Indian J. Applied & Pure Bio. Vol. 38(2), 908-919 (2023). **ISSN: 0970-2091**
A web of Science Journal

Homology modeling, cloning, overexpression, and purification of GST-tagged full-length and C-terminal domain of Human Interferon Regulatory Factor 6

[**]Binita Kumari Sinha, Priyabrata Meher, Sanju Kumari Singh, Sindhuja Snehi, and [*]Tara Kashav

Department of Life Science, Central University of South Bihar, SH-7, Gaya Panchanpur Road, Village- Karhara, Post-Fatehpur, Gaya-824236 (India)
Email: [**]binitasinha@cusb.ac.in, [*]tarakashav@cub.ac.in

Abstract

Human Interferon Regulatory Factor 6 (Hu-IRF6) protein is a transcription factor, a member of the IRF family. IRF6 is involved in palatogenesis, gland development, and acts as a tumor suppressor. The purpose of the present study is to perform homology modeling to predict the structure, clone, overexpress and optimize the purification of IRF6 protein in native conditions. The full-length-IRF6 gene and CTD-IRF6 were cloned, overexpressed, and proteins were purified under non-denaturing conditions. Low temperature (16^{o}C) with a low concentration of IPTG induction helps in the proper folding of protein which enhances protein solubilization. A significant improvement in protein solubility, stability, and yield was observed with additives like L-Arginine, ethylene glycol, LiCl, and fructose. The homology model shows the presence of a long linker connecting the NTD to CTD, this flexible linker region may be the reason for the less stability of full-length IRF6 protein. The addition of additives during purification is helpful to increase the IRF6 protein yield and stability in a soluble fraction which can be scale-up for large-scale purification of protein.

Key words : Interferon Regulatory Factor 6 (IRF6); affinity chromatography; protein purification; protein solubility; additive screening; L-Arg; ethylene glycol; Homology modeling.

Hu-IRF6 belongs to the nine-member IRFs family, consisting of IRF1 to IRF9[29]. All IRF family members share a highly conserved N- terminal, pentad tryptophan repeat, helix-turn-helix, DNA binding domain (DBD), and less well-conserved C-terminal protein binding domain that varies among family members[26,40]. The C-terminal regions of IRFs carry an IRF association domain (IAD) and a transactivation domain (TAD), which mediate protein-protein interactions. The mechanism of activation of IRFs in the cytoplasm is triggered by the

phosphorylation of Serine/Threonine residues in a CTD resulting in dimerization which is a key step in activation and its nuclear translocation[5,20]. Phosphorylation of serine at positions 413 and 424 triggers dimerization and transactivation function which is potentiated by Receptor-interacting protein kinase 4 (RIPK4)[15,16].

Recent studies have illustrated the role of IRF6 in the regulation of several cellular functions such as craniofacial development including palatal development[13], skin development including keratinocyte differentiation[10,15], and development of epidermis[31], cell cycle[1], as a tumor suppressor in various carcinomas, and immune response. IRF6 modulates keratinocyte proliferation, differentiation, and migration which is essentially required for wound healing[2,33]. IRF6 acts as a suppressor for breast carcinoma[1], squamous cell carcinoma[3], gastric cancer[19], urothelial cancer[11], renal cancer[25], nasopharyngeal carcinoma[44], glioma[24], and post-stroke brain damage[8,21]. IRF6 plays a crucial role in regulating TLR-dependent pathways in a cell type-specific fashion[3,9,29].

Mutation in the human *IRF6* gene disrupts orofacial development and causes orofacial clefting as in Van der Woude syndrome (VWS; OMIM #119300) and popliteal pterygium syndrome (PPS; OMIM #119500)[6,14,41]. IRF6 plays a critical role in non-syndromic cleft lip with or without cleft palate (NSCLP) susceptibility[14]. Lip pits and orofacial (cleft lip (CL), cleft lip and palate (CLP), or cleft palate only (P)) are commonly observed in VWS as well as PPS. VWS accounts for approximately 2% of all patients with CL/P, with a prevalence of 1/34,000 live births[4]. Aberrant expression of IRF6 results in the cause of various carcinomas like melanoma[30], oral squamous cell carcinoma[27], Head and Neck squamous cell carcinoma[38], vulvar squamous cell carcinoma[34], lung carcinoma[23], colon and rectal carcinoma[37]. IFN-β1b/IRF6 pharmacogenetic association was identified, which is involved in the promotion of liver damage due to the higher risk of drug-induced liver injury in COVID-19 patients[39].

Here, we are reporting the homology modeling, cloning, overexpression optimization for growth media and temperature, additive screening during cell lysis, and protein purification by affinity chromatography to purify the human Interferon regulatory factor 6. Taking leads from homology modeling for a stable C-terminal domain, the lengths of the deletion construct were decided and the gene fragment was cloned in the pGEX-2TK expression vector. The protein was overexpressed, and purified successfully by using additive screening, growth media, and temperature screening. To enhance the protein solubility and stability, screening of various additives was performed and satisfactory results were obtained by the addition of, L-Arg, LiCl, ethylene glycol, and fructose. The protein purification of GST fused full-length human IRF6 and, its C-terminal domain was performed in non-denaturing conditions.

