Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

Article

Analysis of the Structural Protein Effects Caused by the *PURA* p.Phe233del Mutation Associated to Cognitive Developmental Delay Using Artificial Intelligence and Hybrid Quantum Mechanics-Molecular Mechanics Modelling

Juan Javier López-Rivera[1,2,*], Luna Rodríguez-Salazar[3], Alejandro Soto-Ospina[4], Carlos Estrada-Serrato[2], David Serrano[2], Henry Mauricio Chaparro-Solano[2], Olga Londoño[2], Paula A. Rueda[3], Geraldine Ardila[3], Andrés Villegas-Lanau[4,5], Marcela Godoy-Corredor[6], Mauricio Cuartas[7], Jorge I. Vélez[8], Oscar M. Vidal[8], Mario A. Isaza-Ruget[1] and Mauricio Arcos-Burgos[9,*]

[1] INPAC Research Group, Fundación Universitaria Sanitas, Bogotá, Colombia.
[2] Grupo de Genética Médica, Clínica Universitaria Colombia y Clínica Pediátrica Colsanitas, Bogotá, Colombia
[3] Grupo de Bioinformática de Laboratorio de Clínica Colsanitas
[4] Genética Molecular (GenMol), Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia, Medellín, Colombia.
[5] Grupo de Neurociencias de Antioquia (GNA), Facultad de Medicina, Universidad de Antioquia, Medellín, Colombia.
[6] Laboratorio Clínico y de Patología, Clínica Colsanitas, Bogotá, Colombia.
[7] Grupo de Investigación Estudios en Psicología, Departamento de Psicología, Escuela de Humanidades, Universidad EAFIT, Medellín, Antioquia, Colombia.
[8] Universidad del Norte, Barranquilla, Colombia
[9] Grupo de Investigación en Psiquiatría (GIPSI), Departamento de Psiquiatría, Instituto de Investigaciones Médicas, Facultad de Medicina, Universidad de Antioquia, Medellín, Colombia.
* Correspondence: jjlopez@colsanitas.com, Juan Javier López-Rivera MD. MSc., Professor Fundación Universitaria Sanitas, Bogotá, Colombia, and mauricio.arcos@udea.edu.co, Mauricio Arcos-Burgos, Professor Universidad de Antioquia, Medellín, Colombia.

Abstract: A whole-exome capture and next-generation sequencing applied to an 11 y/o patient with a clinical history of congenital hypotonia, generalized motor and cognitive neurodevelopmental delay, severe cognitive deficit, without any identifiable Syndromic pattern, and to her parents, disclosed a *de novo* heterozygous pathogenic mutation, c.697_699del p.Phe233del (rs786204835)(ACMG classification PS2, PM1, PM2, PP5), harbored in the *PURA* gene (MIM*600473) (5q31.3), associated to Autosomal Dominant Mental Retardation 31 (MIM # 616158). We used the significant improvement in the accuracy of protein structure prediction recently implemented in AlphaFold that incorporates novel neural network architectures and training procedures based on the evolutionary, physical, and geometric constraints of protein structures. The wild-type (WT) sequence and the mutated one, missing the Phe233, were reconstructed. The predicted local Distance Difference Test (lDDT) for the PURA WT and the PURA-Phe233del showed that the occurrence of the Phe233del affects between 220-320 amino acids. The distortion in the PURA structural conformation in the ~5Å surrounding area after the p.Phe233del produces a conspicuous disruption of the repeat III, where the DNA and RNA helix unwinding capability occurs. PURA Protein-DNA Docking corroborated these results *in silico* Analysis that showed a loss of the contact of the PURA-Phe233del III repeat domain model with the DNA. Together, *i)* the energetic and stereochemical, *ii*) the hydropathic indices and polarity surfaces, and *iii*) the hybrid Quantum Mechanics-Molecular Mechanics (QM-MM) analyses of the PURA molecular models demarcate at the atomic resolution the specific surrounding region affected by these mutations and paves the way for future cell-based functional analysis. To the best of our knowledge, this is the first report of a *de novo* mutation underpinning a PURA syndrome in a Latin American patient and highlights the importance of predicting the molecular effects in protein structure using artificial intelligence algorithms and molecular and atomic resolution stereochemical analyses.

© 2022 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

Keywords: purine-rich element binding protein A gene; PURA; transcriptional activator protein Pur-alpha; cognitive developmental delay; mental retardation; Mutation c.697_699del p.Phe233del

1. Introduction

The *Purine-rich element-binding protein A* gene (*PURA*; MIM: 600473) is a single exon gene harbored on the chromosome 5q31.2 that encodes the transcriptional activator protein Pur-alpha (PURA; Uniprot: Q00577), a multifunctional protein, member of the Pur family of nucleic acid-binding proteins which consist of a glycine-rich flexible amino terminus, a central core region, and a potential carboxy-terminal protein binding region.(1–3) The *PURA* gene is expressed almost ubiquitously, including in the brain, muscle, heart, and blood.(4–6)

All human Pur proteins have three highly conserved sequence-specific repeats (Pur repeats I–III) of 64–80 amino acids that are the hallmark of the Pur proteins.(4) Pur-alpha has the helix-unwinding capability and has been shown to bind specific sequences of ssDNA, dsRNA, and ssRNA with a preference for GGN-repeats (*i.e.,* the GGCGGA sequence derived from the myelin essential protein (MBP) proximal regulatory region),(7) to regulate a variety of cellular processes including DNA replication, gene transcription, RNA transport, and mRNA translation.(8,9)

Pur-alpha plays a major role in developing the central nervous system acting in the process such as proliferation, dendrite maturation, and the transport of mRNA to translation sites in hippocampal neurons.(10,11) Knockout mice models of *PURA* show failure to thrive and develop neurologic features, including ataxic gait, hind limb weakness, and abnormal movements such as tremors.(11,12)

In humans, *PURA* heterozygous mutations underlay a complex dominant phenotype (PURA syndrome) characterized by moderate to severe hypotonia, neurodevelopmental, motor, and language delay, feeding difficulties, apneas, epileptic seizures, abnormal non-epileptic movements, visual problems, and less common, congenital heart defects, urogenital malformations, skeletal abnormalities, and endocrine disorders.(4–6,13,14) Since the description of severe mental retardation in four patients with de novo mutations in the *PURA* gene, up to date, ~60 different mutations have been characterized in 75 individuals with PURA syndrome.(4–6,15–24)

Prediction of the functional impact of *PURA* mutations is difficult, considering the pleiotropic nature of the protein. Furthermore, PURA is expressed in almost every cell, making it more challenging to create a mutational profile.(8,9) Therefore, the characterization of nonsynonymous mutations in PURA requires further study to address the role of PURA mutations in the etiology of the disease and potential pathways affected to induce its pathogenesis.

In this article, we present the identification of a *de novo* heterozygous pathogenic mutation, c.697_699del p.Phe233del (rs786204835) harbored in the *PURA* gene in an 11 y/o girl with cognitive developmental delay. As this mutation has been previously described, we explore different *in silico* protein structure prediction tools such as AlphaFold,(25) to study the impact of this deletion. We showed that after the p.Phe233del mutation, a distortion of the ~5Å surrounding area occurs, generating a disruption of the repeat III and a structural alteration of the β-sheet holding Phe233. We validated the interactions between PURA and the human *c-MYC* and *MBP* human upstream promotor regions. Our analysis for the predicted structures of both the AlphaFold PURA wild type (PURAwt) and the PURA-Phe233del successfully showed DNA-Protein biding impairment.

These results will pave the way for future prediction/validation of DNA and RNA binding proteins affected in human disorders through AlphaFold-predicted structures modeling and refined experiments of stereochemistry at atomic resolution. This article represents the first report of a *de novo* mutation underpinning a PURA syndrome in a Latin

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

American patient and highlights the importance of predicting the protein structure molecular effects with artificial intelligence and advanced molecular chemistry algorithms.

