Int J Pharm Bio Sci 2014 Oct; 5(4): (B) 855 - 862

Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

HOMOLOGY BASED 3D- STRUCTURE MODELING OF AQUAPORIN-2 PROTEIN

MONICA SONI[*1] AND ARVIND PAREEK[2]

[1] Mewar University, Gangrar, Chittorgarh, Rajasthan, India
[2] Vardhaman Mahaveer Open University, Kota, Rajasthan, India

ABSTRACT

The aquaporin-2 (AQP2) water channel plays an important role in re-absorption of water in the kidney collecting duct and consequently in concentrating urine. Mutations in AQP2 can cause to disorder such as nephrogenic diabetes insipidus, congestive heart failure, liver cirrhosis, renal cystic disease, etc. Homology modeling allows building three-dimensional protein structure models using experimentally determined 3D-structures of related homologus protein as templates. In this study, it is assumed that the 3-D structure of the protein aquaporin-2 gave the structural analysis. That's why; we modeled the 3-D structure of aquaporin-2 with Swiss - Model by using the crystal structure of Aquaporin-2 as the template. Template AQP2 search with Blast and HHBlits has been performed against the SMTL and have highest identity sequence comparatively others template sequence. Its 3-D structure, obtained from this modling, was evaluated and validated by using PSVS. 3D- structure of human AQP2 protein structure is only reported in a minority of missense mutation cases in the nephrogenic diabetes insipidus.

KEYWORD: **AQP-2, Mutation, Homology Modeling, Template.**

MONICA SONI
Mewar University, Gangrar, Chittorgarh, Rajasthan, India

INTRODUCTION

Aquaporins (AQPs) are membrane proteins allocation in the transfer of water and small solutes transversely cellular membranes. Aquaporin-2 is the vasopressin-regulated channel protein, which is responsible for the regulated water reabsorbtion in the kidney collecting duct[1]. The antidiuretic hormone vasopressin was secreted by the pituitary gland regulate the expression of this aquaporin 2 (AQP2) protein. Binding arginine vasopressin (AVP) to its V2 receptor (AVPR2), leads to an increase in the intracellular cAMP levels, resulting in phosphorylation of AQP2, and possibly other proteins, by protein kinase A and subsequent redistribution of AQP2 from subapical storage vesicles to the apical plasma membrane[2]. AQP2 gene is located in chromosome region 12q13 and has four exons and three introns[3,4]. AQP-2 is a very hydrophobic membrane-integral protein of a molecular mass of 29 kDa. It is a member of the MIP protein family and is closely related to aquaporin-5; AQP2 and AQP-5 mapped to 12q13, suggesting an aquaporin family gene cluster at this locus [5]. AQP5 is localized at the apical membrane of acinar cells, and sometimes in duct cells. The homology modeling is currently the most accurate computational method to generate reliable 3-D structural models of amino acid sequence[6] and are routinely used in many biological applications . Homology modeling also known as comparative modeling of protein which allows building three-dimensional protein structure models using experimentally determined 3D-structures of related homologus protein as templates. Homology modeling can built high-quality structural models when the target and template are closely related. It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structure[7].

MATERIALS AND METHODS

Retrieval of Sequence

For the present study, the AQP2 protein sequence was retrieved from KEGG (Kyoto Encyclopedia of Genes and Genomes) server. The AQP2 Homo sapiens protein sequence having KEGG ID- HSA 359 with 271 amino acid sequence length[8] as showed in table 1.

Table 1

Primary amino acid sequence for which templates were searched and models were built.

MWELRSIAFSRAVFAEFLATLLFVFFGLGSALNWPQALPSVLQIAMAFGLGIGTLVQALGHISGAHI NPAVTVACLVGCHVSVLRAAFYVAAQLLGAVAGAALLHEITPADIRGDLAVNALSNSTTAGQAVT VELFLTLQLVLCIFASTDERRGENPGTPALSIGFSVALGHLLGIHYTGCSMNPARSLAPAVVTGKF DDHWVFWIGPLVGAILGSLLYNYVLFPPAKSLSERLAVLKGLEPDTDWEEREVRRRQSVELHSP QSLPRGTKA

Template Search and Selection

Template search with Blast and HHBlits were performed against the SWISS-MODEL template library (SMTL)[9]. The target sequence was searched with BLAST[10] against the primary amino acid sequence contained in the SMTL. A total of 55 templates was found. For each identified template, the template's quality has been predicted from features of the target-template alignment. Among of them; 4nef, Chain C with the highest sequence identity, sequence similarity and lowest 2-value had then selected for model building as template.

