Int. J. Pharm. Investigation, 2019;9(2):43-46 *Original Article*

Comparative Modeling, Characterization and Energy Minimization Studies of Aquaporin 9: An Exclusive Target Protein for Rheumatoid Arthritis

Surojeet Das[1], Parul Johri[2], Ruchi Sharma[1], Monika Kashyap[1], Shivani Singh[1], Sachidanand Singh[1,*]
[1]Faculty of Biotechnology, Institute of Bio-Sciences and Technology, Shri Ramswaroop Memorial University, Barabanki- 225003, Uttar Pradesh, INDIA.
[2]Amity Institute of Biotechnology, Amity University Uttar-Pradesh, Lucknow Campus, Malhour Gomti Nagar Extension, Lucknow-227105, Uttar Pradesh, INDIA.

ABSTRACT

Background: In the absence of the experimentally determined structure, computer aided protein structure prediction, evaluation and their energetically stable structure identification is the only way out of the problem. The main objective of the study was to perform the structure prediction of Aquaporin 9 (APQ9) the most targeted protein for rheumatoid arthritis, using *in-silico* methodology and validate the generated models. **Methods:** Secondary structure prediction of AQP9 was performed using GOR4, SOPMA and CFSSP algorithm. This was followed by the three-dimensional structure identification from MODELLER, LOMETS and MUSTER server. Many models were built and the best amongst them was identified on the basis of their DOPE score. RAMPAGE was used to validate these models and finally the selected model was energetically stabilized. **Results:** Amongst the 4 predicted models, model predicted using MODELLER software with 1FX8 PDB template (MODELER MODEL 2) was selected as the best. This model showed the best results in Ramachandran plot validation. In the Ramachandran Plot, 223 residues (95.7%) were found to be in the favored region, 9 residues (3.9%) in the Allowed region and the rest 1 residue (0.4%) in the Outlier region. Energy minimization calculations were also done for the four models using SPBDV software and Modeler Model 2 model showed the least energy (E= 3484.038 KJ/mol). **Conclusion:** The accurate three-dimensional structure prediction of proteins is a grand challenge now. Massive amount of sequence and structural data is available now with low cost. The choice of one or other method depends solely on the type of protein sequence and the quality of the predicted structure. The accurate structure prediction, fold recognition, energy calculation, side chain modeling and target template identification are the crucial edges of the molecular modeling process which need to be scrutinized for the best predicted model.

Key words: Homology Modeling, MODELLER, LOMETS, MUSTER, RAMPAGE, Energy minimization, Secondary structure prediction.

Correspondence
Dr. Sachidanand Singh (M Tech, PhD),
Associate Professor and Dean, Faculty of Biotechnology, Institute of Bio-Sciences and Technology, Shri Ramswaroop Memorial University, Barabanki-225003, Uttar Pradesh, INDIA.
Phone no: +91-9205782877
Email: dean.fobt@srmu.ac.in
DOI: 10.5330/ijpi.2019.2.9

INTRODUCTION

Rheumatoid Arthritis (RA) is a chronic systemic inflammatory disorder that affects several tissues and organs, but affects synovial joints in particular.[1] In the process it produces an inflammatory response of synovium (synovitis), which is secondary to synovial cell hyperplasia (excess synovial fluid) and aids in the progression of synovial pannus.[2] The world prevalence of rheumatoid arthritis is about 1%. The disease process's pathology often leads to the destruction of joint cartilage and ankylosis. Rheumatoid arthritis can also cause diffuse inflammation of the lungs, pericardium, pleura and sclera and nodular lesions, which are most common in subcutaneous tissue.[3]

Despite major progress in our understanding of RA, it still remains a disease of unidentified etiology. In recent times, Aquaporin (AQP) proteins have been addressed for their possible involvement in RA.[4] The Aquaporins (AQP) are a family of small hydrophobic proteins (about 25-34 kDa) and integral membrane channels that facilitate homeostasis of water in cells of all organisms. Aquaporin-9 (AQP9) not only facilitates the transport of water, but also of neutral solutes, including glycerol, urea and other small non-electrolytes. It is well-known that the principal pathological phenomena associated with RA are characterized by increased levels of inflammatory cytokines secreted by activated B and T cells. This in turn causes damage to bones and cartilages. Correspondingly, different AQPs have been detected in cartilage cells where they control the traffic of ions and molecules and hence, regulate cartilage physiology.[5] Recent evidences have suggested that in patients of RA, TNF-α could regulate either Aquaporin 9 (AQP9) mRNA and protein expression.[6]