2. Patients and Methods

The propositus is an 11 y/o female, the unique daughter of non-consanguineous parents, with a history of congenital hypotonia, failure to thrive, severe cognitive and motor deficit, and a current language with sparse disyllables. At the physical examination in good general condition: *Heart Rate*: 74 beats/min, *Respiratory Rate*: 20 breaths/min, *Systolic Blood Pressure*: 95 mmHg, *Diastolic Blood Pressure*: 50 mmHg, *Mean Blood Pressure*: 65 mmHg, *Temperature*: 36 ºC. *Sense Organs:* Oral mucosa moist and pink, oropharynx not congestive, without plaques or exudates, right and left eye without alterations, right and left ear without alterations. *Neck:* No masses, *Cardiovascular:* Rhythmic heart sounds, no presence of murmurs. *Pulmonary*: Respiratory sounds present in both pulmonary fields, without aggregates. *Abdomen and pelvis:* Soft, depressible, not painful, no signs of peritoneal irritation, bowel sounds present. *Upper extremities:* Peripheral pulses present, rhythmic and regular. *Lower extremities:* Peripheral pulses present, rhythmic and regular, without edema. *Osseo-muscular-articular:* No alterations. *Neurologic*: Alert, oriented, without apparent deficits. *Skin and Fanners:* Normal. *Paraclinical Tests*: Brain MRI was reported as normal, FISH to discard Prader-Willi syndrome reported as normal.

2.1. Whole Exome Capture

Three methods were applied to DNA quantification and qualification: (1) DNA purity was checked using the Nanodrop (OD260/280 ratio); (2) DNA degradation and contamination were monitored on 1% agarose gels; (3) DNA concentration was measured using Qubit. DNA samples with OD260/280 ratio between 1.8~2.0 and concentration above 1.0ug were used to prepare sequencing libraries.

2.2. Next Generation Sequencing

Next-Generation Sequencing (NGS) of the whole exome (the entire coding region of the genome and intron-exon junction regions, corresponding to approximately 23,000 nuclear genes) was performed on a DNB-SEQ400 next-generation mass sequencer using an MGI-V5 exome library. All sequenced data were quality assessed (base quality distribution, nucleotide distribution, and presence of adapters, chimeras, and other contaminants) to identify/remove low-quality data/samples from further analysis.

2.3. Bioinformatic analysis

All high-quality data was then mapped to the human genome assembly using the `bwa-mem` algorithm. Aligned files were processed using Genome Analysis Tool Kit (GATK) for base quality recalibration, indel realignments, and duplicate removal. This was followed by SNP and INDEL discovery and genotyping (plus phasing where applicable) according to GATK Best Practices recommendations.

All variant calls were subject to variant quality score recalibration and filtering to remove low-quality variants. The remaining high-quality variants were annotated for predicted functional consequences using the Voting Report Index, including SIFT, PolyPhen2 HVAR, Mutation Taster, Mutation Assessor, FATHMM, and FATHMM MKL Coding. For example, for a conservative filter, simply keep things that have 0, 1, or maybe two tolerated predictions. A more conservative filter would keep based on 3, 4, or 5 damaging predictions. Many variants do not have five algorithms with non-missing values.

Subsequently, variants identified in genes with the known clinical association and related to monogenic and mitochondrial diseases were analyzed in contrast to our patient's phenotype and clinical manifestations. Only genes with an average coverage of more than 98% and a minimum depth of 20x were analyzed for accomplishing quality standards. As a note of caution, it is essential to mention that this study may not cover all possible pathogenic variants in each gene, and for technical and scientific reasons, whole-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

exome sequencing may not completely cover the entire coding region of the human genome. The sequencing results were bioinformatically analyzed using the DreamGenics-GenomeOne software, and the aligned and filtered sequences (fulfilling specific quality criteria) were compared against the GRCh37/hg19 reference genome for annotation and variant calling.

Bioinformatics analyses aimed to identify variants included in exonic regions or splicing regions (at least 20 bp), small insertions and deletions, and, in particular cases, genes related to the patient's clinical phenotype. Parameters such as the population frequency of the variant (i.e., in ExAC, gnomAD, and 1000 Genomes databases), its allelic influence, the associated OMIM phenotype, its classification in ClinVar, and the inheritance model were used to filter the variants to be reported. This sequential, comprehensive, and sometimes targeted analysis allowed the identification of exonic deletions and duplications (also known as Copy Number Variants, or CNVs) and variants involving large gene regions. If CNVs were identified, a secondary confirmation method was either applied or recommended, as NGS is an indirect method for their identification.

Variants outside of coding regions or harbored out of the intron-exon junctions of the gene cannot be identified (e.g., variants in promoter regions, enhancers, regulatory regions far from the exonic region, unique variants including dynamic mutations, complex recombination structures, structural variants of genes [e.g., inversion-type rearrangements], and epigenetic effects). Identified variants have been evaluated against HGMD, ClinVar, LOVD, dbSNP, and gnomAD databases. For missense and splicing site variants of uncertain significance, *in silico* prediction tools were used, and their analysis is reported in the interpretation of the results. Finally, the association of the identified variants with the syndromes described in OMIM and the clinical association with the phenotype described in the patient are evaluated.

We followed the recommendations of the American College of Medical Genetics and Genomics (ACMG) and reported pathogenic or probably pathogenic variants in 59 clinically significant genes. According to ACMG and the current knowledge, mutations identified as benign are not reported; these variants, in general, have an allele frequency greater than or equal to 1% and result in a synonymous amino acid change or occur in 5' or 3' untranslated regions. This information can be made available upon request. A negative result does not rule out the possibility that the tested individual has a rare untested mutation in an undetectable region. In these cases, reanalyzing the data in the future could generate new results.

2.4. Protein Reconstruction

We use the significant improvement in the accuracy of protein structure prediction recently implemented in AlphaFold, which incorporates novel neural network architectures and training procedures based on the evolutionary, physical, and geometric constraints of protein structures. 3D protein structure reconstruction with AlphaFold is vastly more accurate than those obtained by competing methods, *i.e.*, median backbone accuracy, highly accurate side chains reconstruction, accurate domains, domain-packing prediction, and precise, per-residue estimates of its reliability. The wild-type (WT) sequence (Uniprot: Q00577) and the mutated one, missing the phe233, were reconstructed.

2.4.1. PURA Protein-DNA Docking in silico Analysis

We applied HDOCK (http://hdock.phys.hust.edu.cn/), a highly integrated suite for several in silico features such as macromolecular docking. HDOCK supports protein–RNA/DNA docking with an intrinsic scoring function. This tool, based on a hybrid algorithm of template-based modeling and ab initio free docking, delivers both template- and docking-based binding models of two molecules and allows interactive visualization.(26). A PURA sequence-specific single-stranded DNA-binding was reported in cell culture and brain extracts.(27–30) We used these reported purine-rich sequences found within a zone to initiate DNA replication upstream of the human c-MYC and MBP genes.(31)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

We used the *c-Myc* upstream region Genomic chr8: 5′ tttctcttttggaggtggtggagggagagaaaaagtttacttaaaatgcct 3′ and MBP upstream region Genomic chr18 specifically, 5′cagggagccgcccccacttgatccgcctcttttcccgagatgccccggggagggagggacaacaccttcaaagacaggccctctgagtccgacgagctcca 3′.

