

Full Length Research Paper

Structure modeling and mutational analysis of gap junction beta 2 (GJB2)

Samina Bilal^{1,3}, Hamid Rashid¹, Jabar Zaman Khan Khattak², Shehzad Ashraf Ch³, Tahira Sultana⁴ and Asif Mir^{2*}

¹Department of Bioinformatics, Mohammad Ali Jinnah University, Islamabad, Pakistan.

²Department of Bioinformatics and Biotechnology, International Islamic University, Islamabad, Pakistan.

³Department of Computer Science and Software Engineering, International Islamic University, Islamabad, Pakistan.

⁴Department of Environmental Sciences, International Islamic University, Islamabad, Pakistan.

Accepted 22 March, 2012

The genome sequencing accomplishes complete genetic blue prints for hundreds of organisms, including humans. In the current era, we are trying to focus on analyzing, controlling and modifying functions of proteins encoded by these genomes. This task is attained by protein three dimensional structures. Three dimensional (3 D) structure is very useful for understanding biological functions. Gap junction beta 2 (GJB2), human gene encoding for gap junction beta 2 protein is involved in various hearing disorders in Pakistani families. After the first report of GJB2 involvement in Pakistani families, it was necessary to further study this protein. Therefore, a 3D structure of GJB2 was developed using comparative modeling approach. For modeling, a template was selected by blastp at NCBI and the best template selected was 2ZW3. By comparing the template-target sequence, a model was created using MODELLER, a program for homology modeling. The accuracy of the predicted structure was checked using Ramachandran plot which showed that the residue falling in the favored region was 92.4%. The predicted GJB2 model can be used to understand the defects that lead to deafness and eventually in drug designing. Domains and different properties of GJB2 were analyzed by applying online servers. Most frequent mutations of GJB2 were discussed by differentiating between damaging and benignity.

Key words: GJB2, 3D structure, 2ZW3, DFNB1, MODELLER.

INTRODUCTION

Deafness is a state in which aptitude to perceive certain frequency of sound is completely or partially impaired. In humans, the term hearing impairment is usually reserved for people who have virtual insensitivity to sound in the speech frequencies. Research in deafness became real necessity in our common life (Karen, 2000; Sterkers et al., 1982). Indeed, this is a real and serious problem causing social disintegration of a wide percent of the active society. Hearing impairment is the result of abnormal ear development, abnormal ear function or both. It is the most common sensory disorder affecting 1 in 1000 newborns world-wide (Cohen and Gorlin, 1995). It is estimated that the prevalence of profound bilateral

hearing loss is 1.6 per 1000 in Pakistani population. (http://www.jpma.org.pk/full_article_text.php?article_id=2089).

More than 100 loci have been associated with the nonsyndromic hearing loss, with majority of the cases (~80%) being autosomal recessive in inheritance (Cryns et al., 2004). DFNB1 at 13q12 is the first locus identified for hearing impairment and is subsequently focused more because of its complexity and clinical relevance. It harbors the gene GJB2 (Denoyelle et al., 1997; Green et al., 1999), a human gene encoding for Gap junction protein, beta 2, 26kDa or Connexin 26. Defects in this gene lead to the most common form of congenital deafness in developed countries, called DFNB1, also known as Connexin 26 deafness or GJB2-related deafness. The GJB2 gene is a member of the gap junction or connexin family. This family of genes produces protein

*Corresponding author. E-mail: mir77uspk@gmail.com.

Table 1. Percentage similarity between target and template sequence.

Model number	Tool used	Template	Similarity (%)	Number of residues modeled
1	Modeller	2ZW3	100	226
2	Modeller	1SJ2	28	226
3	Swiss-Model	2ZW3E	93	216
4	Swiss PDB viewer	2ZW3	100	225
5	ESyred3d	2ZW3A	100	216
6	CPH models	2ZW3A	93	216

subunits for channels (gap junctions) that connect neighboring cells. The channels (which are made from several protein subunits) permit the movement of nutrients, charged atoms (ions), and communication signals between cells. The size of the channel opening and the specific particles that move through the channel are determined by the protein subunits that make up the channel. Connexin26 is believed to play a critical role in the recycling of potassium ions at their entry into hair cells during sensory transduction from the endolymph through to the stria vascularis where other potassium channels pump potassium back into the endolymph (Scott et al., 1998).