Progress in evaluating the role of AQP9 in RA has been primarily hindered due to the absence of its experimentally determined structure. X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are majorly the two experimental methods for evaluating protein structure. However, these techiques face drabacks like high consumption of time and manpower. These techniques also possess critical limitations for different protein targets. However, with the advent of moden day sequencing techniques, it is rather simple to acquire protein sequences than acquiring protein structure. This in turn, has led to databases such as UniProt(https://www.uniprot.org/) and TrEMBL (Translated EMBL) (https://www.uniprot.org/statistics/TrEMBL) acquiring more than 85 million of protein sequences.[7] In the late 20th century computational methods for predicting protein structure from amino acid sequence came into focus. Research in the area depicted that the information in a protein which is required for its appropriate folding is encoded in its amino acid sequence (Anfinsen's dogma).[8] Currently, the major computational methods to determine protein structure include homology modeling (based on sequence comparison), threading (based

on fold-recognition) and *ab initio* (does not rely on any previously solved structure).[9]

The paper aims at devoloping efficient computational protein structure models of AQP9. This in turn, would aid further research and analysis of the role of AQP9 in RA. In the absense of its experimentally deduced sturucture structure, comparitive modelling was performed using MODELLER (https://salilab.org/modeller/), LOMETS (Local Meta-Threading Server) (https://zhanglab.ccmb.med.umich.edu/LOMETS/) and MUSTER (MUlti-Sources ThreadER) (https://zhanglab.ccmb.med.umich.edu/MUSTER/) software. The modelled structure was then put through RAMPAGE (Ramachandran Plot Assessment) (http://mordred.bioc.cam.ac.uk/~rapper) for its evaluation. Energy minization of the four modelled structures was performed through SPDBV (Swiss PDB Viewer) (https://spdbv.vital-it.ch/) software. Protein secondary structuctures were also generated using GOR4 (Garnier-Osguthorpe-Robson) (https://npsa-prabi.ibcp.fr/NPSA/npsa_gor4.html), CFSSP (Chou and Fasman Secondary Structure Prediction Server) (http://www.biogem.org/tool/chou-fasman/) and SOPMA (Self-Optimized Prediction Method with Alignment) (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) algorithms.

MATERIALS AND METHODS

The comparative modelling for protein structure prediction consisted of four major steps. Firstly, target identification followed by alignment of target and template sequence. Once the template sequence alignmnet was completed, the model was generated. The model was finally assessed for its energy, stearic clashes and stability (Figure 1).

Protein secondary structure prediction

The Protein sequence of AQP9 (Accesion number: O43315 (AQP9_HUMAN)) was retrived from Uniprot database and was subjected to secondary structure prediction using GOR4, SOPMA and CFSSP at expasy server.

Template identification for protein tertiary structure prediction

An extensive search was performed against PDB (Protein Data Bank) (http://www.rcsb.org/) to identify potentially related sequences for which experimentally determined structures are already known. Two

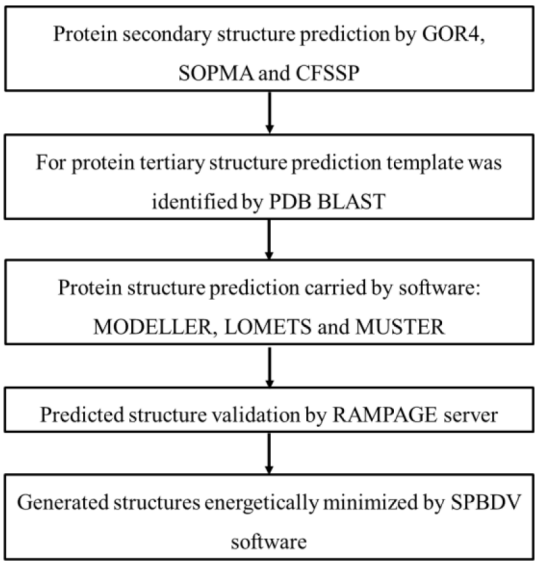

Figure 1: Workflow of molecular modelling.

template protein structures (1LDF_A and 1FX8_A) were identified as the best fit templates for the computational three dimentional protein structure prediction, on the basis of various parameters like the e value, percentage identity, score of the allignment and query coverage.