2.4.2. Energetic and Stereochemical Characterization of the PURA molecular Models

Models obtained by molecular modeling were energetically and stereochemically characterized. For the energetic characterization, we used SAVES v6. 0 coupled to the ERRAT parameter. Additionally, we used the SWISS_MODEL server to characterize the QMEAN6 parameter to define a normalized value (linear combination of six terms in favor of its stability). We also estimated the *Z*-score parameter to indicate whether the model is in trend or not in comparison with experimentally solved structures (either with X-ray diffraction [XRD], cryogenic imaging microscopy [CryoEM], or solid nuclear magnetic resonance [NMR] techniques).(32,33) The stereochemical characterization was performed with the PROCHECK software, which allows measuring the stereochemical viability of the model by considering the dihedral angles Phi (φ) and Psi (ψ) of the planes generated by the peptide bonds of the amide functional groups that link the amino acids in the proteins, all this with the α-carbons that constitute the mobility in the protein system.(34)

2.4.3. Calculation of Hydropathic Indices and Polarity Surfaces of PURAwt and Mutated-Phe233del

Because a deletion with potential effects was found on the conformation of the secondary structure, polarity evaluation was performed through the hydropathicity index, and the Molecular Lipophilicity Potential (MLP) maps for proteins. The polarity changes in the surroundings of the mutation were calculated with the Protscale software of the Swiss ExPASY suite, which compares the primary sequence of the PURAwt and the Mutated-Phe233del variant using the Kyte and Dolittle coefficients. Depending on the score, and its sign, a hydrophilic environment (negative) or a hydrophobic environment (positive) can be favored, being zero the threshold that separates them.(35) The molecular surface lipophilicity potential was obtained with the Chimera software and a range between [-20, 20] defined with a color code; negative hydrophilic values are represented in green, while the positive hydrophobic are represented in gold.(36)

2.4.4. 3D Alignment Structural Viewers

This 3D alignment is intended to obtain the coincidence structurally in a significant part of the spatial distribution of both models. The spatial coordinates and the selection of the molecular regions were made with the Deepview/Swiss-PdbViewer software version 3.7.(37) The Chimera UCSF v 1.1.1 software was used for visualization, and the structural alignments were made with the Needleman-Wunsch global algorithm and the BLOSUM62 matrix. The quantitative alignment to find the RMSD value in the structural comparison was made using PyMol.(36,38,39)

2.3.5. Simulation by the Hybrid Quantum Mechanics-Molecular Mechanics QM-MM Method

To better quantify the Phe233 mutation effects at the atomic functional resolution, we applied a combination of hybrid quantum chemistry theory (focusing on the deletion vicinity region) and molecular mechanics force field theory (applied to other PURA regions far from the vicinity of the Phe233 deletion). Both theories amalgamated in the QM/MM semi-empirical method (preferred for macromolecular systems).(40–44) We used the model's parameterization with the *Z* matrix of connectivity and experimental data derived from empirical crystallographic studies, focusing on those targeting the Phe233del position. Thus far, the quantum analysis (QM) focuses on the antiparallel β-sheets of the secondary structure, which contemplates the amino acids Asn228-Thr252 for the wild type

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

and Asn228-Tyr252 for the mutated structure. This region was delimited from the saturation with hydrogen atoms using the Austin model1 (AM1) (base of the software spartan20' of Wavefunction acquired license, https://www.wavefun.com/corporate/more_spartan.html) to measure, from the electronic correlation, changes of bond distances and angles, as well as energetic changes.(44–46) The QM region also undergoes a molecular mechanics $(MM)_{QM}$ approach with the $MMFF_{aq}$ forcefield to better understand the second solvation sphere, given that the surrounding targeted region is a polyatomic system. The rest of the structure undergoes an approach with classical physics from molecular mechanics (MM) with the $MMFF_{aq}$ base, which favors systems that can be measured in terms of the properties from the electronic correlation at the atomic level, *i.e.*, energy changes, bond distances and changes in the angles.(45,47) In the simulation, a geometric optimization was performed, and the lowest energy conformer was obtained, which serves as a model for calculating the surface properties of the electronic structure (electrostatic potential maps in ranges from -200kJ/mol to 200kJ/mol). This allows us to understand the electronic distribution and the possibility of interacting or anchoring with non-covalent interactions in the multiscale model.(45–47)

3. Results

3.1. WES Bioinformatics

We found a *de novo* (no present in the parents) heterozygous mutation in the proband, c.697_699del (p.Phe233del), NM_005859.5, PURA (MIM*600473) (Het, Autosomal dominant mental retardation 31)(MIM #614563), located at Chr5:139,494,453-139,494,455 (rs786204835)(additional data regarding next-generation sequencing quality parameters are presented in **Table 1S** of the **Supplementary Material**). This deletion of a complete codon located from position 697 to 699 of the cDNA causes the deletion of a phenylalanine at residue 233 of the protein (inframe), which, according to predictions, affects a functional domain of the protein (**Figure 1a**).

3.2. Protein Reconstruction

The 3D reconstruction of the PURAwt structure following the prediction model generated by the AlphaFold resulted in a very high per-residue estimate of its confidence **(**pLDDT) score > 90 within the Phe233 area (β-sheet), as presented in the β-sheet composition with three Phenylalanine's expansion (**Figure 1b**). The PyMol software showed the composition of the β-sheet containing the Phe233. In agreement with protein–DNA binding interactions, protein structures, like the protein region containing the deletion, may induce a deformation of the DNA due to binding in specific regions of the DNA (48). For instance, it was binding to the minor groove that, in addition to electrostatic interactions and other non-covalent interactions, may also play an essential role in generating such deformation. This will allow the transcription initiation to generate mRNA, *i.e.*, AlphaFold analysis of the region harboring the Phe233 change showed its hydrogens in a 3D conformation (**Figure 1c**) and the Phe233 polar contacts measurements less than 4Å (**Figure 1d**). **Table 2S** of the **Supplementary Material** summarizes the evaluation parameters and scores obtained by different software on the available models.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

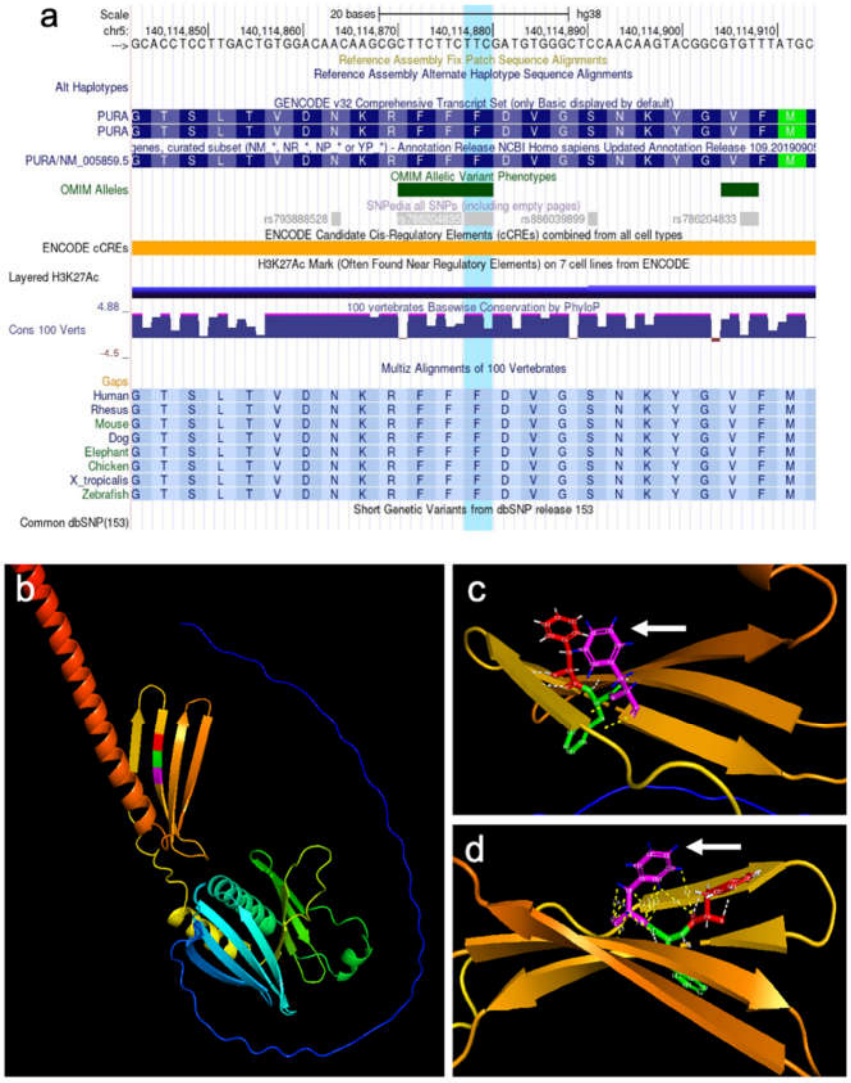

Figure 1. a. *de novo* heterozygous mutation found in the proband, c.697_699del (p.Phe233del), NM_005859.5, PURA (MIM*600473), located at Chr5:139,494,453-139,494,455 (rs786204835). This deletion causes the loss of Phenylalanine at residue 233 of the protein (in-frame), codon located from position 697 to 699 of the cDNA. **b.** β-sheet composition with three Phenylalanine expansions: Phe231 (red), Phe232 (green), and Phe 233 (magenta). **c.** Phe233 shows its hydrogens in a #D conformation (white arrow). **d.** Phe233 polar contact measurements are less than 4Å (white arrow).