Gap junction beta 2 protein is found in cells throughout the body, particularly in the inner ear and the skin. Because of its presence in the inner ear, especially the snail-shaped structure called the cochlea, researchers have focused on the role of this protein in hearing. Some studies indicate that channels made with gap junction beta 2 protein help to maintain the correct level of potassium ions. Other research suggests that the GJB2 gene is required for the maturation of certain cells in the cochlea. Kelsell et al. (1997) identified the first non-syndromic hearing impairment (NSHI) gene, the gap junction beta-2 gene (GJB2), which encodes for Connexin 26 (Cx26). At present, there are >100 known sequence variants for GJB2, of which 56 are reported to be associated with ARNSHI (Calvo et al., 2004)

Comparative modeling is a practical procedure in bioinformatics and computational biology because this process constructs three dimensional models that are related to known protein structure (template) (Sali and Blundell, 1993; Mart-Renom et al., 2000). Thus, this approach is relevant to structural based functional annotation. As a result, it enhances impact of structure and function on biology and medicine. By using various bioinformatics tools, three dimensional structure of GJB2 was constructed in the present study through comparative homology modeling approach. Our predicted model for GJB2 reduces the need for acquiring protein structure through experimental protocols (x-ray crystallography and nuclear magnetic resonance (NMR)). Furthermore, mutations in GJB2 were analyzed through mutant models.

MATERIALS AND METHODS

The aminoacid sequence of GJB2 was retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/nuccore/195539329?report=genbank>). It contains 226 aminoacids. It was confirmed that three dimensional structure of the protein was not available in Protein Data Bank (<http://www.rcsb.org/pdb/results/results.do?outformat=>). Hence, the current task of predicting 3D model of human GJB2 was performed via homology modeling. Then template of protein GJB2 was searched by BLASTP, scanning the non redundant protein sequence database at NCBI with efficient e-value cut off lesser than threshold, and retaining up templates with considerable e-value. Template 2ZW3 was found satisfactorily and was used further.

Web based tools SWISS-MODEL (http://swissmodel.expasy.org/workspace/index.php?func=modelling_simple1) and CPH models (<http://www.cbs.dtu.dk/services/CPHmodels/>) obtained templates automatically without any user interference. Swiss-Model is an automated knowledge-based protein modeling server. CPH models sought templates by iteratively aligning the target sequence to non redundant protein sequence database and searching the template protein data bank (PDB) in protein structure database. ESyPred3D uses PSI-Blast at NCBI. All the obtained templates are listed in Table 1. The target and template sequences were then aligned using the align2d command of MODELLER (<http://www.salilab.org/modeller/8v1/>) which uses global dynamic programming, with linear gap penalty for alignment of two profiles. ESyPred3D use neural network method for increasing the alignment performance between the query and template sequence. CPH model uses profile-profile alignment between target and template. Alignment between target and template (2ZW3) shown in Figure 1 is obtained through ClustaW web based tool.

A three dimensional structure was built from sequence alignment between GJB2 and template protein using MODELLER8v1. It constructs model by satisfaction of spatial restraints. Distance and dihedral angle restraints on target sequence were derived from alignment with template. Stereochemical restraints such as bond angles and bond lengths were extracted from CHARM22 molecular mechanics force field. Statistical correlation of dihedral angles and non-bonded interatomic distance were extracted from database of family alignments that includes proteins with known 3D structures.

CHARMM energy function and these spatial restraints were combined to obtain objective function. Final model was obtained by optimization of objective function using conjugate gradients and molecular dynamics with simulated annealing. 3Djigsaw, CPH models, ESyPred3D automatically build model by using their own set of modeling algorithms. CPH model uses segmod program from the GeneMine package. It further refines the model using encad program from the GeneMine package. The constructed models were subjected to energy minimization by steepest descent, using GROMOS96 force field, implementation of Swiss-pdb Viewer.

The accuracy of the predicted model determines information that

GJB2	-----	
2ZW3	NVCYDHYFPISHIRLWALQLIFVSTPALLVAMHVAYRRHEKKRKFVKGVRIEGLWWTY	120
GJB2	-----	
2ZW3	TSSIFFRVIFEAAFMYVFYVMYDGFMSQRLVKCNAWPCNTVDCFVSRPTEKTVFTVFM	180
GJB2	-----MDWGTLQTLGGVNHSTSIGKIWLTVLFIFRIMLVVAA	40
2ZW3	AVSGICILLNVELCYLLIRYDWGTQTLGGVNHSTSIGKIWLTVLFIFRIMLVVAA	240

GJB2	KEVWGDEQADFVCNTLQPGCKNVCYDHYFPISHIRLWALQLIFVSTPALLVAMHVAYRRH	100
2ZW3	KEVWGDEQADFVCNTLQPGCKNVCYDHYFPISHIRLWALQLIFVSTPALLVAMHVAYRRH	300