Protein structure modelling

The three-dimensional structure of the protein of interest (AQP9) was determined using MODELLER version 9.15, LOMETS and MUSTER server. MODELLER performs comparative protein modelling based on the template identified. LOMETS is meta-threading technique for predicting template-based protein structure. MUSTER is a protein threading algorithm to identify template structures from PDB library. It generates sequence-template alignments by merging sequence profile-profile alignment with multiple structural data. Numerous models were generated using the two template structures and the best model amid them was selected by comparing their DOPE score.

Structure validation

RAMPAGE server was applied to generate Ramachandran plots for validation of the predicted protein structures by assessment of factors such as favoured, allowed and outlier regions of amino acid residues in the predicted protein structure. The pdb files of the best models of the target genes predicted by MODELLER, LOMETS and MUSTER were submitted to RAMPAGE server for generating Ramachandran Plot. Ramachandran plots were generated for the predicted models and the plots were compared for finding the best model among the predicted models.

Energy minimization of the predicted structures

Predicted and evaluated structures were subjected to energy minimization to attain the lowest energy conformation by SPBDV.

RESULTS

Protein secondary structure prediction

Secondary structure prediction of AQP9 was executed by the aid of tools like CFSSP, GOR4 and SOPMA. Information about the secondary structures such as alpha helix, beta strand and random coil for target AQP9 from GOR4, CFSSP and SOPMA tools from expasy was extracted.

CFSSP is an empirical technique for predicting secondary structures in proteins. The method depends on analyses of relative frequencies of each amino acid in alpha helices, beta sheets and turns based on known protein structures solved with X-ray crystallography. CFSSP analysis revealed that AQP9 consisted of 219 alpha helix, 204 extended strands and 32 turns.

The GOR method is an information theory-based technique for predicting secondary structures in proteins. It is based on probability parameters derived from empirical studies of known protein tertiary constructs solved by X-ray crystallography. GOR4 analysis depicted that AQP9 consisted of 103 alpha helix, 60 extended strands and 132 random coils.

SOPMA is a protein secondary structure prediction tool in expasy server. The software lead to majordevelopments in protein secondary structure by utilizing consensus prediction from multiple alignments. SOPMA analysis revealed that AQP9 consisted of 105 alpha helix, 78 extended strands and 80 random coils.

Template identification

PDB Blast was performed for identifying template structures for comparative homology modelling of AQP9. The templates were compared and two of them were selected (1LDF, 1FX8) on the basis of their Query cover, E value and Identity (Table 1). These two templates were down-

Table 1: BLAST parameters for target gene APQ9.

Query Cover	E Value	Identity	Accession
91%	3e-55	40%	1LDF_A
91%	3e-54	40%	1FX8_A

loaded from PDB for modelling of the protein using MODELLER software.

Protein structure modelling using MODELLER

Using the template files ILDF and 1FX8, structures were modelled for the given protein AQP9 using MODELLER version 9.15 software. Fifty models were generated using modeller for ILDF and 1FX8. Keeping DOPE score as criteria one best model was selected for ILDF (Model 1) (Figure 2 A) and 1FX8 (Model 2) (Figure 2 B).

Protein structure prediction using LOMETS server

LOMETS server was also applied to predict the three dimensional structure of the query sequence using a meta threading approach. A total of 10 models were generated and were further analyzed for the best models. The best model amid them was selected by comparing their Zscore and maximum coverage. The best model of the structure was predicted by LOMETS server (Figure 2 C).

Protein structure prediction using MUSTER Server

Further, protein treading was performed by online server of MUSTER. This server generated 10 different models for the protein sequence amongst which the model having the least Z score and maximum coverage over the query was selected as the best model (Figure 2 D).

Structure validation using Ramachandran plot

To analyse the predicted structures, all the four predicted models were uploaded to the online database RAMPAGE, which generated the Ramachandran plots for the predicted protein structures (Figure 3). The amino acids (residues) were distributed in three distinct regions in this plot. The three distinct regions were favoured region, allowed region and outlier region (Table 2).