Additionally, a global alignment with the Needleman Wunsch algorithm and a standard deviation value (RMSD = 0.595) was performed by zooming in the region near position 233 of the wild type R245-D234 protein region (translated from the PURA gene) to evaluate the effect of the deletion and potential changes in the angle of the β-sheets of PURAwt once the PURA-Phe233del occurred (**Figure 1S-a**, **Supplementary Material)**. The acid-base interactions folding the β-sheets with a hydrogen bond molecular interaction at a distance of 1.911Å in the wild type of protein (**Figure 1S-b**) favors interactions between the aspartic acid with arginine by a non-bonding acid-base interaction between R245-D233 decreasing the distance to 1.768Å, once the deletion is present (**Figure 1S-c**).

3.2.1. PURA Protein-DNA Docking in silico Analysis

Interestingly, modeling of the PURAwt protein by AlphaFold and posterior visualization of the three-principal purine-rich element-binding (PUR) interacting with its three PURA RNA/DNA binding domains (**Figure 2a**). It shows a spanner head containing the c-Myc DNA sequence of three different HDOCK generated docking models from the top

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

10 (see DNA in dark blue, light blue, and white). Residues responsible for DNA contact are marked in white (see white arrows) (**Figures 2b** and **2c**). See; specifically, the modification of the residues marked in white, losing contact with the DNA, in the case of the PURA-Phe233del model when compared to the PURAwt.

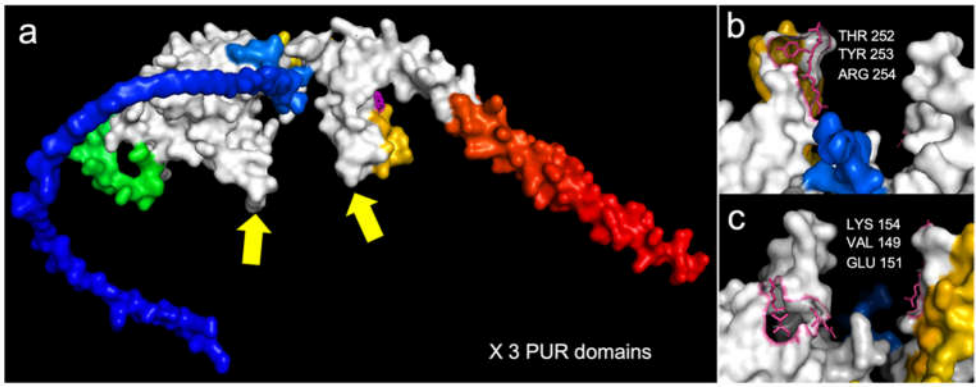

Figure 2. (a) PURAwt AlphaFold modeling, showing the three-principal purine-rich (PUR) elements in white. Yellow arrows show the DNA docking site regions forming a spanner head that will interact with the DNA. **(b)** Zoom of the protein-DNA docking site with three principal residues in contact with the DNA molecule (*c-MYC* promoter region), THR 252, TYR 253, and ARG 254. **(c)** Zoom and 180-degree rotation to visualize the opposite protein-DNA docking site with its three principal residues in contact with the DNA molecule (*c-MYC* promoter region), VAL 149, GLU 151, and LYS 154.

Predicted PURAwt and PURA-Phe233del by AlphaFold, were docked to previously reported DNA promoter regions of the c-MYC and MBP human genes.(31)(27,28) In silico HDOCK results validated previously experimentally reported interaction between PURA and the *c-MYC* and *MBP* DNA promoter region sequences. Ten top *in silico* models aligned perfectly within the PURAwt model (**Figure 3a**), whereas docking between PURA-Phe233del and *c-MYC* and *BMP* sequences showed a disruption in the top 10 HDOCK in silico models (**Figure 3b**).

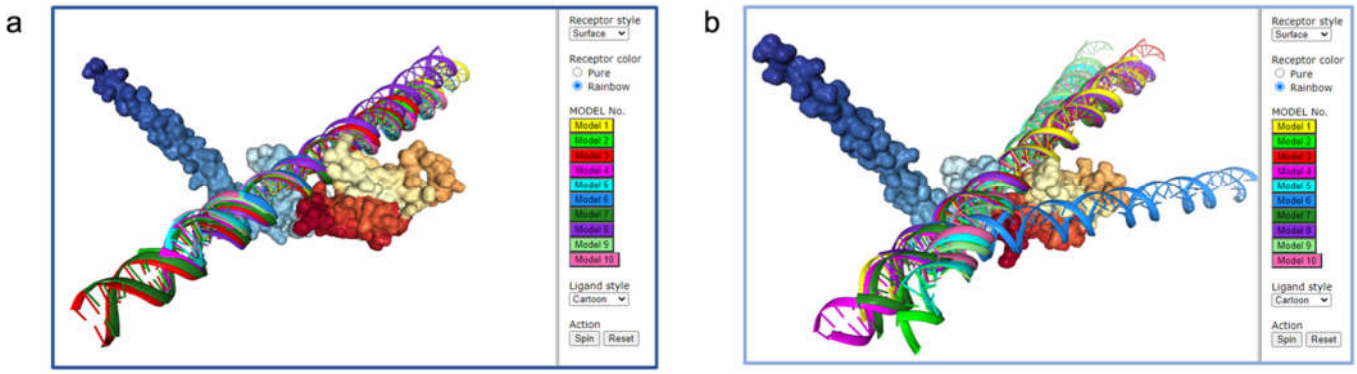

Figure 3. The PURA-DNA comparative analyses show DNA docking damage PURA Protein-DNA Docking. **(a)** *In silico* analysis by HDOCK shows 10 top models between PURA-WT AlphaFold and previously reported DNA promoter regions of the c-Myc. **(b)** HDOCK docking analysis between AlpahFold predicted PURA-Phe233del protein and the c-Myc DNA promoter regions.

3.2.2. Energetic and Stereochemical Characterization of the PURA molecular Models

The structures at the energetic level have QMEAN6 values within the normalized ranges of [0-1], presenting larger energetic values for PURA wild type 0.61 compared to the structure with the deletion 0.59; likewise, they have Z-score values close to the scattering values of the structure plot, contemplating experimentally solved structures with values of -2.41 (PURAwt) and -3.60 (PURA Phe233del). Additional data is shown in **Figure 2S** of the **Supplementary Material**.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

Considering the statistical values and the structure of amino acids for the local groupings of non-bonding interactions, the Errat software estimated high scores, close to 93.89% for PURAwt and 89.66% for PURA Phe233del. At the stereochemical level, measurements of the peptide bonds alpha-carbons-dihedral angles (Cα), using the Ramachandran plots, showed for the PURAwt 219 amino acids in the favored region (84.6%), 23 in the allowed region (8.9%), four in the generously allowed region (1.5%), and 23 in the disallowed region (5.0%). In contrast, for the PURA Phe233del, there are 216 amino acids in the favored region (83.7%), 30 in the allowed region (11.6%), three in the generously allowed region (1.2%), and 9 in the disallowed region (3.5%) (**Figure 3S** of the **Supplementary Material**).

3.2.4. Calculation of Hydropathic Indices and Polarity Surfaces of PURAwt and Mutated-Phe233del

The lipophilicity potential structures show that Phe233 is harbored in a golden region (hydrophobic), indicating that it plays a significant role in that three-dimensional arrangement (**Figure 4a**). Once the Phe233del occurs, the lipophilicity potential (polarity of the protein environment) changes, so the whole environment near the deletion is made green, and the polarity increases, which induces a change in the structure (**Figure 4b**). It can also be observed that the polarity increased in the vicinity of the deletion since more negative values of the environment are obtained for the Kyte and Dolittle coefficients. Indeed, the polarity is -0.056 at position PURAwt, which changes to -0.756 (Asp233)(**Figure 4c**).