Chemicals and reagents :

pGEX-2TK expression plasmid (GE Healthcare Life Science, USA), *Escherichia*

coli DH5α, and BL21 DE3 P_{LYS} S cells (Zymo Research, Epigenetix, USA) were commercially purchased. Gel extraction kit & PCR purification kit (Qiagen, Germany), *Taq* DNA polymerase master mix (G-Biosciences, USA), restriction enzymes (New England Biolabs Ltd. UK), T4 DNA ligase, DNaseI (New England Biolabs Ltd., UK), RNase, PMSF (G-Biosciences, USA), Glutathione agarose affinity resin (Qiagen, Germany) and molecular biology grade reagents were obtained from Sigma Aldrich (USA). Primers were synthesized by G-Biosciences.

In-silico analysis of FL-IRF6 and CTD-IRF6

The protein sequence of IRF6 of *Homo sapiens* (accession no. O14896) was retrieved from the UniProt database (www.uniprot.org/). Homology modeling of IRF6 was performed using the SWISS-MODEL server[36, 43]. In absence of any full-length template, the domain was modeled by the server. CTD-IRF6 was modeled using the C-terminal domain structure of the Interferon Regulatory factor (PDB ID: 3DSH). The 3D model quality was validated by the Ramachandran plot using the online server PROCHECK[17]. The online server PyMOL (https://pymol.org/) was used to visualize, edit, and analyze the .pdb files. Recently, the Hu-FL-IRF6 model is available on the 3D Alpha Fold structure database (AlphaFoldDB)[12,42].

Cloning of FL-IRF6 and CTD-IRF6 :

The codon-optimized 1401 bp coding sequence for human IRF6 was cloned between *Bam*HI and *Eco*RI restriction endonuclease sites of the pUC57 cloning vector, which is a kind gift by Prof. W. Royer (UMASS, U.S.A.).

Double digestion of pUC57_FL-IRF6 was performed using the same restriction enzymes followed by gel extraction of digested FL-IRF6 and was sub-cloned into pGEX-2TK expression vector carrying Glutathione S-Transferase (GST) tag.

The gene corresponding to the CTD of IRF6 was amplified using a semi-nested Polymerase Reaction chain using clone pUC57-IRF6 as a template. The PCR amplification of 661-1275 nucleotides was done using *Taq* polymerase with the following forward and reverse primer:
F.P.- 5' CGGGATCCATGCTGCCGATGA 3',

R.P.- 5' CGGAATTCTTACGTGCTAATTTGCA GGCG 3'. The amplified DNA was subcloned using the same restriction enzymes into the pGEX-2TK expression plasmid vector which includes a GST tag at the N-terminus of the recombinant protein. The confirmation of the clone was done by double digestion of the plasmid using the same restriction enzymes. The calculated molecular weight pGEX-2TK_$CTD_{221-425}$ construct is approximately 49.498 kDa including a 26 kDa GST tag as calculated using ExPASy ProtParam (www.expasy.org/resources/protparam).

DNA sequencing of recombinant clones:

The recombinant clones (FL-IRF6 and CTD-IRF6) were verified by DNA sequencing using universal primers pGEX 5' & pGEX 3'.

Optimization of overexpression of recombinant FL-IRF6 & CTD-IRF6 :

The plasmid of clone pGEX-2TK_FL-IRF6 and pGEX-2TK_$CTD_{221-425}$-IRF6 were

transformed into *E. coli*. strain BL21 DE3 P_{Lys} S cells. A single transformed bacterial cell was inoculated into 10 mL Luria–Bertani (LB) broth containing 100 µg/mL ampicillin and agitated overnight in an incubator shaker at 200 rpm and 37°C. The secondary bacterial culture was then inoculated into 100 mL LB supplemented with 100 µg/mL ampicillin and grown at 220 rpm at 37°C. After the OD_{600} had reached 0.5-0.6, expression of the recombinant FL-IRF6 protein was induced by adding 1 mM isopropyl-d-1-thiogalactopyranoside (IPTG) while the recombinant CTD protein was induced by 0.4 mM IPTG. The culture was induced for 03 hrs at 37°C and samples were collected to analyze expression. The cells were harvested by centrifugation at 6000 rpm and 4! for 10 minutes and frozen. The fractions were detected by SDS-PAGE gel.

Soluble expression and purification of FL-IRF6 & CTD-IRF6 using additives :

The harvested pellets were resuspended in ice-cold lysis buffer consisting of 50 mM HEPES, 150 mM NaCl, 10 mM $MgCl_2$, 5% glycerol, and 0.1% triton X-100. The pH of the lysis buffer for FL-IRF6 and CTD-IRF6 was 6.8 and 8.0 respectively. After adding 10 µg/mL RNase, 10 µg/mL DNase, 1 mM Phenylmethylsulphonyl fluoride (PMSF), 5 mM β-mercaptoethanol, and lysozyme (1 mg/mL), cell lysis was completed by using 5 cycles of the freeze and thaw method. The cell lysate was further centrifuged at 13,000 rpm for 30 minutes at 4°C to get clear supernatant from the cell debris. The supernatant was loaded onto the Glutathione agarose column and equilibrated with lysis buffer. After extensive washing, the GST-tagged recombinant protein was eluted using elution buffer containing 20 mM glutathione reduced, 50 mM HEPES, 150 mM NaCl, 10 mM $MgCl_2$, 5% Glycerol, pH 8.0. Overexpression and GST fusion protein purification were carried out as described by Harper *et al*.[7]. The fractions were loaded on 10% SDS-PAGE gel and results were visualized by Coomassie staining.