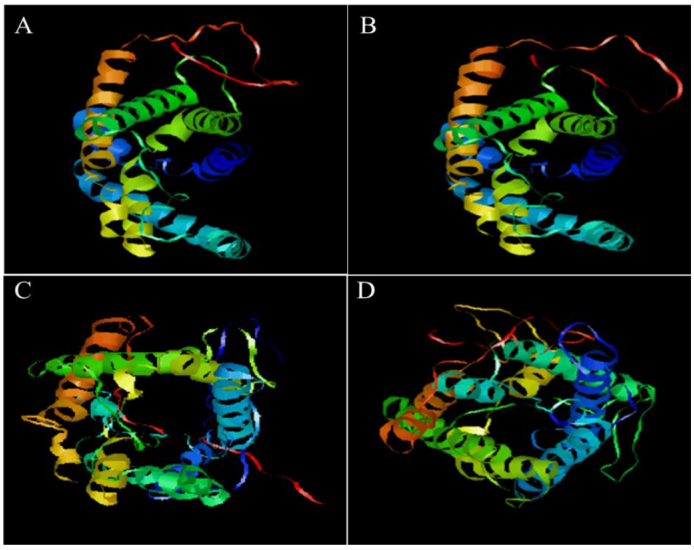

Figure 2: Best predicted model by (A) MODELLER (Model 1); (B) MODELLER (Model 2); (C) LOMETS server; (D) MUSTER server.

Table 2: Comparison of Ramachandran plots of different models.

	MODELLER		LOMETS Model	MUSTER Model
	Model 1	Model 2		
Favored Region	222 (95.3%)	223 (95.7%)	278 (84.6%)	273 (93.2%)
Allowed Region	8 (3.4%)	9 (3.9%)	27 (9.2%)	13 (4.4%)
Outlier Region	3 (1.3%)	1 (0.4%)	18 (6.1%)	7 (2.4%)

Table 3: Energy minimisation values for different models.

	MODELLER		LOMETS Model	MUSTER Model
	Model 1	Model 2		
Energy (KJ/mol)	4416.158	3484.038	14757.951	7031.386

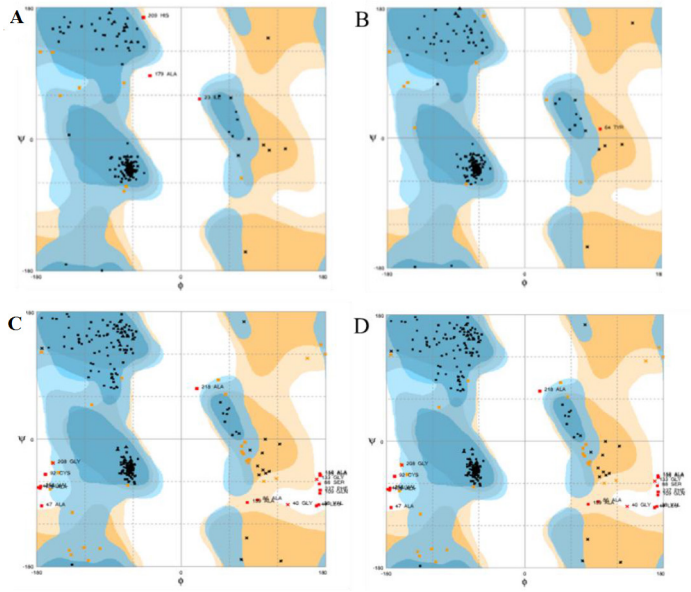

Figure 3: Ramachandran plots for (A) MODELLER (Model 1); (B) MODELLER (Model 2); (C) LOMETS server; (D) MUSTER server.

Energy minimisation

The energy minimisation for the predicted models carried by SPBDV software. Energy minimisation values were compared to finalise the best predicted models (Table 3). MODELLER Model 2 presented the least energy (E= 3484.038 KJ/mol).

DISCUSSION

There are various computational methods to predict the protein structures varing from homology approaches to protein threading and *ab initio* methods.[10] In the present work, the structure of AQP9 is predicted using all these approaches (MODELLER, LOMETS and MUSTER server). The models generated were validated using Ramachandran plot to finalize the best model. Finally, energy minimization was performed for the best model identified. The lowest energy model identified was from modeler software with energy of 3484.038 KJ/mol. This computationally predicted protein model for AQP9, the significant target protein for rheumatoid arthritis could be further used for active site prediction and docking analysis.[11]

CONCLUSION

In recent times, there has been a tremendous increase in the *in-silico* structure prediction of protein using different software and algorithms.[12]