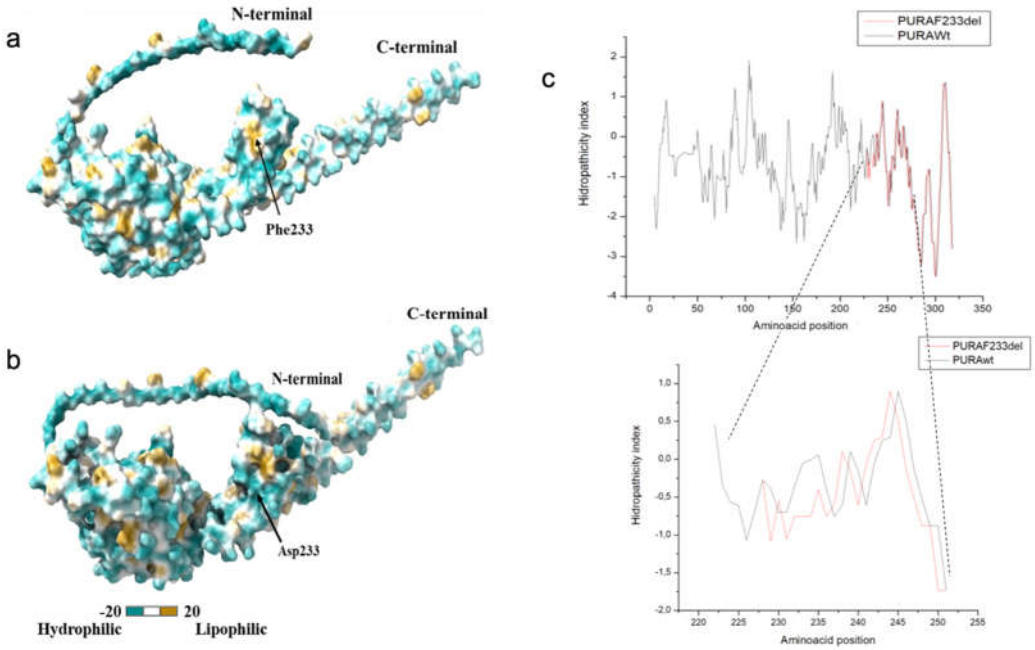

Figure 4. Hydrophobicity analysis (Molecular Lipophilicity Potential, MLP maps) of the **(a)** PURAwt, and **(b)** PURAPhe233del system. The analysis shows that in the wild type variant, the Phe 233 is harbored in a golden region (hydrophobic), implicating that it plays a significant role in the 3D arrangement, while the hydrophobicity changes once the deletion occurs. This is also appreciated in **(c)**, which depicts the Hydrophobicity index plot zooming in the 233 amino acid region (see text).

3.2.5. Simulation by the Hybrid Quantum Mechanics-Molecular Mechanics (QM-MM) Method.

For the PURAwt, the estimated bond angle between amino acids F232-F233 was 37.57°, while for F233-D234, a bond angle of 113.84° was obtained, which distances the reference points on the CH2 carbons constituting the side chain by 7.215Å (reference distance of the amino acids in the β-sheets) (**Figure 5a**). Once the deletion occurs, the bond angles between F231-F232 are 35.51°, and the bond angle between F232-D233 changes to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

112.47°, indicating that for the reference structure, the binding distance decreases to 6.103Å, which affects the accessible area of the structure (**Figure 5b**). The electrostatic potential maps were calculated for the β-sheets involved in the region undergoing variation, and an area and volume values for the electronic region were obtained; the area and volume values for the electronic region analyzed using quantum chemistry were 3032.37 Å^2 and 2930.94 Å^3, respectively (**Figure 5c**). Upon the occurrence of the Phe233 deletion, the area value in the electronic region changes to 2852.30 Å^2 and the volume to 2773.77 Å^3, which is in agreement with the loss of enough electronic topology contributed by the Phe233 amino acid and its aromaticity (**Figure 5d**).

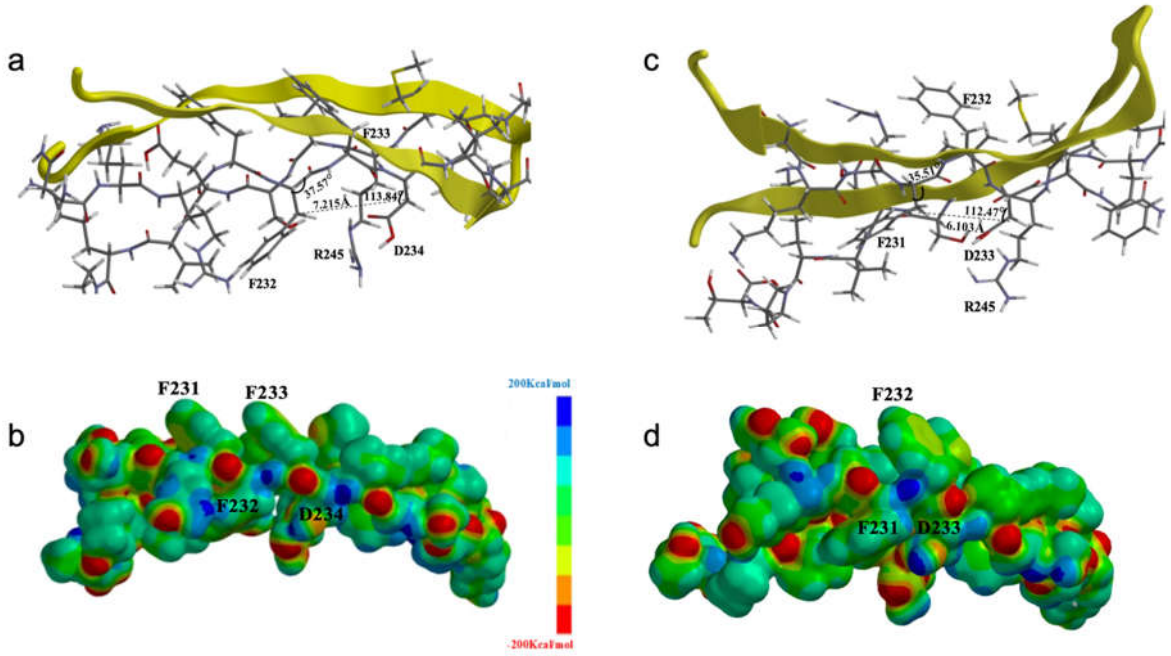

Figure 5. Functional comparison of the connective and electronic structures for the QM/MM simulation of PURAwt and PURA-Phe233del. **a)** PURAwt, the angle between amino acids F232-F233 was 37.57°, while for F233-D234, a bond angle of 113.84° distancing the CH2 carbons constituting the side chain by 7.215Å (reference distance of the amino acids in the β-sheets). **b)** In PURA-Phe233del, the bond angles between F231-F232 are 35.51°, and the bond angle between F232-D233 changes to 112.47°, indicating that for the reference structure, the binding distance decreases to 6.103Å, which affects the accessible area of the structure. **c)** The electrostatic potential maps of the PURAwt β-sheets involved, area, and volume values for the electronic region in the region undergoing variation were 3032.37 Å^2 and 2930.94 Å^3, respectively. **d)** The electrostatic potential maps of the PURA-Phe233del β-sheets involved, area, and volume values for the electronic region in the region undergoing variation change to 2852.30 Å^2 and the volume to 2773.77 Å^3, which is in agreement with the loss of enough electronic topology contributed by the Phe233 amino acid and its aromaticity.

4. Discussion

Variants in the *PURA* gene have been linked to autosomal dominant mental retardation 31, which has an autosomal dominant inheritance pattern and correlates with the patient's phenotype. This *de novo* variant fits the PS2 criteria of the ACMGguidelines and is an in-frame deletion of a highly conserved residue of a functional domain of the protein (PM1) that has no allele frequency (PM2).