Primary additive screening and additives-based purification of recombinant FL-IRF6 & CTD-IRF6 :

The primary additive screening was performed using 1 mL homogenized slurry aliquoted. The same lysis buffer was used to solubilize the pellets. Additives namely 100 mM L-Arginine (L-Arg), 100 mM lithium chloride (LiCl), 5% ethylene glycol, 5% sucrose, and 0.1 M L-proline were used to solubilize both FL & CTD-IRF6 protein. Additionally, 100 mM L-Glutamic acid, 100 mM Sodium citrate, 5% D-Sorbitol, and 5% Xylitol were used to solubilize FL-IRF6 protein only, and for CTD-IRF6, 5% glucose, 100 mM KCl, 5% D-Xylose, 5% fructose, 100 mM glycine, 100 mM ammonium sulfate ($(NH_4)_2SO_4$), and 100 mM dipotassium hydrogen phosphate (K_2HPO_4) were used. Data is not shown for additive, glycine, ($(NH_4)_2SO_4$), K_2HPO_4, and L-proline. After lysis, samples were centrifuged at 4°C for 30 minutes at 13,000 rpm. The supernatant and pellet were analyzed on 10% SDS-PAGE gel. For enhancement of solubilization, bacterial growth media was also changed from LB to Terrific Broth (TB) for culture, and induction temperature was decreased from 30°C to 16°C (low temperature). Selected additives-based purification was done for both recombinant proteins.

In-silico analysis of FL-IRF6 and CTD-IRF6

The schematic in Fig. 1.A. illustrates that FL-IRF6 is composed of 467 amino acids. The N-terminal domain consists of conserved penta-tryptophan and, the C-terminal domain consists of IRF association domain 1 (IAD1) and a transactivation domain (TAD). FL-IRF6 model from AlphaFold defines structurally that it consists of twelve α-helices and fourteen β-sheet (Fig. 1.B.). The NTD consists of four α-helices and four β-sheet which form the helix-turn-helix motif while CTD consists of eight α-helices and ten β-sheet which form the β-sandwich core flanked by α-helices. Alpha Fold model quality of FL-IRF6 was validated by the Ramachandran plot (Fig. 1.C.). The number of residues in the most favored region was 184 (92.5%), the number of residues in the additional allowed region was 14 (7.0%), the number of residues in the generously allowed region was 01 (0.5%) and the number of residues in the outlier region was 0 (0%). The Swiss model server predicts the CTD model of IRF6 (residue no. 214 to 445), dimeric in nature (Fig. 1.D.) taking template Human IRF5, PDB ID: 3DSH sharing 75% homology with CTD-IRF6. The quality of the CTD model was found good as predicted by the Ramachandran plot (Fig. 1.E.). The number of residues in the most favored region was 184 (92.5%), the number of residues in the additional allowed region was 14 (7.0%), the number of residues in the generously allowed region was 01 (0.5%) and the number of residues in the outlier region was 0 (0%).

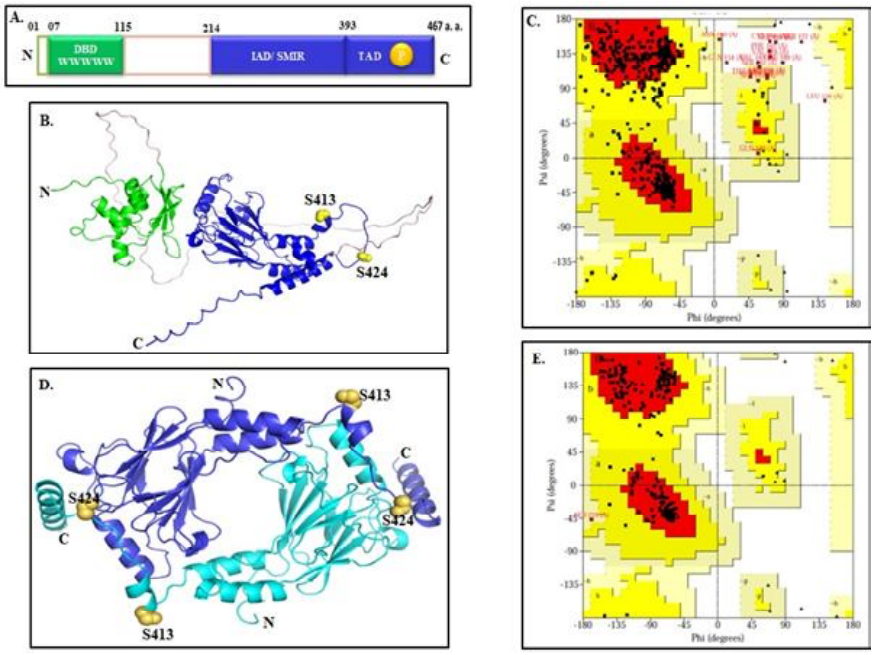

Fig. 1: A. Schematic diagram of Hu-IRF6, **B.** Model of Human FL-IRF6 (AlphaFold server). **C.** Ramachandran plot, for the Alpha FL-IRF6 model. **D.** Model of CTD-IRF6 using SWISS-MODEL server, showing dimerization (SER 413 & SER 424 showing as spheres in yellow color). **E.** Ramachandran plot, for the predicted CTD-IRF6 model.

