

## Original Research Article

<https://doi.org/10.20546/ijcmas.2018.706.257>

## Homology-Based Structure Prediction of the Human Kruppel-Like Factor 1 Protein

Sayed Abdul Azeez\*

Department of Genetic Research, Institution for Research and Medical Consultations (IRMC),  
Imam Abdulrahman Bin Faisal University (Formerly: University of Dammam), Dammam,  
Saudi Arabia

\*Corresponding author

### ABSTRACT

#### Keywords

Protein homology modelling; *KLF1* gene; Ramchandran plot; Insilco methodologies; I-TASSER; 3D structure prediction

#### Article Info

##### Accepted:

14 May 2018

##### Available Online:

10 June 2018

Erythroid Kruppel-like transcription factor (KLF1 or EKLF) plays a vital role in haematopoiesis. The KLF1 protein is composed of 362 amino acids that are coded within three exons of the *KLF1* gene, covering three zinc fingers and two functional domains. The Protein Data Bank contains only one structure of KLF1 (id 2MBH) from *Homo sapiens*, with a length of 41 amino acids, thus lacking a complete 3D model. The present study has been undertaken to predict the 3D structure of KLF1 using automated protein homology modelling servers. These structural attributes were predicted using a hierarchical approach by Iterative Threading ASSEMBLY Refinement, which can be used to predict similarities and comparative awareness of ligand interactions with KLF1. Based on the output of I-TASSER server simulations, a model was generated with a C-score of 3.79. This high value signifies the best predicted 3D structural model of KLF1, with the lowest RMSD value of 1.26 Å and an energy minimization of -532 kcal/mol, which was validated with KoBaMIN. Thus, our structurally and energetically optimized KLF1 3D structure can be used as a model for further theoretical studies in structure-based drug design of novel compounds for the treatment of SCD and beta thalassemia disorders.

### Introduction

Kruppel-like factor 1 (KLF1) or erythroid Kruppel-like transcription factor (EKLF) is a factor responsible for haematopoietic-specific transcription that stimulates adult beta-globin and plays an important role in globin gene switching and erythrocyte maturation (Viprakasit *et al.*, 2014; Tallack *et al.*, 2010; Vassen *et al.*, 2014; Borgio *et al.*, 2016). The *KLF1* gene encodes the Kruppel-like factor 1 protein (protein symbol: Q13351-

KLF1\_HUMAN), which is composed of 362 amino acids that are coded by 3 exons within the gene, which is located on chromosome locus 19p13.2. This protein contains 3 zinc fingers and two functional domains. The N-terminal domain is proline-rich, and the transactivation domain and C-terminal domain contain essential factors for DNA binding. *KLF1* plays a vital role as an erythroid transcription factor, and it binds to the DNA binding sequence, a CACCC motif, in the beta-globin gene. The major biological

function of KLF1 is regulation of erythropoiesis, as it regulates the switch between foetal haemoglobin and adult haemoglobin expression both by direct activation of the  $\beta$ -globin gene and indirect repression of the  $\gamma$ -globin gene expression in human erythroid progenitors (Perkins *et al.*, 2016). KLF1 regulates the cell cycle, haeme synthesis enzymes, and globin chaperones and activates genes that encode metabolic and antioxidant enzymes, structural membrane and ion channels, and water channels (Perkins *et al.*, 2016; Borgio *et al.*, 2016; Allack *et al.*, 2010; Pilon *et al.*, 2011; AbdulAzeez *et al.*, 2016). The main objective of this study is to accurately determine the 3D structure of KLF1, which will help to identify binding sites for the development new drugs. A well-established structure will guide us towards structure-based drug design using novel compounds for the treatment of sickle cell disease and beta thalassemia.