The PURA protein was first described to bind to ssDNA in cell culture and brain extracts (27,31)(31) PURA was shown to play a role in cellular transcriptional activity.(27,31)(31) The *PURA* gene has been reported as part of the 3% of human genes that lack introns.(13) It seems that this gene cannot generate alternative splicing. Evolution has tagged this genomic region as inalterable, which may prove its unique and essential role in interacting with DNA and RNA molecules. It has been reported that any mutation in the PURA protein causes the full spectrum of the human PURA syndrome.(49)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

Predicted PURAwt and PURA Phe233del prediction models derived with AlphaFold, showed a close molecular interaction between the Phenylalanine´s molecules. Hence, changes in this structural conformation may alter PURA protein's biding and transcription initiation capacities. This may suggest that salt-bridge interactions between the phosphate groups and Lys or Arg residues, along with the intercalation of Phe residues between two base-pair stacks, stabilize and opened-up DNA conformation, which can be altered in the PURA Phe233del mutation. Furthermore, the design of super-helical nanostructures by minimal molecular elements has demonstrated the importance of Phenylalanine as molecular blocks, establishing the structure of the phenylalanine zipper, which consists of aromatic side chains from ten phenylalanine residues that are stacked within a hydrophobic core (27). This zipper has been demonstrated to mediate the dimerization of various proteins, including APS, SH2-B, and Lnk (28).

The analysis of the AlphaFold prediction by the Local Distance Difference Test (pLDDT), for the PURAwt and the PURA-Phe233del predicted structure deletion per amino acid, using five different models via AlpahFold2 models, showed that the occurrence of the Phe233del affects between 220-320 amino acids (red arrow) (**Figures 6a** and **b**). Additionally, the 3D reconstruction of the transcriptional activator protein Pur-alpha (PURA; Uniprot: Q00577) from the wild sequence using Alpha fold, and the 3D reconstruction of the PURA-Phe233del (**Figures 6c** and **d**), with a focus on the surrounding area of the Phe233 amino acid (in magenta), disclose the distortion in the PURA structural conformation of the ~5Å surrounding area after the PURA-Phe233del occurs (double head arrows), with a conspicuous disruption of the repeat III.

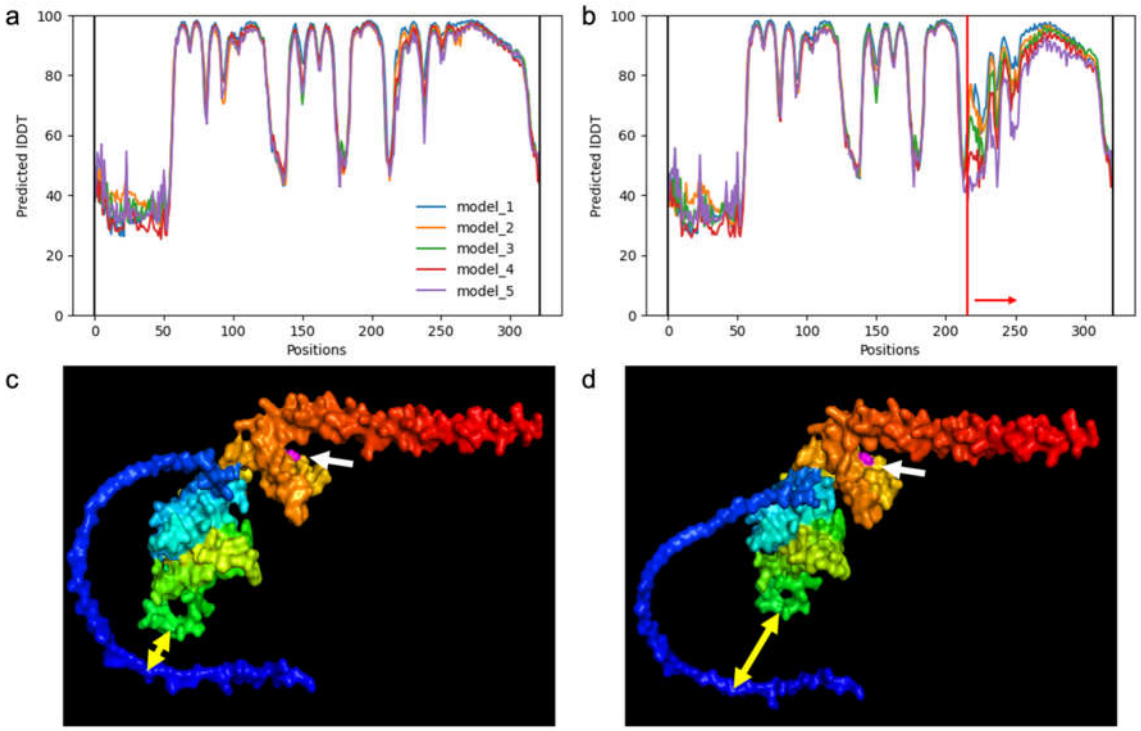

Figure 6. (a) Predicted local Distance Difference Test (lDDT) for the PURA WT and **(b)** PURA-Phe233del predicted structure deletion per amino acid using five different models via AlpahFold2. Models show that the occurrence of the Phe233del affects between 220-320 amino acids (red arrow). **(c)** 3D reconstruction of the transcriptional activator protein Pur-alpha (PURA; Uniprot: Q00577) from the wild sequence using Alpha fold, focusing on the surrounding area of the Phe233 amino acid (in magenta). **(d)** 3D reconstruction of the PURA-Phe233del. Double head arrows highlight the distortion in PURA structural conformation in the ~5Å surrounding area after the p.Phe233del occurs. The effect of the mutation shows a loss of the repeat III.

The Protein PURA has been shown to bind DNA and RNA, implicated in these molecules' transcriptional and intracellular localization. PURA protein molecular variations

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

have been linked to several neurodevelopmental disorders (13,53,54)(13). In this manuscript, we showed that the HDOCK modeling of PURAwt aligns perfectly with the c-Myc DNA promotor region, whereas alteration in PURA-Phel233del causes an impairment in the DNA binding capacity of PURA (**Figures 7 a and b**).

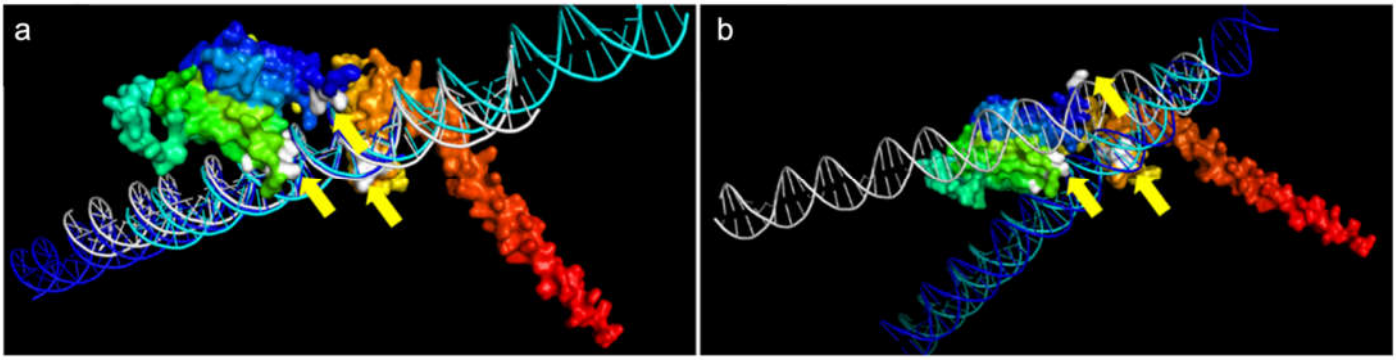

Figure 7. PURA-Phe233del disrupts DNA biding. **(a)** A molecular view of PURA Protein-DNA Docking analysis showing three different HDOCK models between PURAwt AlphaFold and the *c-Myc* DNA promoter region. See yellow arrows indicating specific docking sites in white. **(b)** HDOCK docking analysis shows the c-Myc DNA binding disruption caused by the AlpahFold2 predicted PURA-Phe233del protein. Yellow arrows indicate disruption of specific docking sites (in white).