Cloning of FL-IRF6 and CTD$_{221-425}$-IRF6 into the expression vector :

The full-length IRF6 gene of 1401 bp from plasmid pUC57 was digested with *Bam* HI and *Eco*RI (Fig. 2. A.) and subsequently subcloned into the pGEX-2TK expression vector between the same restriction enzyme (R.E.) sites (Fig. 2.B.). CTD-IRF6 gene was generated by PCR amplification and the amplified product was 618 bp including the start codon (Fig. 2.C.). Both amplified product and expression vector was digested with the same R.E. and ligated. The final construct pGEX-2TK_CTD$_{221-425}$-IRF6 was confirmed by double digestion (Fig. 2.D.). Further, the positive clones were also verified by DNA sequencing.

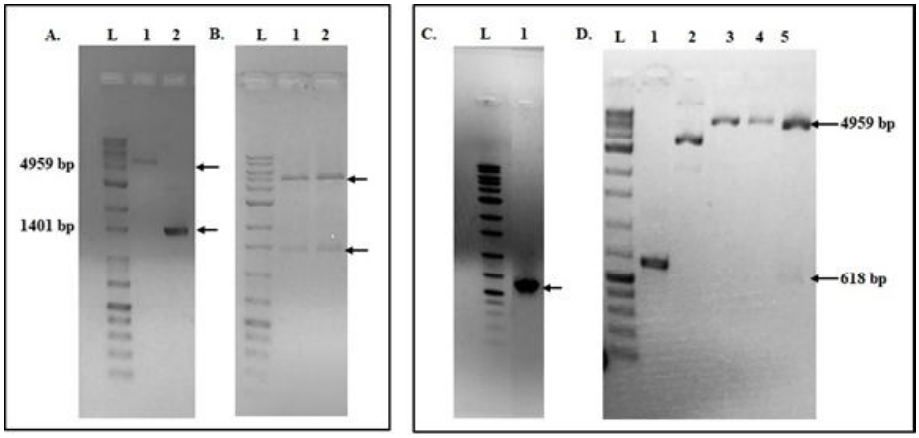

Fig. 2: A. 1% agarose gel showing gel clean-up of L- Ladder, lane 1- digested FL-IRF6, lane 2- digested pGEX-2TK, **B.** Double digestion of pGEX-2TK_FL-IRF6 clone (L- ladder, 1-2: digested products of size 4959 bp pGEX-2TK and 1401 bp FL-IRF6. **C.** PCR amplified product of CTD$_{221-425}$-IRF6 (L-ladder; Lane l: PCR amplified of 681 bp), **D.** Analysis of double digestion of clone CTD$_{221-425}$-IRF6 with *Bam*HI and *Eco*RI; L- ladder; 1: PCR amplified product as positive control; 2: plasmid DNA of clone; 3 & 4: single digestion with *Bam*HI and *Eco*RI respectively; 5: digestion fragments (4959 bp and 618 bp).

Overexpression and purification of FL-IRF6 using additives :

The plasmid of the FL-IRF6 construct was transformed into *E. coli* BL21 (DE3) P_{Lys} S cells. Expression of recombinant FL-IRF6 protein was induced with 1 mM IPTG at 37^{o}C. The result of overexpression of FL-IRF6 protein shows that the molecular weight is 81.42 kDa including a 26 kDa GST tag, clearly indicated on an SDS-PAGE gel (Fig. 3.A.). The two degradation products of size 67 and 45 kDa were also observed which indicated that the recombinant protein is not stable. Further, protein purification was accomplished by affinity chromatography on the Glutathione resin column. The SDS-PAGE analysis showed multiple protein bands of approximately molecular weight 45, 26, and 16 kDa (Fig. 3.B.). The 26 kDa protein band was observed more intense, this could be the GST tag. Due to degradation, the intact recombinant protein of expected molecular mass was not observed during purification.