## Materials and Methods

### Template Search and Sequence Alignment

The amino acid sequence of KLF1 was obtained from NCBI (reference sequence: NP\_006554.1) and the UniProt database (accession number: Q13351 KLF1\_HUMAN). To build a 3D model of KLF1, template identification was performed using a basic local alignment search tool (BLAST) sequence search. We selected PSI-BLAST, and this was searched against the PDB database to screen all human NMR and crystal structures during October 2016, excluding models and culture sequence samples, to find suitable templates for homology modelling. Crystal structures of the human proteins were retrieved with PDB IDs 2WBS\_A, 4M9E\_A, 2WBU\_A, and 2L2I\_B. FASTA sequences of all structures were downloaded from PDB, and template alignment was performed using a multiple sequence alignment program (MAFFT)

(Kato *et al.*, 2002), which detects structurally conserved regions (SCR) common to the KLF1 template and target sequence.

### 3D Structure prediction

Homology modelling of KLF1 was performed using an automated form of the I-TASSER server and the MODELLER program. The KLF1 FASTA sequence was retrieved from the UniProt database (accession number: Q13351) and was submitted to I-TASSER and Modeller. Modeller 9v6 was used for 3D structure prediction (<http://www.salilab.org/modeler>) (Šali *et al.*, 1993). I-TASSER is a hierarchical methodology to predict 3D protein structure and function using LOMETS, a multiple threading approach program.

I-TASSER generates full-length atomic structural models from multiple threading alignments and template fragment assembly simulations. The KLF1 FASTA sequence was submitted to I-TASSER, and 10 multiple threading alignments were generated (Figure 1) based on the Z-score selected. Five 3-dimensional models were generated with the confidence score (C-score), estimated TM-score and RMSD reported. The C-score of Models 1, 2, 3, 4, and 5 are -3.79, -3.82, -3.67, -3.73, and -3.50, respectively. Based on these scores, the best model with a C-score of -3.79, an estimated TM-score of  $0.30 \pm 0.10$ , and an estimated RMSD of  $16.1 \pm 3.1$  Å was selected (Yang *et al.*, 2015).

### Protein structure refinement and energy minimization

Structural refinement of the KLF1 model was performed using the KoBaMIN web server (Rodrigues *et al.*, 2012), which offers an online interface enhancement protocol to predict a consistent and computationally proficient protein structure. To optimally refine the protein structure of KLF1 based on

energy minimization, the KLF1 model was selected and submitted to KoBaMIN (Vyas *et al.*, 2015; Rodrigues *et al.*, 2012).

### Model Validation

Model validation is a key step in homology modelling, as the accuracy of the stereochemical quality of the protein structure is responsible for its suitability for diverse biological and molecular experimental purposes. The quality estimation of the model by I-TASSER was calculated as a confidence score (Roy *et al.*, 2010). The superiority of the model was validated with a Ramchandran plot (RAMPAGE server) (Ramachandran *et al.*, 1990) and VERIFY3D (Eisenberg *et al.*, 1997). The Ramchandran plot explains the visualization of dihedral angles  $\phi$  against  $\psi$  of amino acid residues within the protein structure and visualizes possible conformations of  $\phi$  and  $\psi$  angles for a polypeptide chain. The number of amino acid residues in the allowed and disallowed regions of the Ramchandran plot determines the quality of the model (Lovell *et al.*, 2003).

### Protein-ligand binding site prediction

COACH is a meta-server that predicts complementary ligand binding sites using TM-SITE and S-SITE as comparative methods based on templates from the BioLiP protein database. The COACH algorithm produces multiple function annotation results with the created 3D models, and the obtained output is based on the C-score ranges [0-1]. The C-score is a confidence score of the most reliable prediction (Yang *et al.*, 2015).

## Results and Discussion

### Sequence analysis

Only one KLF1 structure from *Homo sapiens* is present in the Protein Data Bank (PDB) (ID:

2MBH), but the 3D structure of this protein is not fully available as it is only 41 amino acids in length. Human KLF1 is composed of 362 amino acids, and no full-length 3D structure of KLF1\_Human currently exists. Therefore, in this study, I predicted the 3D structure of KLF1. FASTA sequences from UniProtKB - Q13351 were retrieved after an NCBI database search for the *KLF1* gene using PSI-BLAST against PDB. Based on the output of I-TASSER, the 10 best templates were selected by LOMETS threading programs based on the highest Z-score (2.29).