Protein Molecular Modeling

The strategy used in the model building was based on the target-template alignment using Promote-II. Taking 3d9s as the template and done alignment between both the sequences. Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodeled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. The generated model visualized with Deep

Viewer. The global and per-residue model quality has been assessed using the QMEAN4 scoring function[11].

Ligand Modeling

Ligands present in the template structure PS6 are transferred by homology to the model when the ligand was not clashing with the protein and the residues in contact with the ligand were conserved between the target and the template.

Statistical Assessment of the AQP2 Model

The protein structure validation software suite (PSVS) is used for assessment of protein structures generated from NMR, X-ray crystallographic and homology modeling methods. PSVS integrates analyses from several widely-used structure quality evaluation tools, including RPF, PROCHECK, MolProbity, Verify3D, Prosa II, the PDB validation software, and various structure-validation tools[12]. PSVS provides standard constraint analyses, statistics on the PDB validation goodness-of-fit between structures and experimental data, Z-Score values and knowledge-based structure quality scores in a standardized format suitable for database integration.

3D-Dimensional Model Visualization

The generated model visualized in 3D form using the RasMol molecular 3D viewer. The RasMol resulted extract information about chain, atoms, groups and bonds of 3D model[13].

RESULTS AND DISCUSSION

For the complete assessment of recognized 3-D protein structure model, present analysis of 3-D structure modeling of human AQP2 was analyzed and is described into following leads:

Templates Validation

For each identified template, the template's quality has been predicted from features of the target-template alignment as shown in table 2. The templates with the highest quality have then been selected for model building. Template 4nef.1.C validation for the generated model of AQP2 was analyzed was employing structure analysis for template chain.

Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
4nef.1.C	100	homo-tetramer	BLAST	X-ray	2.75Å	0.59	6-241	0.88	Aquaporin-2

Table 2
Multiple sequence alignment result of Aqp2 protein sequence and template 4nef chain C. 100% identity found by BLAST

```
Target    MWELRSIAFSRAVFAEFLATLLFVFFGLGSALNWPQALPSVLQIAMAFGLGIGTLVQALGHISGAHINPAVTVACLVGCH
4nef.1.C  ----ELRSIAFSRAVFAEFLATLLFVFFGLGSALNWPQALPSVLQIAMAFGLGIGTLVQALGHISGAHINPAVTVACLVGCH
Target    VSVLRAAFYVAAQLLGAVAGAALLHEITPADIRGDLAVNALSNSTTAGQAVTVELFLTLQLVLCIFASTDERRGENPGTP
4nef.1.C  VSVLRAAFYVAAQLLGAVAGAALLHEITPADIRGDLAVNALSNSTTAGQAVTVELFLTLQLVLCIFASTDERRGENPGTP
Target    ALSIGFSVALGHLLGIHYTGCSMNPARSLAPAVVTGKFDDHWVFWIGPLVGAILGSLLYNYVLFPPAKSLSERLAVLKGL
4nef.1.C  ALSIGFSVALGHLLGIHYTGCSMNPARSLAPAVVTGKFDDHWVFWIGPLVGAILGSLLYNYVLFPPAKSLSERLAVLKGL
Target    EPDTDWEEREVRRRQSVELHSPQSLPRGTKA
```

Figure 1
PDB structure of template 4nef in ribbon form. The detail information of template model showed

Model Propertied	Data
Number of Atoms	1736
Number of Bonds	1777

Model Analysis

AQP2 model structure was prepared for the target-template alignment, because the template 4nef has the highest quality and alignment with target, so it was selected for AQP2 model as shown in Figure 2. The PSVS results provided the stereo chemical property of the model. The molecular weight of model is 97804. RMS deviation of bond angle is 1.6^{o} and bond length is $0.013A^{o}$. With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution[14] <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28 (Luthy et al. 1992); a positive value indicates a 'better' score for 3Selected residues[15] 3A-725B.

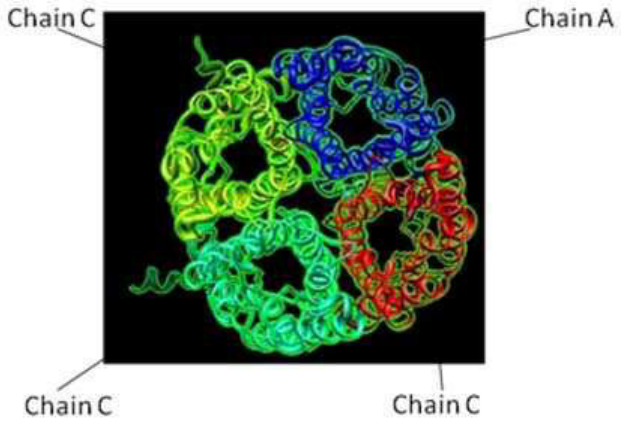

Figure 2
Showing Model Structure of AQP2 with 100% sequence Identity regarding Template sequence.