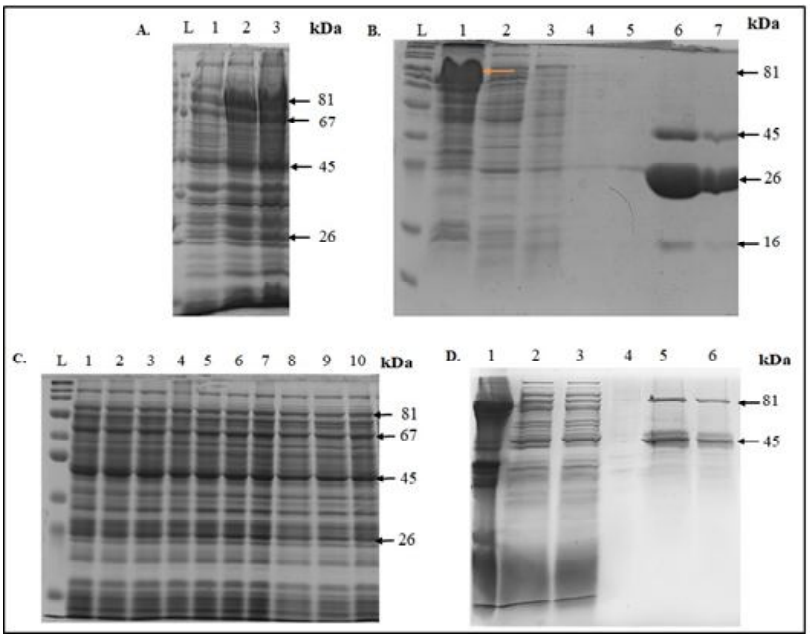

Fig. 3: The 12% SDS-PAGE analysis of FL-IRF6 protein. **A.** Induction with 1 mM IPTG concentration. (L-ladder, 1: Un-induced, 2: induced for 1 hr., 3: induced for 2 hr. **B.** Affinity purification without additives (L: ladder, 1: processed pellet, 2: supernatant, 3: flowthrough, 4: wash, 5-7: fractions of protein elution. **C.** Primary additive screen (L: ladder, supernatant loaded of additive processing 1: L-Arg, 2: LiCl. 3: ethylene glycol 4: L-Glu, 5: Sodium citrate, 6: D-Sorbitol, 7: Xylitol, 8: Sucrose, 9: L-proline, 10: No additive. **D.** Affinity purification with selected additive, 100 mM LiCl (1: processed pellet, 2: Supernatant, 3: flowthrough, 4: wash, 5-7: elution fractions).

To increase the stability and yield of protein, primary screening of cell lysis in presence of ten different additives to facilitate protein stability and solubilization; 7 out of 9 additives had increased the protein yield compared with no additive control which was quantified by Nanodrop spectrophotometer. SDS-PAGE results showed the almost same level of expression in presence of additives and the difference in protein yield was about 2.5-3.5 mg/mL higher in comparison to no additive control (Fig. 3.C.). The additives resulting in increased solubility relatively are ethylene glycol, L-Arg, Sodium citrate, Xylitol, Sorbitol, L-Glu, & LiCl. In the LiCl-based protein purification, the two-protein bands are observed; one of expected molecular mass 81 kDa and other recombinant polypeptides of size 45 kDa on an SDS-PAGE gel. Hence, LiCl enhanced protein stability as well as yield (Fig. 3.D.). However, the N-terminus of the recombinant IRF6 must be intact as GST is fused to the N-terminus facilitating the purification process and degradation might occur from the C-terminal end.

Overexpression, solubility optimization, and purification of CTD$_{221-425}$-IRF6 :

The overexpression of CTD-IRF6 was induced with 0.4 mM IPTG concentration (Fig. 4.A.). The result shows that the molecular mass of recombinant CTD protein is 49.5 kDa including a 26 kDa GST tag which corresponds to the predicted size of pGEX-2TK_CTD-IRF6 as calculated by the ProtParam tool. In addition to the 49 kDa band, a 26 kDa band was observed which could be GST or recombinant polypeptide of size 23 kDa. Due to instability, some amount of GST fusion protein was observed to be degraded. As a result of affinity protein purification, a very less intense protein band was observed (Fig. 4.B.).

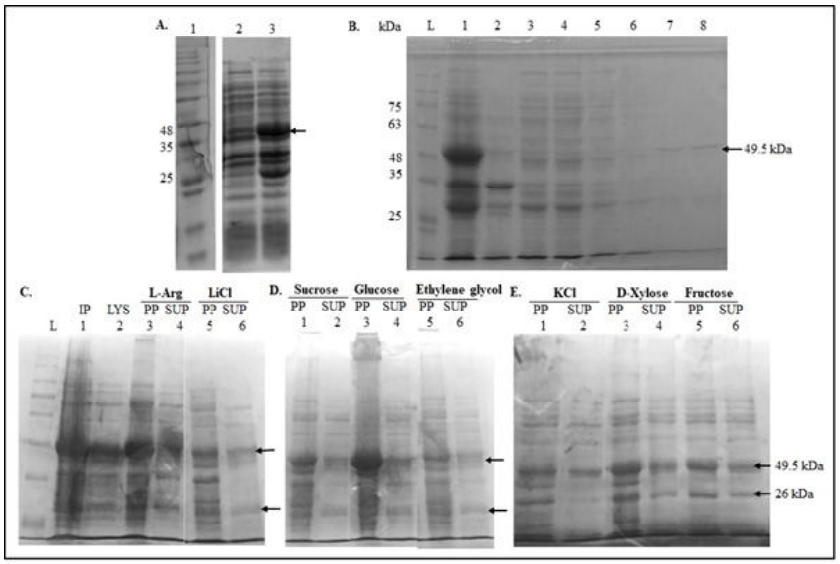

Fig. 4: The 10% SDS-PAGE analysis of CTD-IRF6 protein. **A.** Induction with 0.4 mM IPTG concentration. (L: ladder, 1: uninduced, 2: induced). **B.** Affinity purification without additives (L: ladder, 1: Induced pellet, 2: processed pellet, 3: supernatant, 4: flowthrough, 5: wash, 6-8: elution fractions. **C.** Analysis of the primary additive screen. L: ladder, 1: Induced pellet, 2: Total cell lysate without additive (LYS), 3: L-Arg- processed pellet, 4: L-Arg- supernatant, 5: LiCl- PP 7: LiCl- SUP. **D-** 1: Sucrose- PP, 2: Sucrose- SUP, 3: Glucose- PP, 4: Glucose- SUP, 5: Ethylene glycol- PP, 6: Ethylene glycol- SUP. **E-** 1: KCl- PP, 2: KCl- SUP, 3: D-Xylose- PP, 4: D-Xylose- SUP, 5: Fructose- PP, 6: Fructose- SUP. [Processed pellet (PP) & Supernatant (SUP)].