But the accuracy of the structure prediction, the magnitude of errors in the fold assignment, modelling of side chains and loops still need a considerable body shift. Aquaporin-9s are a family of proteins encoded by AQP9 gene in humans. They are usually present at the plasma membrane, where they control the influx and outflow of water and small molecules. These proteins were detected in synovial tissues of people affected with Rheumatoid Arthritis (RA). In the absence of their experimentally determined structure, their functional studies were a subject of concern. Amongst the 4 predicted models, model predicted using MODELLER software with 1FX8 PDB template (MODELER Model 2) was selected best. This model showed the best results in Ramachandran plot validation. In the Ramachandran Plot, 223 residues (95.7%) were found to be in the favoured region, 9 residues (3.9%) in the allowed region and the rest 1 residue (0.4%) in the Outlier region. Energy minimization calculations were also carried for the four models using SPBDV software and Modeller Model 2 model showed the least energy (E= 3484.038 kJ/mol). Hence, this structure can be used for structure and functional analysis of AQP9 protein.

ACKNOWLEDGEMENT

This work was supported by Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

APQ9: Aquaporin 9; **RA:** Rheumatoid Arthritis; **AQP:** Aquaporin; **NMR:** Nuclear Magnetic Resonance; **TrEMBL:** Translated EMBL; **LOMETS:** Local Meta-Threading Server; **MUSTER:** MUlti-Sources Threader; **RAMPAGE:** Ramachandran Plot Assessment; **SPDBV:** Swiss PDB Viewer; **GOR4:** Garnier-Osguthorpe-Robson; **CFSSP:** Chou and Fasman Secondary Structure Prediction Server; **SOPMA:** Self-Optimized Prediction Method with Alignment; **PDB:** Protein Data Bank.

REFERENCES

1. Almaaytah A, Elhajji FD. Comparative Cost Efficiency of the Originator Drug of Infliximab and its Biosimilar for the Treatment of Rheumatoid Arthritis in the MENA Region. Int J Pharm Investig. 2019;9(1):12-5.
2. Turesson C, O'Fallon WM, Crowson CS, Gabriel SE, Matteson EL. Extra-articular disease manifestations in rheumatoid arthritis: Incidence trends and risk factors over 46 years. Ann Rheum Dis. 2003;62(8):722-7.
3. O'Neil LJ, Kaplan MJ. Neutrophils in Rheumatoid Arthritis: Breaking Immune Tolerance and Fueling Disease. Trends Mol Med. 2019;25(3):215-27.
4. Lindskog C, Asplund A, Catrina A, Nielsen S, Rützler M. A Systematic Characterization of Aquaporin-9 Expression in Human Normal and Pathological Tissues. J Histochem Cytochem. 2016;64(5):287-300.
5. Mobasheri A, Trujillo E, Bell S, Carter SD, Clegg PD, Martín-Vasallo P, *et al.* Aquaporin water channels AQP1 and AQP3, are expressed in equine articular chondrocytes. Vet J. 2004;168(2):143-50.
6. Nagahara M, Waguri-Nagaya Y, Yamagami T, Aoyama M, Tada T, Inoue K, *et al.* TNF-alpha-induced aquaporin 9 in synoviocytes from patients with OA and RA. Rheumatology. 2010;49(5):898-906.
7. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al.* The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.
8. Anfinsen CB. Principles that govern the folding of protein chains. Science. 1973;181(4096):223-30.
9. Deng H, Jia Y, Zhang Y. Protein structure prediction. Int J Mod Phys B. 2018;32(18):1840009.
10. Dawson WM, Rhys GG, Woolfson DN. Towards functional de novo designed proteins. Curr Opin Chem Biol. 2019;52:102-11.
11. Desai M, Chauhan JB. Predicting the functional and structural consequences of nsSNPs in human methionine synthase gene using computational tools. Syst Biol Reprod Med. 2019;65(4):288-300.
12. Singh S, Vennila JJ, Snijesh VP, George G, Sunny C. Implying Analytic Measures for Unravelling Rheumatoid Arthritis Significant Proteins Through Drug-Target Interaction. Interdiscip Sci. 2016;8(2):122-31.

Cite this article: Das S, Johri P, Sharma R, Kashyap M, Shivani S, Sachidanand S. Comparative Modeling, Characterization and Energy Minimization Studies of Aquaporin 9: An Exclusive Target Protein for Rheumatoid Arthritis. Int. J. Pharm. Investigation. 2019;9(2):43-6.