Ramchandran Plot Statistics

The Ramchandran plot displays the psi and phi backbone conformational angles for each residue in AQP2 protein as shown in figure 3. The darkest region is (shown in red) correspond to the **"**core" region represent the most favorable combination of phi-psi values. Few residues found in allowed region. The percentage of residue in the core regions are described as follows in table 3 –

Table 3
Percentage of residue in the core regions

	from PROCHECK % target	from Richardson Lab's Molprobity % target	
Most favoured region	87.3%	93.5%	
Additional allowed region	11.2%	5.3%	
Generously allowed region	1.5%	0.0%	
Disallowed region	0.0%	1.2%	
Structure Quality Factors – overall statistics	**Mean Score**	**SD**	**Z-scores**
Procheck G-factor (phi/psi only)	0.07	N/A	0.59
Procheck G-factor (all dihedral angles)	0.01	N/A	0.06
Verify 3D	0.45	0.0000	-0.16
Prosall (-ve)	0.46	0.0000	-0.79
MollProbity clashscore	11.11	0.0000	-0.38

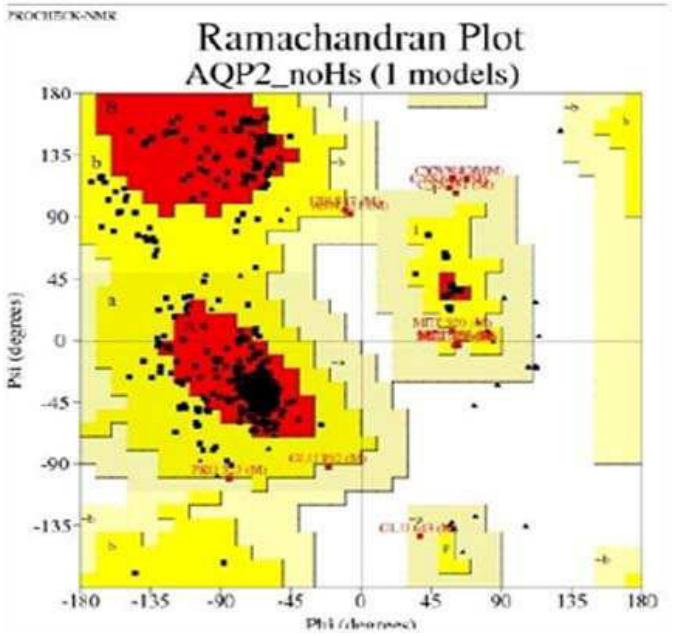

Figure3
Showing the Ramachandran plot of modelled protein AQP2.

Output From PROCHECK

For the analysis of the stereochemical quality of protein chains in AQP2, PDB structure was analysied by PROCHECK. Procheck G-factor evaluated (Fig. 4) probability of dihedral angles of a residue type to be within a given range as below-