Like FL-IRF6, the primary additive screen was applied to increase yield and stability. In this additives trial, the supernatant and pellet in presence of 8 additives were analyzed on 10% SDS-PAGE gel (Fig. 4.C-E.). Some amount of recombinant protein gets solubilized as observed in the supernatant, clearly visible on the gel. But still, a high amount of protein remains in inclusion bodies or pellets. Additive-based protein purification

in presence of Ethylene glycol (Fig. 5.B.), fructose (Fig. 5.C.), LiCl (Fig. 5.D.), and L-proline (Fig. 5.E.) was employed to increase the yield of affinity-purified protein. As expected, the yield increased in comparison to without additive-based purified protein (Fig. 5.A.). Although, growth in TB media with induction at low temperature in the presence L-Arg as additive results in a little higher protein yield (Fig. 5.F.) as a significantly intense protein band of concentrated recombinant protein was observed (Fig. 5.G.).

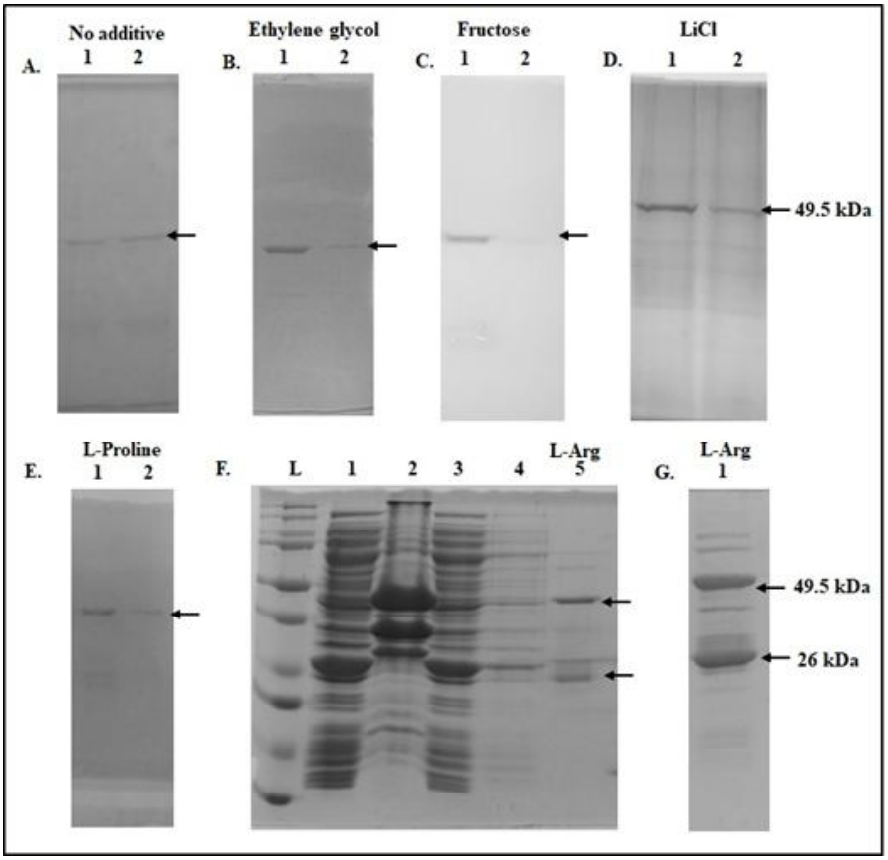

Fig. 5: The 10% SDS-PAGE gels for the analysis for affinity purified protein. **A.** Affinity purification without additives. **B.** Protein purification in presence of additives 3% ethylene glycol (1 & 2: fractions of elution). **C.** Purification in presence of additives 100 mM LiCl (1 & 2: fractions of elution). **D.** Purification in presence of additive 5% D-Fructose (1 & 2: fractions of elution), **E.** Purification in presence of additive 100 mM L-Proline (1 & 2: fractions of elution), **F.** Purification in presence of additive 25 mM L-Arg (L: ladder, 1: supernatant, 2: processed pellet, 3: flow-through, 4: wash, 5: protein elution fractions. **G.** concentrated protein (1: protein elution). The solubility of purified protein enhances and the protein band of 49.5 kDa was observed in fractions of elution.