GJB2	EKKRKFVKGEIKSE-----	114
2ZW3	EKKRKFVKGVRIEGLWWTYSSIFFRVIFEAAFMYVFYVMYDGFMSQRLVKCNAWPCP	360

GJB2	-----	
2ZW3	NTVDCFVSRPTEKTVFTVFMIAVSGICILLNVELCYLLIRYDWGTQTLGGVNHST	420
GJB2	-----	
2ZW3	IGKIWLTVLFIFRIMLVVAAKEVWGDEQADFVCNTLQPGCKNVCYDHYFPISHIRLWAL	480
GJB2	-----FKDIEEK--TQVRIEGLWWTYSSIFFRVIFEAAFMYVF	154
2ZW3	QLIFVSTPALLVAMHVAYRRHEKKRKFVKGVRIEGLWWTYSSIFFRVIFEAAFMYVF	540
	:: *:: . *****	
GJB2	YVMYDGFMSQRLVKCNAWPCNTVDCFVSRPTEKTVFTVFMIAVSGICILLNVELCYLL	214
2ZW3	YVMYDGFMSQRLVKCNAWPCNTVDCFVSRPTEKTVFTVFMIAVSGICILLNVELCYLL	600

GJB2	IRYCSGKSKKPV-----	226
2ZW3	IRYDWGTQTLGGVNHSTSIGKIWLTVLFIFRIMLVVAAKEVWGDEQADFVCNTLQ	660
	*** *:::	
GJB2	-----	
2ZW3	GCKNVCYDHYFPISHIRLWALQLIFVSTPALLVAMHVAYRRHEKKRKFVKGVRIEGLW	720
GJB2	-----	
2ZW3	WYTYSSIFFRVIFEAAFMYVFYVMYDGFMSQRLVKCNAWPCNTVDCFVSRPTEKTVFTV	780
GJB2	-----	
2ZW3	FMAVSGICILLNVELCYLLIRYDWGTQTLGGVNHSTSIGKIWLTVLFIFRIMLV	840
GJB2	-----	
2ZW3	VAAKEVWGDEQADFVCNTLQPGCKNVCYDHYFPISHIRLWALQLIFVSTPALLVAMHVAY	900
GJB2	-----	
2ZW3	RRHEKKRKFVKGVRIEGLWWTYSSIFFRVIFEAAFMYVFYVMYDGFMSQRLVKCNAW	960
GJB2	-----	
2ZW3	PCPNTVDCFVSRPTEKTVFTVFMIAVSGICILLNVELCYLLIRYDWGTQTLGGVNH	1020
GJB2	-----	
2ZW3	STSIGKIWLTVLFIFRIMLVVAAKEVWGDEQADFVCNTLQPGCKNVCYDHYFPISHIRL	1080
GJB2	-----	
2ZW3	WALQLIFVSTPALLVAMHVAYRRHEKKRKFVKGVRIEGLWWTYSSIFFRVIFEAAF	1140
GJB2	-----	
2ZW3	YVFYVMYDGFMSQRLVKCNAWPCNTVDCFVSRPTEKTVFTVFMIAVSGICILLNVELC	1200
GJB2	-----	
2ZW3	YLLIRY	1206

Figure 1. Multiple sequence alignment between target (GJB2) and template (2ZW3).

can be derived from it; therefore, all the models were evaluated via applying different model assessment web servers. Stereochemical properties were evaluated through procheck (Laskowski et al., 1993). Backbone conformation was evaluated by investigating

Psi/Phi Ramachandran plot using Procheck and RAMPAGE (Laskowski et al., 1993; <http://www-cryst.bioc.cam.ac.uk/servers.html>). Packing quality and RMS of model was evaluated using Whatif packing quality control and protein analysis.

Table 2. Ramachandran plot values obtained through Procheck, Rampage and Whatif servers.

Model number	Procheck				Rampage			Whatif
	Core (%)	Allowed (%)	Generously (%)	Disallowed (%)	Number of residues in favored region (%)	Number of residues in allowed region (%)	Number of residues in outlier region (%)	Z-Score
1	87.4	9.7	2.9	0.0	92.4	6.2	1.3	-1.124
2	78.6	15.0	4.4	1.9	88.8	8.0	3.1	-1.033
3	83.9	14.7	1.1	0.3	84.7	13.4	1.9	-5.909
4	83.4	15.6	1.0	0.0	85.1	13.4	1.6	-6.237
5	89.9	8.6	1.5	0.0	92.5	5.1	2.3	-0.534
6	86.9	10.1	2.0	1.0	90.2	7.0	2.8	-6.045