These results provide new evidence, validating previous reports, which is significant considering that most studies of PURA molecular interaction were performed before the X-ray structure of PURA was available. Moreover, reported interactions were disclosed with methods that did not distinguish between direct or indirect interactions.(55) Thus far, *in silico* protein prediction and molecular interaction tools such as docking are essential to investigate further whether those interactions represent direct binding events and identify their physiological relevance.

The energetic and stereochemical molecular modeling of the PURAwt and the mutated PURA-Phe233del showed good scores. The QMEAN6 is a method conceived as a linear combination of stability terms involving energy of the torsion angles, coupling energy of atom pairs, solvation energy, beta carbon energy, secondary structure agreement, and solvent accessibility agreement. These parameters generated a high global value for the two structures in terms of stability. Likewise, the *Z*-score parameter is a term that allows a comparison of the predicted hypothetical models to those experimentally obtained by X-ray diffraction techniques with a monochromator, NMR nuclear magnetic resonance, and CryoEM cryogenic imaging microscopy. We can observe that both structures leave the model located in the trend of the graph, as shown in **Figure 1S** of the **Supplementary Material.** The Errat software has a statistical estimation algorithm that allows to group nine amino acids and performs a characterization, following along the whole primary sequence, to generate the summation of a value for the global of the proteins, producing a very high score for the two hypothetical models obtained.

In stereochemical terms, the measurement of the spatial behavior of the models was made from the dihedral angles considering the alpha carbons Cα), belonging to the side chain, measuring the residues belonging to specific regions within the Ramachandran graph, dependent on the twists and angles of the secondary structure, they are adopting. These values depend on the movement of the angles Phi (φ) and Psi (ψ), whose rotations are measured with forbidden rotations by the amino acids glycine and proline located in the N-terminal and C-terminal loops. Likewise, the values obtained show a total percentage for PURAwt of 95% in the favored regions and 96.5% for PURA-Phe233del since it

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

loses an amino acid in an antiparallel beta-sheet folding. This occurs because it has hydrogen bonding interactions every 180° for every 1-3 amino acid that belongs to that secondary structure and becomes a flexible loop, releasing forbidden conformations.

After analyzing the polarity changes and their effect at the structural level (Kyte and Dolittle coefficients) and using the Needleman Wunsch algorithm to visualize the effect at the 3D level, we concluded that a twisting effect, mediated by the Phe233del deletion, produces a practical approach of the antiparallel β-sheets by non-covalent interactions, *i.e.*, hydrogen bonding that decreases the interaction distance with the variant causing a structural torsion. Consequently, these chemical effects favor the formation of adducts (the interaction distance between the amino acids D233 and R245 is decreased), causing a change in the torsion angle in the F232-D233 peptide bond. This topological effect alters the electronic surface of the antiparallel β-sheet (since an amino acid with a high electron density is lost) that provides more access to ligands, allowing the recognition of the nucleotide chains associated with the III region of the DNA recognition domain. Our conclusion is corroborated by the area and volume decrease of 175.07Å2 and 157.17Å3, respectively. This change is quantified with the accessible area that a ligand can have, and this exposed area goes from 1071.62 Å2 in the case of the PURAwt to 990.97 Å2 in the case of the PURA-Phe233del mutated protein, which affects the protein functionality.

Author Contributions: Conceptualization, JJ.L-R. and M.A-B.; methodology, JJ.L-R., C.E-S., D.S., HM.C-S., O.L., JI.V., OM.V., LR, MA.I-R., M.A-B.; software, JI.V., OM.V., L.R-S., PA.R, G.R., A.S-O., A.V-L., M.A-B.; .validation, JJ.L-R., CE, DS, JI.V., OM.V., L.R-S, M.A-B.; formal analysis, JI.V., OM.V., L.R-S, A.S-O., A.V-L., M.A-B.; investigation, JJ.L-R., CE, DS, JI.V., OM.V., L.R-S, MA.I-R., M.A-B.; resources, JJ.L-R., M.C., M.G-C., MA.I-R., M.A-B.; data curation, JI.V., OM.V., L.R-S., PA.R, G.R., M.G-C., M.A-B.; writing—original draft preparation, JJ.L-R. and M.A-B.; methodology, JJ.L-R., CE, DS, JI.V., OM.V., L.R-S., PA.R., G.R., A.S-O., A.V-L., M.A-B.; writing—review and editing, JJ.L-R., JI.V., OM.V., L.R-S, M.C., M.A-B.; visualization, OM.V., L.R-S., PA.R, G.R., A.S-O., A.V-L., M.A-B.; supervision, JJ.L-R. and M.A-B.; project administration, JJ.L-R. and M.A-B.; funding acquisition, JJ.L-R., JI.V., MA.I-R., M.A-B. All authors have read and agreed to the published version of the manuscript.

Funding: Keralty (https://www.keralty.com/) funded this research.

Institutional Review Board Statement: Informed consent was obtained from all individual participants included in the study. Statement on Human Ethics The procedures performed in our study involving human participants were following the ethical standards of the institutional research committee **(Comité de Ética COLSANITAS, Act Number. 014-22, see supplementary information),** and the 1964 Helsinki declaration and its later amendments. All participants provided informed consent, and parents provided informed consent for their daughter (younger than 16 years of age).

Acknowledgments: Our most profound acknowledgment to the patient and her parents.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the study's design, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Tanaka AJ, Bai R, Cho MT, Anyane-Yeboa K, Ahimaz P, Wilson AL, et al. De novo mutations in PURA are associated with hypotonia and developmental delay. Molecular Case Studies. 2015;1(1).
2. S.R. L, J. Z, C.P. S, C.W. B, P. M, A.C.-H. T, et al. Mutations in PURA cause profound neonatal hypotonia, seizures, and encephalopathy in 5q31.3 microdeletion syndrome. American Journal of Human Genetics. 2014;95(5).
3. Hunt D, Leventer RJ, Simons C, Taft R, Swoboda KJ, Gawne-Cain M, et al. Whole exome sequencing in family trios reveals de novo mutations in PURA as a cause of severe neurodevelopmental delay and learning disability. Journal of Medical Genetics. 2014;51(12).
4. Tanaka AJ, Bai R, Cho MT, Anyane-Yeboa K, Ahimaz P, Wilson AL, et al. De novo mutations in PURA are associated with hypotonia and developmental delay . Molecular Case Studies. 2015;1(1).
5. S.R. L, J. Z, C.P. S, C.W. B, P. M, A.C.-H. T, et al. Mutations in PURA cause profound neonatal hypotonia, seizures, and encephalopathy in 5q31.3 microdeletion syndrome. American Journal of Human Genetics. 2014;95(5).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