In this study, the structure of the C-terminal domain of IRF5 was selected as a template for the homology structure prediction of IRF6. The tertiary structure prediction of full-length IRF6 by the AlphaFold server shows that the structured N-terminal domain is separated from the C-terminal domain by a long unstructured linker that exhibits a high level of disordered structure. Full-length IRF6 and CTD-IRF6 were successfully cloned in the pGEX-2TK expression vector followed by its overexpression and protein purification in non-denaturing conditions. As our target was to purify IRF6 recombinant protein under native conditions unlike earlier studies performed under denaturing conditions[22] and, various reports suggest the importance of additive screening to increase protein solubilization and its stability[18,28]. Additives-based screening helped to improve the stability and solubilization of IRF6 protein. The GST-tagged FL-IRF6 protein was completely forming inclusion bodies as recombinant protein expression was found in the pellet (Fig. 3.B.) and the soluble fraction increased in presence of additive LiCl (Fig. 3.D.). Overexpression of $CTD_{221-425}$-IRF6, GST-fused protein is of 49.5 kDa observed (Fig. 4.A.). Unfortunately, the GST moiety was falling off as one polypeptide band observed of 26 kDa due to protein instability, like the IRF6-RIPK4 studies[15]. The higher concentrations of IPTG increased the amount of protein in the insoluble fraction due to the increased expression rate[35], therefore IPTG concentration is optimized for the purification of CTD protein. During protein purification of the CTD, it was observed that a low yield of recombinant proteins was found in soluble form and a large amount of protein remains in an insoluble form or pellets. After solubilization trials with different additives, the protein solubility enhances using standardized additives such as ethylene glycol, and LiCl, especially in the case of L-arginine. The protein solubility is also enhanced by changing the broth and using a standardized additive. During induction, low temperature with low IPTG concentration helps in the proper folding of protein which increases solubilization. Protein yield does increase in soluble fraction but is not completely absent in the lysed pellets.

The full-length IRF6 homology model has the presence of a long unstructured linker region separating the helix-turn-helix motif of NTD from the CTD consisting of eight α-helices and ten β-sheet which form the β-core. The initial overexpression and purification trials of GST-fused FL-IRF6 and CTD-IRF6 both resulted in very less yield along with the degradation of the desired protein. The addition of additives like ethylene glycol, L-arginine, fructose, and LiCl improved the stability and expression of the FL-IRF6 protein. Changing the growth medium from LB to TB media and, the decrease in temperature during overexpression also have positive effects on the expression of proteins. Overall, a promising protein yield could be obtained that can be utilized further for performing biochemical assays.

Dr. T.K. is thankful to the DBT Bio-CARe project, B.K.S. is thankful to the CSIR fellowship and P.M. is thankful to UGC non-net fellowship for supporting this work.

Conflict of interest: The authors declare no conflict of interest.

References :

1. Bailey, C.M., D.E. Abbott, N.V. Margaryan, Z. Khalkhali-Ellis, and M. J. Hendrix, (2008). *Molecular and cellular biology*, *28*(7): 2235-2243.
2. Biggs, L. C., R. L. Naridze, K. A. DeMali, D. F. Lusche, S. Kuhl, D. R. Soll, M. Dunnwald, (2014). *Journal of cell science*, *127*(13): 2840-2848.
3. Botti, E., G. Spallone, F. Moretti, B. Marinari, V. Pinetti, S. Galanti, and A. Costanzo, (2011). *Proceedings of the National Academy of Sciences,* U S A, *108*(33): 13710-13715.
4. Burdick, A. B. (1986). *Journal of craniofacial genetics and developmental biology*, *2:* 99-105.
5. Chen, W., and W. E. Royer, Jr. (2010). *Cellular Signalling*, *22*(6): 883-887.
6. Gorlin, R.J., H.O. Sedano and J. Cervenka (1968). *Pediatrics*, *41*(2): 503-509.
7. Harper, S., and D. W. Speicher, (2011). *Methods in Molecular Biology*, *681:* 259-280.
8. Huang, R., Z. Hu, Y. Feng, L. Yu, and X. Li, (2017). *Scientific reports*, *7*(1): 1-14.
9. Huynh, J., G. M. Scholz, J. Aw, M. Q. Kwa, A. Achuthan, J. A. Hamilton, and E. C. Reynolds, (2016). *Journal of Immunology*, *196*(5): 2230-2238.
10. Ingraham, C. R., A. Kinoshita, S. Kondo, B. Yang, S. Sajan, K. J. Trout, and M. Lovett, (2006). *Nature Genetics*, *38*(11): 1335-1340.
11. Jou, Y.-C., G.-L. Lin, H.-Y. Lin, W.-H. Huang, Y.-M. Chuang, R.-I. Lin, and M. W. Chan (2021). *Cancer cell international*, *21*(1): 1-12.
12. Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, and D. Hassabis (2021). *Nature*, *596*(7873): 583-589.
13. Ke, C.-Y., W.-L. Xiao, C.-M. Chen, L.-J. Lo, and F.-H. Wong, (2015). *Scientific reports*, *5*(1): 1-14.
14. Kondo, S., B. C. Schutte, R. J. Richardson, B. C. Bjork, A. S. Knight, Y. Watanabe, and J.C. Murray (2002). *Nature Genetics*, *32*(2): 285-289.
15. Kwa, M. Q., J. Huynh, J. Aw, L. Zhang, T. Nguyen, E. C. Reynolds, and G. M. Scholz, (2014). *Journal of Biological Chemistry*, *289*(45): 31077-31087.
16. Kwa, M. Q., T. Nguyen, J. Huynh, D. Ramnath, D. De Nardo, P. Y. Lam, and G.M. Scholz (2014). *Journal of Biological Chemistry*, *289*(28): 19758-19768.
17. Laskowski, R., M. MacArthur, and J. Thornton, (2006). *Journal of Applied Crystallography* 1993; *26*(2): 283–291.
18. Leibly, D. J., T. N. Nguyen, L. T. Kao, S. N. Hewitt, L. K. Barrett, and W. C. Van Voorhis, (2012). *PLoS One*, *7*(12): e52482.
19. Li, D., P. Cheng, J. Wang, X. Qiu, X. Zhang, L. Xu and S. Qin (2019). *Frontiers in Oncology*, 9: 220.
20. Lin, R., C. Heylbroeck, P. M. Pitha, and J. Hiscott, (1998). *Molecular and Cell Biology*, *18*(5): 2986-2996.
21. Lin, Y., D. Xu, X. Li, C. Liu, X. Liu, S. Huang, and X. Liu, (2016). *Cellular and Molecular Neurobiology*, *36*(1): 27-36.
22. Little, H. J., N. K. Rorick, L. I. Su, C. Baldock, S. Malhotra, T. Jowitt, and P. Shore, (2009). *Human Molecular Genetics*, *18*(3): 535-545.
23. Liu, Y., G. Shao, Z. Yang, X. Lin, X. Liu, B. Qian, and Z. Liu, (2021). *Journal of Pharmacy and Pharmacology*, *73*(5): 682-691.