RESULTS AND DISCUSSION

Three dimensional structure plays a chief role in studying disease related mutations and in drug designing process. Protein sequence of GJB2 was obtained through NCBI. Templates were obtained using blastp at NCBI. Web based tools obtained templates (<http://www-cryst.bioc.cam.ac.uk/servers.html>) automatically and are shown in Table 1. Comparative modeling builds a three dimensional structure of the target protein based on sequence identity to known protein structures (template) (<http://www-cryst.bioc.cam.ac.uk/servers.html>; Sali, 1998). Therefore, sequence identity is a good determinant for the quality of the model. Sequence of at least one related structure must have more than 30% identity. Sequence identity between target and templates is shown in Table 1. Among the different alignments, the more related alignment is of models obtained through MODELLER. The template is 2ZW3. MODELLER and web-based tools were used for building the model and global energy minimization. After model building, the structures were validated through energy minimization. Refined models were checked through RAMPAGE and

PROCHECK.

Values for the Ramachandran plot obtained through Procheck are shown in Table 2. The plot is subdivided in core, allowed, generously allowed and disallowed regions. The models obtained through MODELLER and Esypred3D showed better Ramachandran plot values, as core region (>80%) accounts for better structure (Moris et al., 1992). Rampage assessment is shown in Figure 3. Rampage derives Phi/Psi plots for Gly, Pro, Pre-Pro and other residues. The plot was divided into three regions - the favored, allowed and outlier regions. The result for models obtained through MODELLER and Esypred were significant as denser number of residues in favored region (>90%) is the measure of good quality of a model (Morris et al., 1992), but Esypred3D created the model for 216 residues while MODELLER created the model for all 226 residues.

The values for Ramachandran plot obtained through Whatif Server are shown in Table 2. The score expressing how well the backbone conformations of all residues are corresponding to the known allowed areas in the Ramachandran plot is within expected ranges for well-refined structures. These results demonstrate that prediction of the

best possible target would be a difficult task because the target performing well in one case was not found good in other cases. Esypred3D model tends to have better stereochemistry, whereas it does not hold good sequence similarity and is modeled for 216 residues only.

For all the targets described herein, the structure obtained through MODELLER, using 2ZW3 template was found to be satisfactory based on the above results. This model is shown in Figure 2. Ramachandran plot analysis through procheck showed that 87.4% residues are within the core region. RMS and packing quality was evaluated through Whatif and found satisfactory for this model. The predicted structure would be helpful for molecular characterization of proteins. Different regions of GJB2 were determined by web server SMART (http://smart.emblheidelberg.de/smart/job_status.pl?jobid=11973310764861289571183-tfEfZVzneN) which shows some important domains that include signal peptide ranging from residue no 1 to 40, Pfam connexin domain from 2 to 124 and four transmembrane regions from 21 to 40, 76 to 98, 132 to 154 and 193 to 215. The interaction network of GJB2 is shown in Figure 4. The



Figure 2. Three dimensional structure of GJB2 protein in RasMol 2.7.5.

physical and chemical properties of GJB2 were analyzed utilizing web server Protparam (<http://expasy.org/cgi-bin/protparam>). The results showed that molecular weight: 26215.0 Da, theoretical pI: 9.11, formula: $C_{1216}H_{1876}N_{302}O_{311}S_{16}$, total number of atoms: 3721, extinction coefficient: 52410 (280 nm). The estimated half-life was: 30 h (mammalian reticulo-lyocytes, *in vitro*). The instability index (II) was computed to be 42.80; this classifies the protein as unstable. Aliphatic index: 98.67. Grand average of hydropathicity (GRAVY): - 0.288. The GJB2 is composed of 20 kinds of Aminoacids. The most abundant components are Val, Ile, Lys and Phe but low content residues are His, Asn and Gln.

Congenital hearing loss occurs in approximately 1 in 1000 live births and 50% of these cases are hereditary. Most cases of hereditary hearing loss are non-syndromic sensorineural hearing loss (Morton, 1991). Recently, significant progress has been made in identifying the genes for non-syndromic hearing loss. Since the mutation of connexin26 (Cx26) gene (*GJB2*) in a deaf family was identified (Kelsell et al., 1997), half of autosomal recessive non-syndromic deafness was found to be caused by *GJB2* mutations (Hong-Joon et al., 2000). Gap junctions are believed to play a role in the recycling of potassium ions back to the endolymph of the cochlear duct after stimulation of the sensory hair cells. The loss of Cx26 would be expected to disrupt this potassium ion flow, thereby leading to hearing loss (Yeager et al., 1998; Steel, 1998). The mutation of the Cx26 gene is a major contributor to autosomal recessive deafness as well as a small percentage of autosomal dominant deafness. GJB2

variants 95G>A(R32H) and 269T>C(L90P) that occurred at conserved residues were deemed possibly damaging, while those at non-conserved residues, variants 341A>G(E114G), 380G>A(R127H), 457G>A(V153I), and 493C>T(R165W) were considered benign. The only exception was 79G>A(V27I), which occurred at a conserved residue based on the homology search but was predicted to be functionally benign. This can be explained by its location at a transmembrane region which for hydrophobic residues is variable in conservation according to a predicted hydrophobic and transmembrane matrix (Ng, 2000). In contrast, 95G>A(R32H) which also occur at the transmembrane region was considered conserved and damaging due to polarity of residues.