From these 10 best templates, 2IL3 was selected. The I-TASSER simulations generated a KLF1 model with a C-score of 3.79 (C-score range of -5 to 2), and this higher value signifies that the created model of KLF1 is predicted accurately.

### KLF1 structure construction and validation

A total of 5 KLF1 3D structural models were determined using the I-TASSER server, and another 5 models were generated from MODELLER 9.1. The final model was selected after analysing and validating the models.

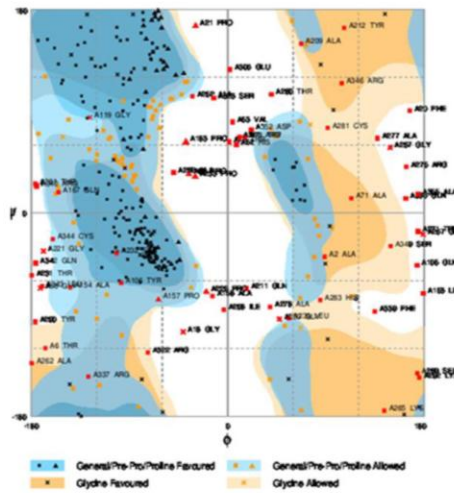
The RMSD of the models was calculated with the template structure using SPDBV software and SuperPose v1.0. Additional refinement of KLF1-modelled structures was performed using the KoBaMIN web server, and the lowest RMSD of 1.26 Å was obtained, an energy value of  $E=7219$ , and an energy minimization of -532 kcal/mol (Rodrigues *et al.*, 2012). The selected KLF1 model was evaluated with a Ramchandran plot using RAMPAGE and PROCHECK (Lovell *et al.*, 2003). The Ramchandran plot showed 221 amino acid residues in the favoured region (61.4%), 75 amino acid residues in the allowed region (20.8%) and 64 amino acids (17.8%) in the outlier region (Figure 1).

**Fig.1** Validation of KLF1 protein structure. A. Multiple threading alignments. B. Ramachandran plot of KLF1, The plot calculations were based on the 3D model of KLF1 using RAMPAGE C. The Ramachandran plot D. Verified 3D results plot shows 80% of the amino acids (scored  $\geq 0.2$ )

