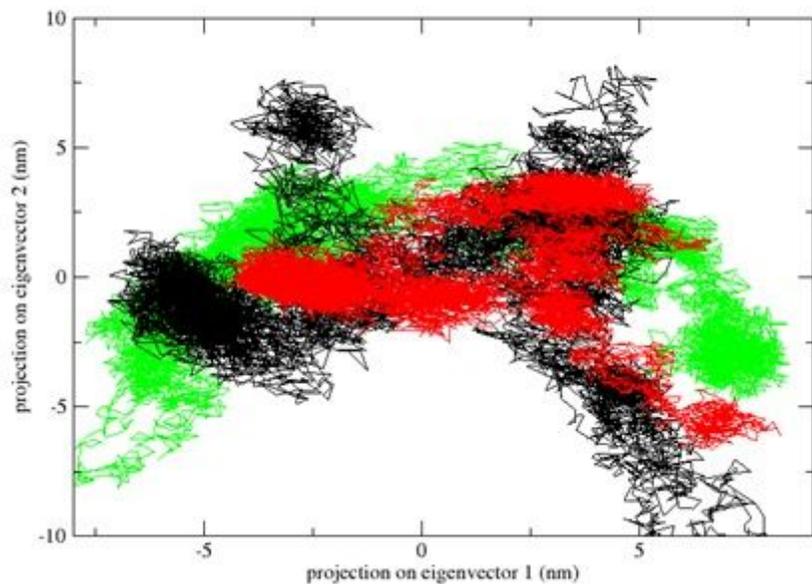
Figure 7

The above figure represents the Root Mean Square Fluctuation (RMSF) during the time period of 100ns. Black: Native, Red: P2683S and Green: R2702C. (X axis = time and Y axis = RMSF peaks)



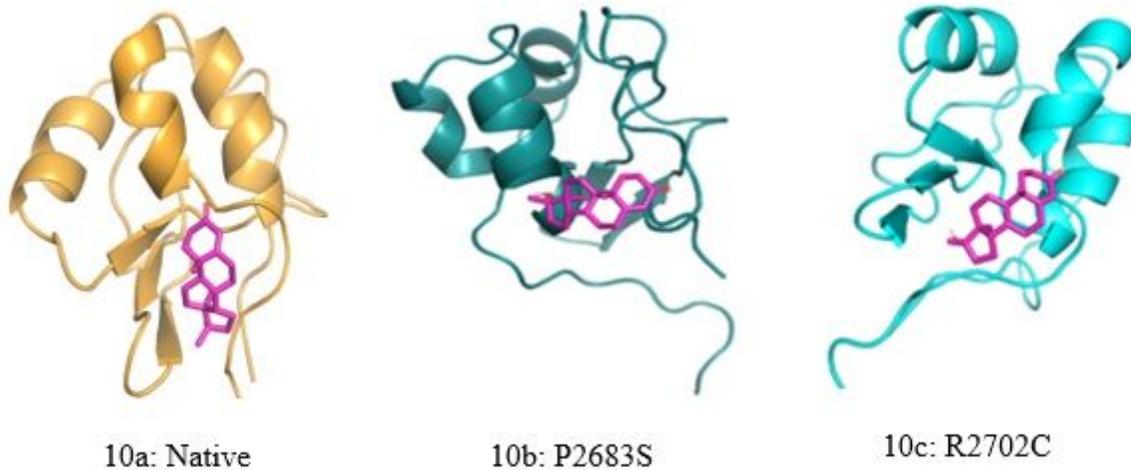
**Figure 8**

(a-c) Eigen vectors 1-5, 10 & 20 plotted for 100ns MDS: (a) Native, (b) P2683S, (c) R2702C  
 2D projection of trajectory



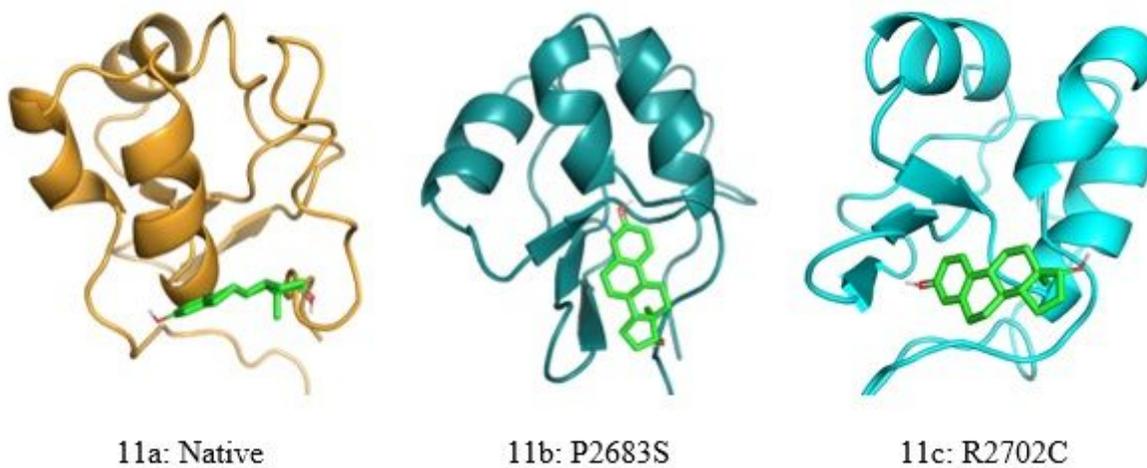
**Figure 9**

2D projection of eigenvector 1 & 2 for 100ns MDS: Black: Native, Red: P2683S and Green: R2702C



**Figure 10**

Interaction of Methyltestosterone with CHD7 proteins. (Magenta: Methyltestosterone, Yellow orange: Native, Deep teal: P2683S, Cyan: R2702C) 10a: Interaction of native structure with Methyltestosterone, 10b: Interaction of mutant structure (P2683S) with Methyltestosterone and 10c: Interaction of mutant structure (R2702C) with Methyltestosterone.



**Figure 11**

Interaction of Estradiol with CHD7 proteins. (Green: Estradiol, Yellow orange: Native, Deep teal: P2683S, Cyan: R2702C) 11a: Interaction of native structure with Estradiol, 11b: Interaction of mutant structure (P2683S) with Estradiol and 11c: Interaction of mutant structure (R2702C) with Estradiol.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFiles.pdf](#)