24. Lu, J., X. Liu, J. Zheng, J. Song, Y. Liu, X. Ruan, and D. Wang (2020). *Cell Death and Disease*, *11*(6): 1-16.
25. Ma, X., X. Wang, Q. Dong, H. Pang, J. Xu, J. Shen, and J. Zhu, (2021). *Cancer cell international*, *21*(1): 1-14.
26. Mamane, Y., C. Heylbroeck, P. Génin, M. Algarté, M. J. Servant, C. LePage, and J. Hiscott, (1999). *Gene*, *237*(1): 1-14.
27. Melath, A., G. K. Santhakumar, S. S. Madhavannair, B. M. Nedumgottil, and A. Ramanathan, (2013). *Asian Pacific Journal of Cancer Prevention*, *14*(11): 6803-6806.
28. Mishra, S. K., R. Verma, P. Kumari, A. Kumari and K. Prakash (2021). *Molecular Biology Reports*, *48*(3): 2307-2314.
29. Nguyen, H., J. Hiscott, and P. M. Pitha, (1997). *Cytokine and Growth Factor Reviews*, *8*(4): 293-312.
30. Nobeyama, Y., and H. Nakagawa, (2017). *PLoS One*, *12*(9): e0184444.
31. Oberbeck, N., V. C. Pham, J. D. Webster, R. Reja, C. S. Huang, Y. Zhang, and A. Birnberg, (2019). *Nature*, *574*(7777): 249-253.
32. Ramnath, D., K. Tunny, D.M. Hohenhaus, C. M. Pitts, A. S. Bergot, P. M. Hogarth, . . . and G. M. Scholz, (2015). *Immunology and cell biology*, *93*(9): 771-779.
33. Rhea, L., F. J. Canady, M. Le, T. Reeb, J. W. Canady, D. S. F. Kacmarynski, and M. Dunnwald, (2020). *Developmental Dynamics*, *249*(4): 509-522.
34. Rotondo, J. C., A. Borghi, R. Selvatici, E. Magri, E. Bianchini, E. Montinari, and F. Martini, (2016). *JAMA Dermatology*, *152*(8): 928-933.
35. Santosh, K. M., K. Nitish, K. Gautam, K. Tara, and P. Krishna, (2018). *Molecular Biology Reports*, *45*(5): 1367-1374.
36. Schwede, T., J. Kopp, N. Guex, and M.C. Peitsch, (2003). *Nucleic Acids Research*, *31*(13): 3381-3385.
37. Slattery, M. L., A. Lundgreen, K. L. Bondurant, and R. K. Wolff, (2011). *Carcinogenesis*, *32*(11): 1660-1667.
38. Stransky, N., A. M. Egloff, A. D. Tward, A. D. Kostic, K. Cibulskis, A. Sivachenko, and A. McKenna (2011). *Science*, *333*(6046): 1157-1160.
39. Takahashi, T., J. A. Luzum, M. R. Nicol, and P. A. Jacobson, (2020). *NPJ Genomic Medicine*, 5: 35.
40. Tamura, T., H. Yanai, D. Savitsky, and T. Taniguchi, (2008). *Annual Review of Immunology*, *26:* 535-584.
41. Van der Woude, A. (1954). *American journal of human genetics*, *6*(2): 244.
42. Varadi, M., S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, and S. Velankar, (2021). *Nucleic Acids Research*, *50*(D1): D439-D444.
43. Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, and T. Schwede (2018). *Nucleic Acids Research*, *46*(W1): W296-w303.
44. Xu, L., T.-J. Huang, H. Hu, M.-Y. Wang, S.-M. Shi, Q. Yang, and Y.-H. Lang, (2018). *Cancer letters*, *431:* 230-243.