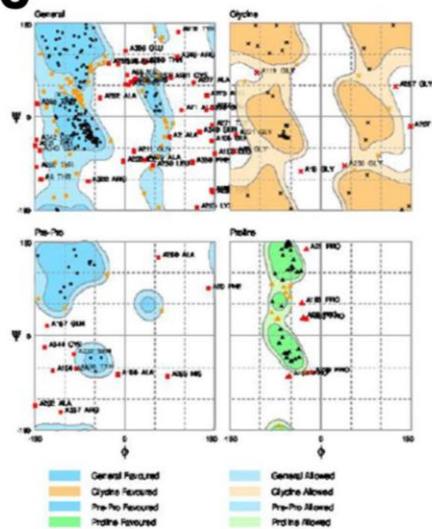
**A**



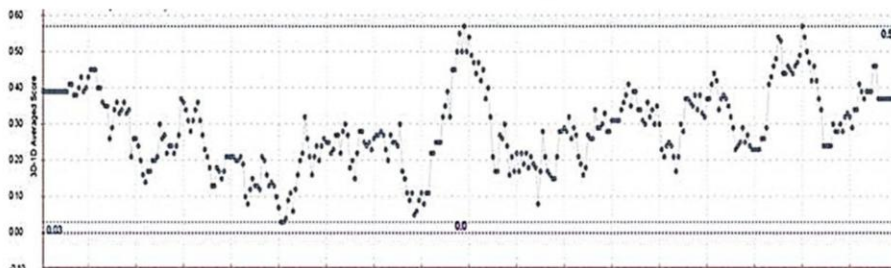
**B**



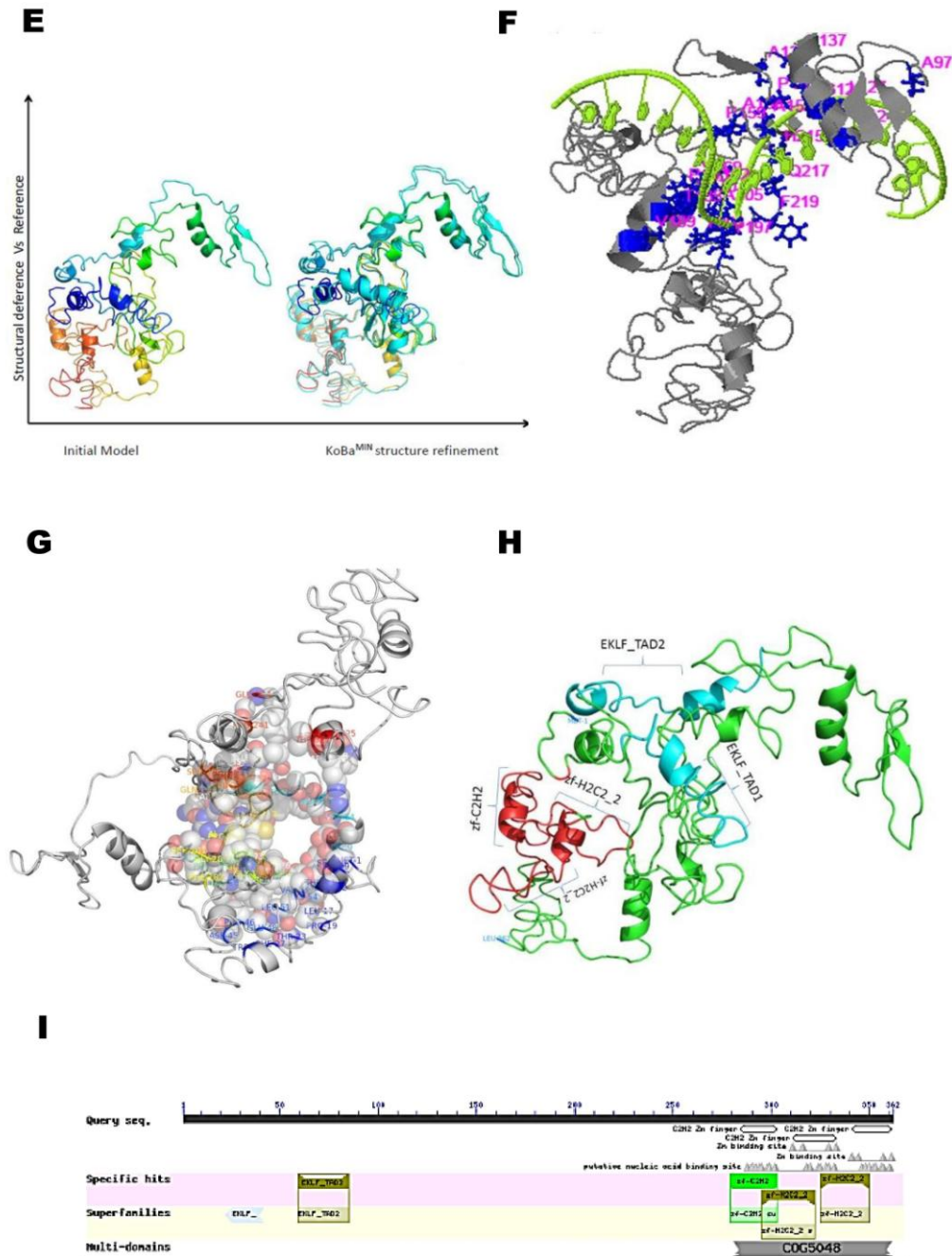
**C**



**D**



**Fig.2** Prediction of the structure of human KLF1 protein. E. Structural differences between the initial and final models of KLF1 with RMSD 1.653. F. Active site of KLF1. Green represents the DNA ligand, blue denotes active site molecules of KLF1, and grey indicates KLF1 protein. G. Active site of KLF1 predicted by CASTp and Visualized by Pymol. H. Homology models of 3D structures of KLF1. The  $\alpha$ -helices and  $\beta$ -sheets are shown as helices and ribbons, and the remaining structure is shown as loops. The 3D structure was generated using PyMOL. I. Conserved domains on KLF1