(a) Procheck G-factor for phi-psi for ordered residues overall - 0.069

(b) Procheck G-factor for all dihedral angles for ordered residues overall 0.012

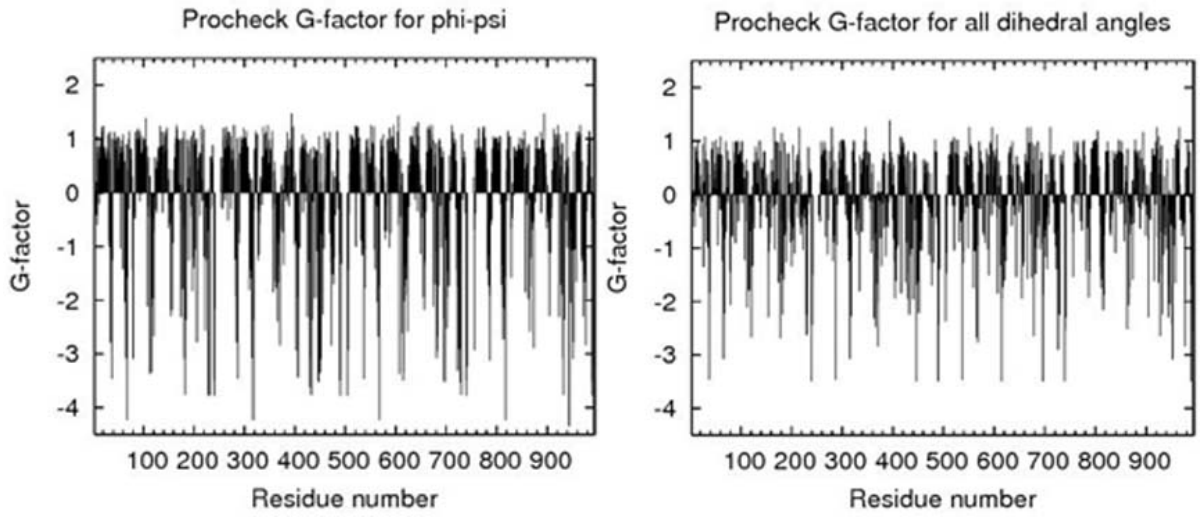

Figure 4
Procheck G-factor evaluated probability of dihedral angles of a residue.

Output From Verify 3D

Structure quality validation analysis in PSVS, was done by Verify 3D tool. Verify 3D gave the interaction probability of the amino acid sequences to have the three dimensional packing seen in the model structure and prepare overall average score 0.448 as shown in Figure 5.

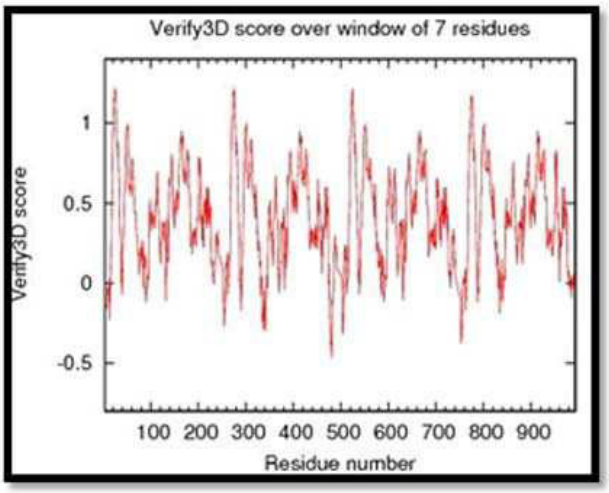

Figure5
Showing Verify 3D gave the interaction probability of the amino acid. Output From Prosall:

ProsaII relies on empirical pseudo-conformational energy potential derived from the pair-wise interactions observed in defiend AQP2 protein structure. In PSVS result calculation, ProsaII overall average score 0.448 shows in Figure 6 pseudo energy of pair-wise interactions from the spatial separation of residues.

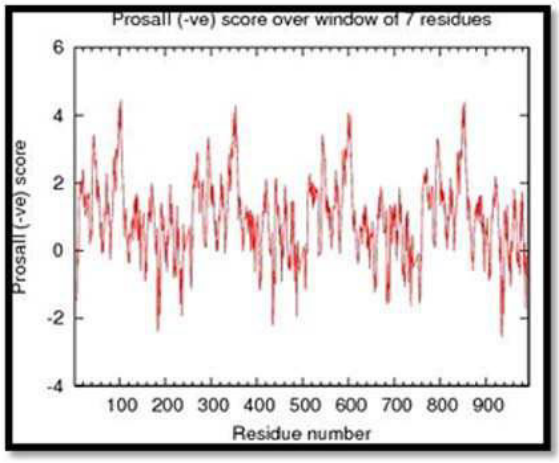

Figure6
Showing Prosall overall average score 0.414 shows pseudo energy of pair-wise interaction from the spatial separation of residues.

Output From MolProbity

In PSVS server, the MolProbity server is a valuable structure validation tool in the final stage of structure refinement. VDW violations from MAGE calculate 0.320 MolProbity (Fig. 7) clash score and visualize atomic overlap and Cbeta position deviations.

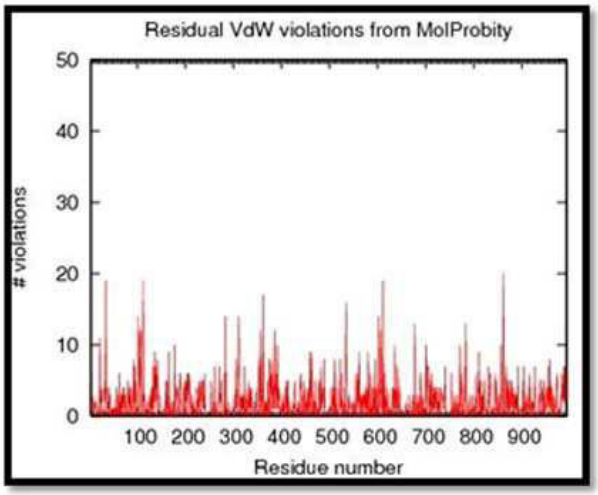

Figure 7
Showing VDW violations from MAGE calculate MolProbity clash score.