The 269T>C(L90P) substitution at the transmembrane region, though with a hydrophobic residue, results in a negative PHAT matrix score and was thus considered possibly damaging. The mutations 269T>C(L90P) failed to form functional gap junction channels in cellular studies (D'Andrea et al., 2002; Thonnissen et al., 2002; Bruzzone and Veronesi, 2003). On the other hand, 380G>A(R127H) and 341A>G(E114G) were not different from wild type in functional studies on transfected HeLa cells (D'Andrea et al., 2002; Thonnissen et al., 2002; Bruzzone and Veronesi, 2003). The different mutant models for GJB2 are shown in Figure 5. Furthermore, 380G>A(R127H) and 457G>A(V153I) were mostly observed in the heterozygous state among the hearing-impaired, and in addition occurred with a relatively high frequency in the hearing control population (Roux et al.,

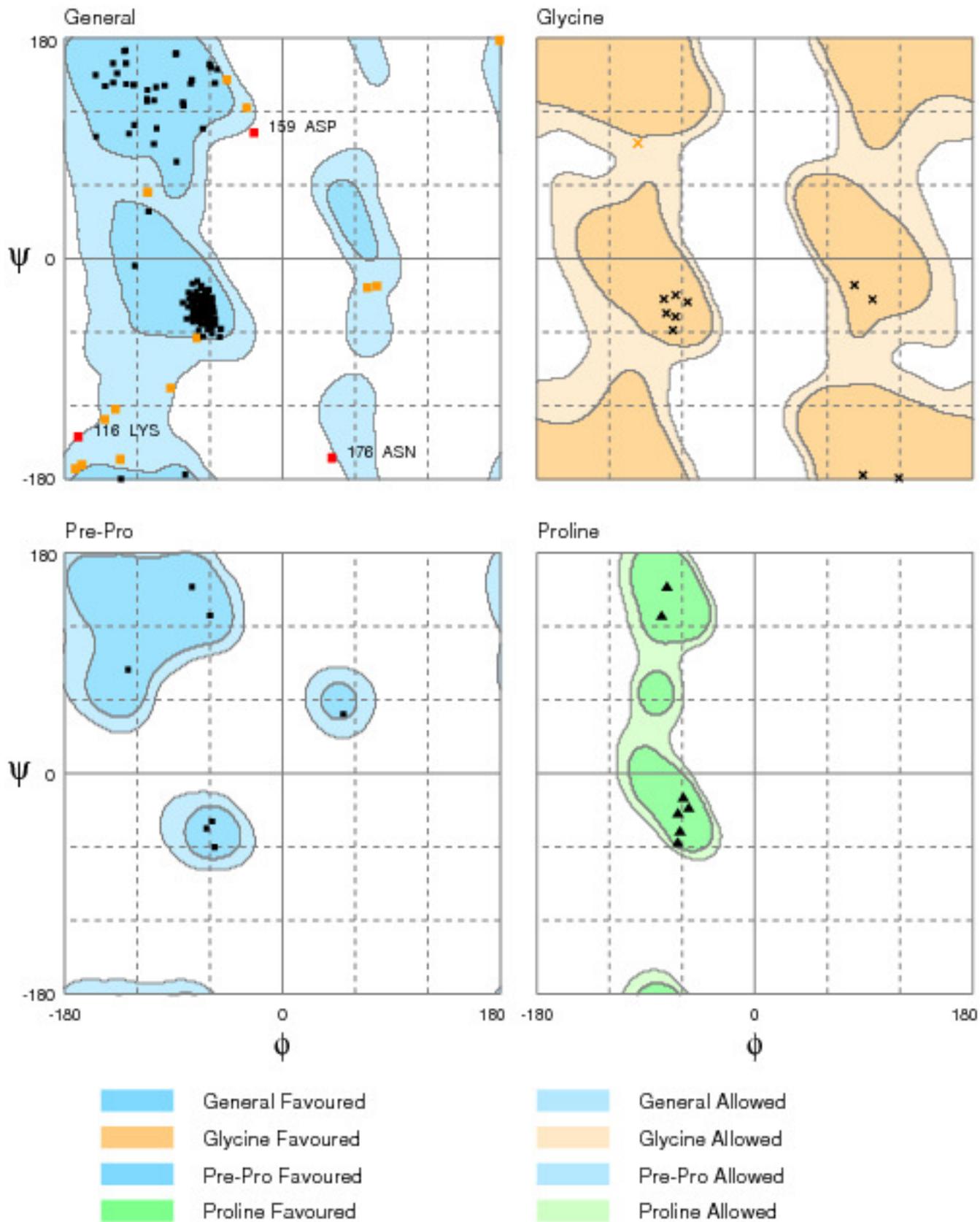


Figure 3. Ramachandran plot values showing number of residues in favored, allowed and outlier region through RAMPAGE evaluation server.

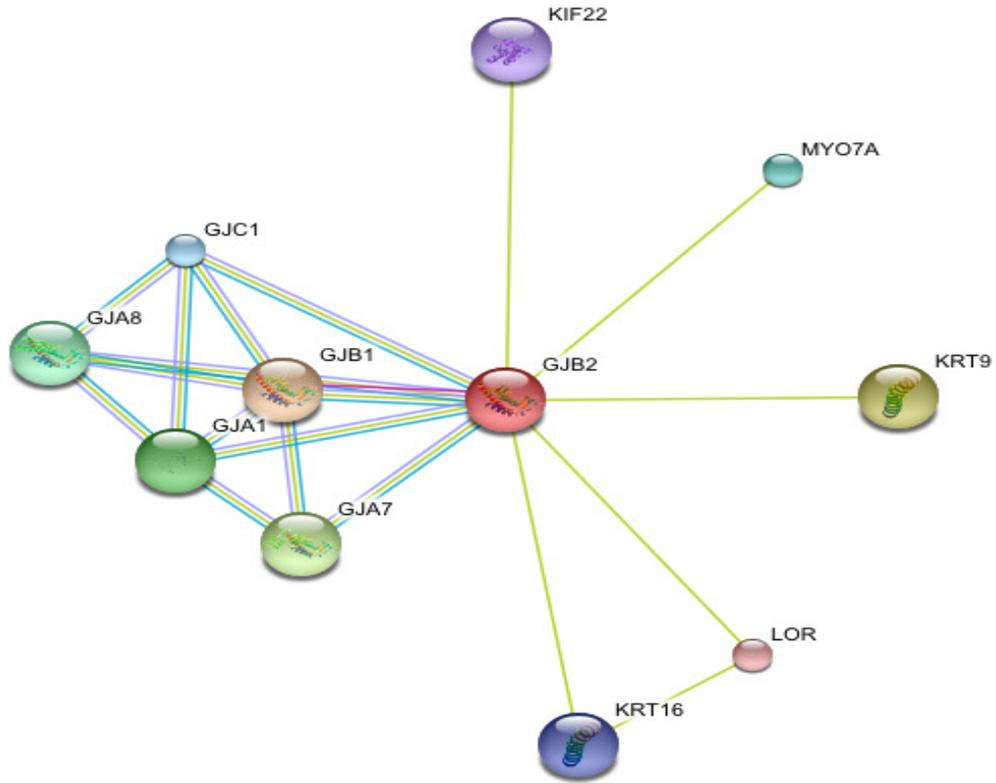


Figure 4. Interaction Network of GJB2 by SMART web server.