**Table.1** Overall statistics and structure quality factors of KLF1

S. No	Structure Quality Factors	Mean score	Z-score g
1	Procheck G-factor e (phi / psi only)	-2.03	-7.67
2	Procheck G-factor e (all dihedral angles)	-1.70	-10.05
3	Verify3D	0.28	-2.89
4	ProsaII (-ve)	-0.20	-3.52
5	MolProbityclashscore	54.76	-7.87

**Table.2** The active site residues of KLF1 predicted

Servers	Residues
COACH	A97, A124, L127, G128, R137, A139, P151, A152, A158, P159, V169, Y170, P182, P183, T185, V189, Y197, Y203, A205, H215, Q217, F219
CastP	M1, P9, S10, S12, T13, L17, P19, T23, F27, T31, D45, P46, F48, L51, V53, K54, Q56, D57, G59, D64, D65, G68, A69, T72, W73, L75, L78, F82, A139, L140, R143, P157, E160, L166, P168, V169, Y170, P171, G172, G174, S178, P182, S201, G202, P204, A205, Y207, P208, A209, Q211, Y212, H215, Q217, G221, Q223, S234, L236, P241, G314, D324, Q325, T327, Y330.

The quality of the structure was analysed by PSVS Version 1.5 (Luthy *et al.*, 1992), which identifies the secondary structure alpha helices (60-63aa A, 77-80 aa, 123-134 aa, 185-192 aa, 241-248 aa, 295-302 aa, 325-328 aa) and the beta strands (100-102 aa, 111-113 aa, 139-140 aa, 147-148 aa, 170-171 aa, 178-179 aa). All dihedral angles were measured by Procheck G-factor (phi/psi only), and the overall Z-score was found to be -7.67. At least 80% of the amino acids have scored  $\geq 0.2$  in the 3D/1D profile (Figure1) (Table 1). Knowledge-based potential refinement was performed for KLF1 using the KoBaMIN web server. The KLF1 structure was refined through energy minimization using C $\alpha$ RMS, GDT-HA, and GDT-TS based on the calculated potential of the mean force, and the final refinement structure was superimposed with the initial and final structures (Figure2). Prediction of ligand binding sites in KLF1 determined that there are 22 amino acids in the active site (A97, A124, L127, G128, R137, A139, P151, A152, A158, P159, V169, Y170, P182, P183, T185, V189, Y197, Y203,

A205, H215, Q217, and F219) as predicted by COACH and CASTp servers (Dundas *et al.*, 2006) (Table 2) (Figure2).

### Conserved domains

The final predicted KLF1 structure was subjected to identification of the conserved regions using NCBI conserved domain tools. A zinc-finger double domain at residues 295-32 and 325-350 and one classical zinc finger domain at residues 279-303 were predicted. EKLF\_TAD1 was predicted as the first segment of the minimal transactivation domain of erythroid-specific transcription factor at residues 22-42. EKLF\_TAD2 was predicted as the first segment of the minimal transactivation domain of erythroid-specific transcription factor at residues 59-85 (Figure 2) (Marchler-Bauer *et al.*, 2014).

In the current study, homology modelling of KLF1 was designed using *in silico* methodologies. The *KLF1* gene is responsible for coding a haematopoietic-specific

transcription factor that stimulates adult beta-globin, globin gene switching and erythrocyte maturation. Mutations in KLF1 can ameliorate the severity of  $\beta$ -thalassemia, borderline HbA<sub>2</sub>, congenital dyserythropoietic anaemia and sickle cell disease. The obtained homology results of the KLF1 model shows a high C-score of 3.79, which signifies that it is the best predicted model of KLF1 using homology modelling principles. The optimal 3D structural model was validated with the KoBaMIN web server, which obtained the lowest RMSD of 1.26 Å. Structure quality analysis of the KLF1 model in PSVS resulted in a Procheck G-factor Z-score of -7.67 and a Verify3D Z-score of -2.89 (Figure 1). The obtained protein-ligand binding site prediction shows that 22 amino acids are active (Figure1). The DNA ligand tightly binds with KLF1, and two conserved domains were predicted to be important for ligand binding: first, a zinc finger double domain at residues 295-32 and 325-350, and second, a classical zinc finger domain at residues 279-303. These observations of the KLF1 protein structure and its binding sites could aid in the development of new drugs through structure-based drug design of novel compounds to treat sickle cell disease and beta thalassemia.