Output From PDB Validation Software

After the PDB validation software analysis 2.2 Angstroms are considered as close contacts for heavy atoms, 1.6 Angstroms for hydrogen. The RMS deviation for covalent bonds relative to the standard dictionary is 0.013 Angstroms. The RMS deviation for covalent angles relative to the standard dictionary is 1.6 degrees.

CONCLUSION

In this study, we have modeled a 3D structure of the Aquaporin-2 protein by homology modeling and visualized with help of online tools. Three dimensional model of AQP2 shows significant amino acid sequence similarity with the target sequence. Homology modeling was suggested similarity between target and template sequences. In this modeling a template is a homologous protein

that can be identified by a sequence similarity, with proteins sharing 100% identity. Protein Validation indicates about the region where residues are present. Ramachandran Plot analysis from PROCHECK indicated maximum of the residues present in most favoured region, i.e. 87.3% of the residues were in the most favoured region and from Richardson's lab Molprobity present 93.5% favoured region for selected residues.3D-structure of human AQP2 protein structure is only reported for a minority of missense mutation cases in the nephrogenic diabetes insipidus. In the reference of Homology modeling characterization, many previous studied were found, which was based on AQP1 template. But in the present study we used AQP2 template 4nef for the modeling of AQP2 protein sequences, because at the time of the template selection procedure we found that the AQP2 template provides more sequence identity in comparison to other template sequences. This study gives us prospect to illustrate consideration towards a computational approach for 3D molecular modeling and generated model were expected to be most accurate and computer generated model is only a visualization in different model; it's cannot a substitute of a crystal structure. The 3D model representation of AQP2 should prove a useful model for exploiting the increasing amount of information in the nephrogenic diabetes insipidus mutation database and residue or derivatives relationship and expermental verification would be helpful for the modeled 3-D structure of AQP2 protein.

REFRENCES

1. Schlesinger LB. "Physiognomic perception: empirical and theoretical perspectives". Genet Psychol Monogr; 101: 71-97, 1980.
2. Bougacha-Elleuch N., Ben Lassoued M. and Zouari S. et al., "Characterization of V71M mutation in the aquaporin-2 gene causing nephrogenic diabetes insipidus"; Journal of Genetics, Vol. 87, No.3, December 2008.
3. Deen PM, Verdijk MA, Knoers NV, Wieringa B, et al. "Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine". Science; 264(5155):92-5, 1994.
4. Sasaki S, Fushimi K, Saito H, et al., "Cloning, characterization, and hromosomal mapping of human aquaporin of collecting duct". J. Clin. Invest. 93:1250–56, 1994.
5. Ma T, Yang B, Umenishi F et al. "Closely spaced tandem arrangement of AQP2, AQP5, and AQP6 genes in a 27-kilobase segment at chromosome locus 12q13"; Genomics 43 (3), 387-389, 1997.
6. Bagchi P, "Ayurinformatics: Selecting the best ligand for Schizophrenic Protein D(3) Dopamine Receptor (DRD3) from the Compounds of Valeriana wallichii, Asparagus racemosus and Acorus calamus". Int. J Pharm. Bio. Sci.Vol:1; Issue;1, No 03; 2014.
7. Chothia, C; Lesk, AM. "The relation between the divergence of sequence and structure in proteins". EMBO J 5(4): 823–6. 1986,PMC 1166865. PMID 3709526.
8. http://www.genome.jp/dbget-bin/www_bget?hsa:359 accessed on 22.01.2014.
9. http://swissmodel.expasy.org/ accessed on 25.03.2014.
10. Altschul S.F., Madden T.L., Schaffer A.A., et al. , "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". Nucleic Acids Res, 25, 3389-3402, 1997.
11. Benkert P., Biasini M. and Schwede T. "Toward the estimation of the absolute quality of individual protein structure models". Bioinformatics, 27, 343-350, 2011.
12. http://psvs-1_5-dev.nesg.org/results/AVPR2/OUTPUT.html accessed on 28.06.14
13. http://rasmol.org/ accessed on 25.03.2014
14. Luthy R, Bowie J U and Eisenberg D, "Assessment of protein models with three-dimensional profiles", Nature: 83-85, 356, 1992.
15. Sippl M J, "Recognition of Errors in Three-Dimensional Structures of Proteins", Proteins 355-362,17, 1993.