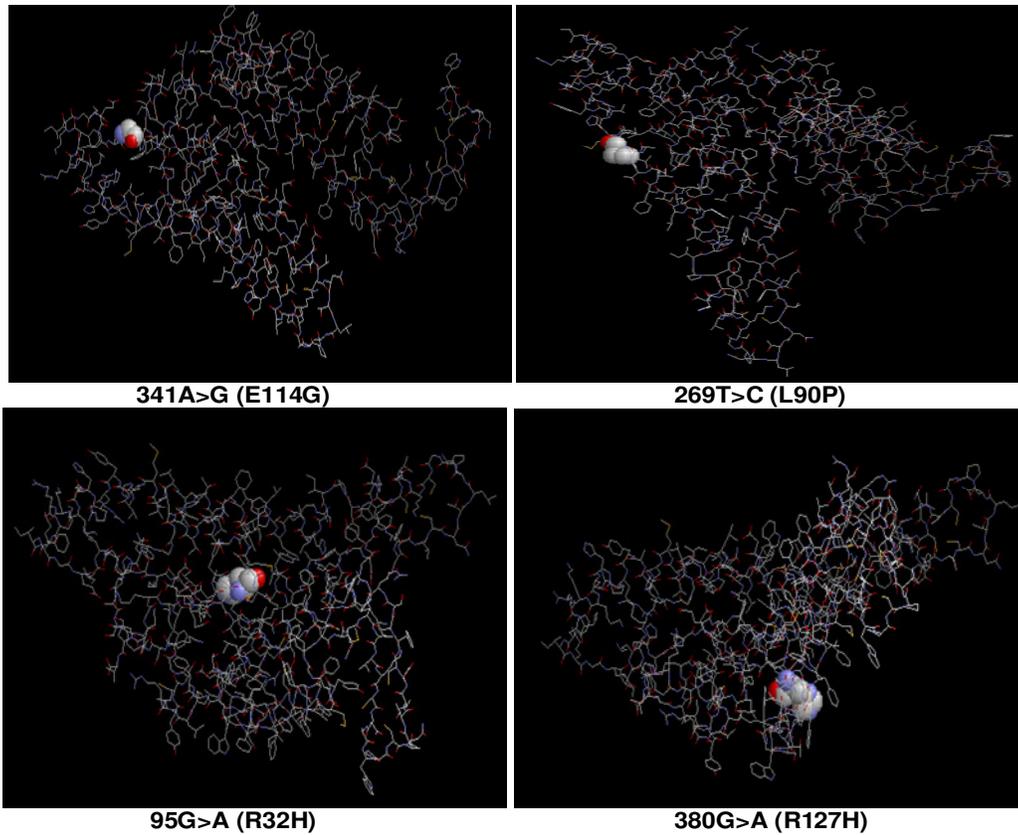


Figure 5. The reported mutations in different regions of GJB2.

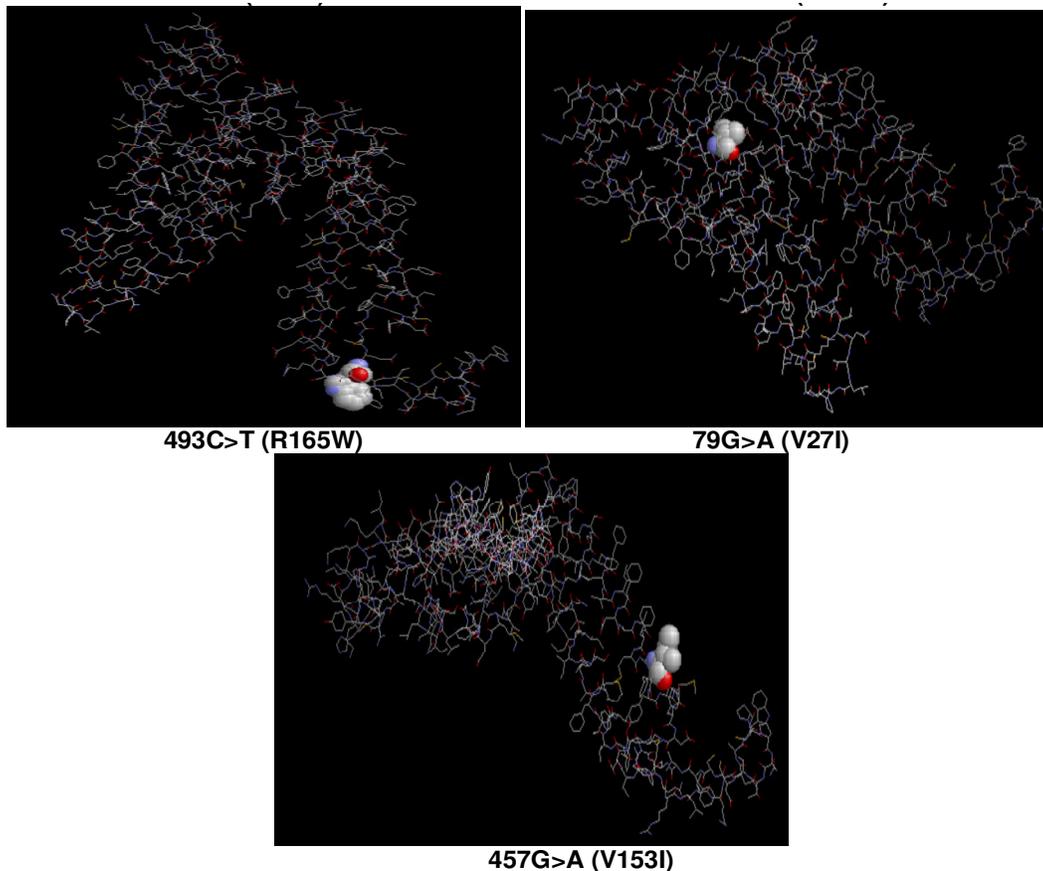


Figure 5. Contd.