### **Acknowledgements**

I thank Dr. J. Francis Borgio for his valuable suggestions.

### **References**

- AbdulAzeez S, Borgio JF (2016) In-Silico computing of the most deleterious nsSNPs in HBA1 gene. *PloS one* 11: e0147702.
- Borgio JF, AbdulAzeez S, Naserullah ZA, Al-Jarrash S, Al-Ali RA, Al-Madan MS, Al-Muhanna F, Al-Suliman AM, Al-Nafie A, Steinberg MH, Al-Ali AK (2016) Mutations in the  $\beta$ -globin gene from a Saudi population: an update. *Int J Lab Hematol* 38: e38-40.
- Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34: W116-8.
- Eisenberg D, Lüthy R, Bowie JU (1997) [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol* 277: 396-404.
- Katoh K, Misawa K, Kuma KI, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059-66.
- Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by C $\alpha$  geometry:  $\phi$ ,  $\psi$  and C $\beta$  deviation. *Proteins: Structure, Function, and Bioinformatics* 50: 437-50.
- Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356: 83.
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Res* gku1221.
- Perkins A, Xu X, Higgs DR, Patrinos GP, Arnaud L, Bieker JJ, Philipson S (2016) Krüppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood* 127: 1856-62.
- Pilon AM, Ajay SS, Kumar SA, Steiner LA, Cherukuri PF, Wincovitch S, Anderson SM, Center NC, Mullikin JC, Gallagher PG, Hardison RC (2011) Genome-wide ChIP-Seq reveals a dramatic shift in the

- binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. *Blood* 118: e139-48.
- Ramachandran GN, Ramakrishnan C, Sasisekharan V (1990) Stereochemistry of polypeptide chain configurations. *Current Science* 59: 813-8.
- Rodrigues JP, Levitt M, Chopra G (2012) KoBaMIN: a knowledge-based minimization web server for protein structure refinement. *Nucleic Acids Res* 7: gks376.
- Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5: 725-38.
- Šali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* 234: 779-815.
- Tallack MR, Whittington T, Yuen WS, Wainwright EN, Keys JR, Gardiner BB, Nourbakhsh E, Cloonan N, Grimmond SM, Bailey TL, Perkins AC (2010) A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells (vol 20, pg 1052, 2010). *Genome Research* 20: 1748.
- Vassen L, Beauchemin H, Lemsaddek W, Krongold J, Trudel M, Möröy T (2014) Growth factor independence 1b (gfi1b) is important for the maturation of erythroid cells and the regulation of embryonic globin expression. *PloS one* 9: e96636.
- Viprakasit V, Ekwattanakit S, Riolueang S, Chalaow N, Fisher C, Lower K, Kanno H, Tachavanich K, Bejrachandra S, Saipin J, Juntharanyom M (2014) Mutations in Krüppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression. *Blood* 123: 1586-95.
- Vyas VK, Ghatge M, Patel K, Qureshi G, Shah S (2015) Homology modeling, binding site identification and docking study of human angiotensin II type I (Ang II-AT 1) receptor. *Biomedicine & Pharmacotherapy* 74: 42-8.
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. *Nature methods* 12: 7-8.
- Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* 43: W174-81.

#### **How to cite this article:**

Sayed Abdul Azeez. 2018. Homology-Based Structure Prediction of the Human Kruppel-Like Factor 1 Protein. *Int.J.Curr.Microbiol.App.Sci.* 7(06): 2169-2176.  
doi: <https://doi.org/10.20546/ijemas.2018.706.257>