6. Hunt D, Leventer RJ, Simons C, Taft R, Swoboda KJ, Gawne-Cain M, et al. Whole exome sequencing in family trios reveals de novo mutations in PURA as a cause of severe neurodevelopmental delay and learning disability. Journal of Medical Genetics. 2014;51(12).
7. Khalili K, del Valle L, Muralidharan V, Gault WJ, Darbinian N, Otte J, et al. Pur Is Essential for Postnatal Brain Development and Developmentally Coupled Cellular Proliferation As Revealed by Genetic Inactivation in the Mouse. MOLECULAR AND CELLULAR BIOLOGY,. 2003;23(19):6857–6875.
8. Wortman MJ, Hanson LK, Martínez-Sobrido L, Campbell AE, Nance JA, García-Sastre A, et al. Regulation of PURA gene transcription by three promoters generating distinctly spliced 5-prime leaders: a novel means of fine control over tissue specificity and viral signals. BMC Molecular Biology. 2010;11.
9. White MK, Johnson EM, Khalili K. Multiple roles for Purα in cellular and viral regulation. Vol. 8, Cell Cycle. 2009.
10. Hokkanen S, Feldmann HM, Ding H, Jung CKE, Bojarski L, ller IRM, et al. Lack of pur-alpha alters postnatal brain development and causes megalencephaly. Human Molecular Genetics. 2012;21(3).
11. Khalili K, Del Valle L, Muralidharan V, Gault WJ, Darbinian N, Otte J, et al. Pur Is Essential for Postnatal Brain Development and Developmentally Coupled Cellular Proliferation As Revealed by Genetic Inactivation in the Mouse. MOLECULAR AND CELLULAR BIOLOGY,. 2003;23(19):6857–6875.
12. Hokkanen S, Feldmann HM, Ding H, Jung CKE, Bojarski L, ller IRM, et al. Lack of pur-alpha alters postnatal brain development and causes megalencephaly. Human Molecular Genetics. 2012;21(3).
13. Reijnders MRF, Janowski R, Alvi M, Self JE, van Essen TJ, Vreeburg M, et al. PURA syndrome: Clinical delineation and genotype-phenotype study in 32 individuals with review of published literature. Journal of Medical Genetics. 2018;55(2).
14. PURA Gene. In: Definitions. 2020.
15. Okamoto N, Nakao H, Niihori T, Aoki Y. Patient with a novel purine-rich element binding protein A mutation. Congenital Anomalies. 2017;57(6).
16. Shimojima K, Okamoto N, Ohmura K, Nagase H, Yamamoto T. Infantile spasms related to a 5q31.2-q31.3 microdeletion including PURA. Human Genome Variation. 2018;5.
17. Bonaglia MC, Zanotta N, Giorda R, D'Angelo G, Zucca C. Long-term follow-up of a patient with 5q31.3 microdeletion syndrome and the smallest de novo 5q31.2q31.3 deletion involving PURA. Molecular Cytogenetics. 2015;8(1).
18. Qiao Y, Bagheri H, Tang F, Badduke C, Martell S, Lewis SME, et al. Exome sequencing identified a de novo mutation of PURA gene in a patient with familial Xp22.31 microduplication. European Journal of Medical Genetics. 2019;62(2).
19. Okorie U, Monceaux B, Smalley M, Roberts E, Liendo C, Chernyshev O. 1233 Mitochondrial Myopathy Making It Hard To Sleep! OSA management in Mitochondrial Myopathy with a variant in SNAPC4 and PURA genes. Sleep. 2020;43(Supplement_1).
20. Spangenberg L, Guecaimburú R, Tapié A, Vivas S, Rodríguez S, Graña M, et al. Novel frameshift mutation in PURA gene causes severe encephalopathy of unclear cause. Molecular Genetics and Genomic Medicine. 2021;9(5).
21. Mayorga L, Gamboni B, Mampel A, Roqué M. A frame-shift deletion in the PURA gene associates with a new clinical finding: Hypoglycorrhachia. Is GLUT1 a new PURA target? Molecular Genetics and Metabolism. 2018;123(3).
22. Rodríguez-García ME, Cotrina-Vinagre FJ, Arranz-Canales E, de Aragón AM, Hernández-Sánchez L, Rodríguez-Fornés F, et al. A novel de novo mutation in the PURA gene associated with a new clinical finding: large brainstem. Journal of Genetics. 2020;99(1).
23. Reijnders MRF, Janowski R, Alvi M, Self JE, Van Essen TJ, Vreeburg M, et al. PURA syndrome: Clinical delineation and genotype-phenotype study in 32 individuals with review of published literature. Journal of Medical Genetics. 2018;55(2).
24. PURA Gene. In: Definitions. 2020.
25. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. AlphaFold. Nature. 2020;577(7792).
26. Yan Y, Tao H, He J, Huang SY. The HDOCK server for integrated protein–protein docking. Nature Protocols. 2020;15(5).
27. Bergemann AD, Johnson EM. The HeLa Pur factor binds single-stranded DNA at a specific element conserved in gene flanking regions and origins of DNA replication. Molecular and Cellular Biology. 1992;12(3).
28. Bergemann AD, Ma ZW, Johnson EM. Sequence of cDNA comprising the human pur gene and sequence-specific single-stranded-DNA-binding properties of the encoded protein. Molecular and Cellular Biology. 1992;12(12).
29. Wortman MJ, Johnson EM, Bergemann AD. Mechanism of DNA binding and localized strand separation by Purα and comparison with Pur family member, Purβ. Biochimica et Biophysica Acta - Molecular Cell Research. 2005;1743(1–2).
30. Ma ZW, Bergemann AD, Johnson EM. Conservation in human and mouse Purα of a motif common to several proteins involved in initiation of DNA replication. Gene. 1994;149(2).
31. Haas S, Gordon J, Khalili K. A developmentally regulated DNA-binding protein from mouse brain stimulates myelin basic protein gene expression. Molecular and Cellular Biology. 1993;13(5).
32. Studer G, Biasini M, Schwede T. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). In: Bioinformatics. 2014.
33. Benkert P, Tosatto SCE, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. Proteins: Structure, Function and Genetics. 2008;71(1).
34. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 1993;26(2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2022 doi:10.20944/preprints202205.0280.v1

35. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. The Proteomics Protocols Handbook - Chapter 52: Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook. 2005.
36. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry. 2004;25(13).
37. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis. 1997;18(15).
38. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. Journal of Molecular Graphics. 1996;14(1).
39. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. Vol. 26, Journal of Computational Chemistry. 2005.
40. Soto-Ospina A, Marín PA. Redox mechanism of Trypanosoma cruzi Resistance to Nitro Prodrugs Benznidazole and Nifurtimox. International Journal of Bioinformatics and Computational Biology. 2020;5(1).
41. Soto-Ospina A, Araque Marín P, Bedoya GDJ, Villegas Lanau A. Structural Predictive Model of Presenilin-2 Protein and Analysis of Structural Effects of Familial Alzheimer's Disease Mutations. Biochemistry Research International. 2021;2021.
42. Soto-Ospina A, Araque Marín P, Bedoya G, Sepulveda-Falla D, Villegas Lanau A. Protein Predictive Modeling and Simulation of Mutations of Presenilin-1 Familial Alzheimer's Disease on the Orthosteric Site. Frontiers in Molecular Biosciences. 2021;8.
43. Cano L, Soto-Ospina A, Araque P, Caro-Gomez MA, Parra-Marin MV, Bedoya G, et al. Diffusion Mechanism Modeling of Metformin in Human Organic Cationic Amino Acid Transporter one and Functional Impact of S189L, R206C, and G401S Mutation. Frontiers in Pharmacology. 2021;11.
44. Ahmadi S, Barrios Herrera L, Chehelamirani M, Hostaš J, Jalife S, Salahub DR. Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review. Vol. 118, International Journal of Quantum Chemistry. 2018.
45. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. J Am Chem Soc. 1985;107(13).
46. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. AM1: A New General Purpose Quantum Mechanical Molecular Model1. J Am Chem Soc. 1985;107(13).
47. Foresman J, Frisch a. Exploring chemistry with electronic structure methods, 1996. Gaussian Inc, Pittsburgh, PA. 1996.
48. Baker CM, Grant GH. Role of aromatic amino acids in protein-nucleic acid recognition. Biopolymers. 2007;85(5–6).
49. Molitor L, Bacher S, Burczyk S, Niessing D. The Molecular Function of PURA and Its Implications in Neurological Diseases. Vol. 12, Frontiers in Genetics. 2021.
50. Mondal M, Mukherjee S, Bhattacharyya D. Contribution of phenylalanine side chain intercalation to the TATA-box binding protein–DNA interaction: molecular dynamics and dispersion-corrected density functional theory studies. Journal of Molecular Modeling. 2014;20(11).
51. Mondal S, Adler-Abramovich L, Lampel A, Bram Y, Lipstman S, Gazit E. Formation of functional super-helical assemblies by constrained single heptad repeat. Nature Communications. 2015;6.
52. Dhe-Paganon S, Werner ED, Nishi M, Hansen L, Chi YI, Shoelson SE. A phenylalanine zipper mediates APS dimerization. Nature Structural and Molecular Biology. 2004;11(10).
53. Trau SP, Pizoli CE. PURA Syndrome and Myotonia. Pediatric Neurology. 2020;104.
54. Jia P, Wang L, Meltzer HY, Zhao Z, Lee PH, O'dushlaine C, et al. Pervasive Sharing of Genetic Effects in Autoimmune Disease. Nature Genetics. 2013;41(1).
55. Grzybowska EA. Human intronless genes: Functional groups, associated diseases, evolution, and mRNA processing in absence of splicing. Vol. 424, Biochemical and Biophysical Research Communications. 2012.