2004; RamShankar et al., 2003). Polymorphisms 79G>A(V27I) and 341A>G(E114G) have been observed independently and as a haplotype. The 493C>T(R165W) variant has not been noted among hearing controls (Ram Shankar et al., 2003; Santos et al., 2005), nevertheless its predicted effect on the protein product point to its benignity (Santos et al., 2005).

REFERENCES

- Brenner SE (2000). Target selection for structural genomics. *Nat. Struct. Biol.* 7(Suppl.): p. 967.
- Bruzzone R, Veronesi V (2003). Loss-of-function and residual channel activity of connexin26 mutations associated with non-syndromic deafness. *FEBS Lett.* 533: 79-88.
- Calvo J, Rabionet R, Gasparini P, Estivill X (2004). Connexins and deafness homepage. (Retrieved on 31 August from <http://www.crg.es/deafness>).
- Choung YH, Moon SK, Park HJ (2002). Functional study of GJB2 in hereditary hearing loss. *Laryngoscope*, 112: 1667-1671.
- Cryns K, Orzan E, Murgia A, Huygen PL, Moreno F, del Castillo I, Chamberlin GP, Azaiez H, Prasad S, Cucci RA, Leonardi E, Snoeckx RL, Govaerts PJ, Van de Heyning PH, Van de Heyning CM, Smith RJ, Van Camp G (2004). A genotype-phenotype correlation for GJB2 (connexin 26) deafness. *J Med Genet* 41:147-154
- D'Andrea P, Veronesi V, Bicego M (2000). Hearing loss: frequency and functional studies of the most common connexin26 alleles. *Biochem. Biophys. Res. Commun.* 296: 685-691.
- Hong-Joon P, Houn Si, Young-Myoung (2000). Connexin26 Mutations Associated With Nonsyndromic Hearing Loss *Laryngoscope*, 110: 1535-1538.
- Kelsell DP, Dunlop J, Stevens HP (1997). Connexin 26 mutations in hereditary non syndromic sensorineural deafness. *Nature*, 387: 80-83.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. (1993). Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26: 283-291. Server:<http://www.csb.yale.edu/userguides/datamanip/procheck/><http://www-cryst.bioc.cam.ac.uk/servers.html>
- Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29: 291-325. <http://www.ncbi.nlm.nih.gov/nuccore/195539329?report=genbank> <http://www.rcsb.org/pdb/results/results.do?outformat=> http://swissmodel.expasy.org/workspace/index.php?func=modelling_simple1 <http://www.cbs.dtu.dk/services/CPHmodels/http://www.salilab.org/modeler/8v1/>
- Morris AL, MacArthur MW, Hutchinson, EG, Thornton JM (1992). Stereochemical properties of protein structure coordinates. *Proteins*, 12: 345-346 http://smart.emblheidelberg.de/smart/job_status.pl?jobid=11973310764861289571183tfEfZVzneN. <http://expasy.org/cgi-bin/protparam>
- Morton NE (1991). Genetic epidemiology of hearing impairment. *Ann. N. Y. Acad. Sci.* pp. 630:16-31.
- Ng PC, Henikoff JG, Henikoff S (2000). PHAT: a transmembranespecific substitution matrix. *Bioinformatics*, 16: 760-766.

- Sali A, Blundell TL (1993). Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* 234: 779-815.
- Sali A (1998). 100,000 protein structures for the biologist. *Nat. Struct. Biol.* 5: 1029-1032.
- Steel KP (1998). A new era in the genetics of deafness. *N. Engl. J. Med.* 339: 1545-1547.
- Thonnissen E, Rabionet R, Arbones ML, Estivill X, Willecke K, Ott T (2002). Human connexin26 (GJB2) deafness mutations affect the function of gap junction channels at different levels of protein expression. *Hum. Genet.* 111: 190-197.
- Roux AF, Pallares-Ruiz N, Vielle A (2004). Molecular epidemiology of DFNB1 deafness in France. *BMC Med.Genet.* 5: p. 5.
- RamShankar M, Girirajan S, Dagan O (2003). Contribution of connexin26 (GJB2) mutations and founder effect to non-syndromic hearing loss in India. *J. Med. Genet.* 40: e68.
- Santos RLP, Wajid M, Pham TL (2005). Low prevalence of Connexin 26 (GJB2) variants in Pakistani families with autosomal recessive non-syndromic hearing impairment. *Clin. Genet.* 67: 61-68.
- Yeager M, Unger VM, Falk MM (1998). Synthesis, assembly and structure of gap junction intercellular channels. *Curr. Opin. Struct. Biol.* pp. 517-524.