

Structural Modeling and DNA Binding Autoinhibition Analysis of Ergp55, a Critical Transcription Factor in Prostate Cancer

Shanti P. Gangwar¹, Sharmistha Dey², Ajay K. Saxena^{1*}

1 Structural Biology Lab, School of Life Sciences, Jawaharlal Nehru University, New Delhi, India, **2** Department of Biophysics, All India Institutes of Medical Sciences, New Delhi, India

Abstract

Background: The Ergp55 protein belongs to Ets family of transcription factor. The Ets proteins are highly conserved in their DNA binding domain and involved in various development processes and regulation of cancer metabolism. To study the structure and DNA binding autoinhibition mechanism of Ergp55 protein, we have produced full length and smaller polypeptides of Ergp55 protein in *E. coli* and characterized using various biophysical techniques.

Results: The Ergp55 polypeptides contain large amount of α -helix and random coil structures as measured by circular dichroism spectroscopy. The full length Ergp55 forms a flexible and elongated molecule as revealed by molecular modeling, dynamics simulation and structural prediction algorithms. The binding analyses of Ergp55 polypeptides with target DNA sequences of E74 and cfos promoters indicate that longer fragments of Ergp55 (beyond the Ets domain) showed the evidence of auto-inhibition. This study also revealed the parts of Ergp55 protein that mediate auto-inhibition.

Significance: The current study will aid in designing the compounds that stabilize the inhibited form of Ergp55 and inhibit its binding to promoter DNA. It will contribute in the development of drugs targeting Ergp55 for the prostate cancer treatment.

Citation: Gangwar SP, Dey S, Saxena AK (2012) Structural Modeling and DNA Binding Autoinhibition Analysis of Ergp55, a Critical Transcription Factor in Prostate Cancer. PLoS ONE 7(6): e39850. doi:10.1371/journal.pone.0039850

Editor: Vladimir N. Uversky, University of South Florida College of Medicine, United States of America

Received: May 5, 2012; **Accepted:** May 31, 2012; **Published:** June 28, 2012

Copyright: © 2012 Gangwar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The work was supported by grant from the Department of Science and Technology (DST) New Delhi, India (No-SR/SO/HS-05/2009). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ajaysaxena@mail.jnu.ac.in

Introduction

The Ets family proteins share highly conserved winged helix-turn-helix DNA-binding domain and bind to consensus DNA core sequence 5'-GGA (A/T)-3' [1]. The Erg proteins belong to Ets family of transcription factor. Erg gene is rearranged in human myeloid leukemia [2] and in 5–10% of patients with Ewing's sarcoma [3]. In both cases, chromosomal translocations results in the expression of oncogenic fusion proteins composed of Erg and member of Tet subfamily of RNA binding proteins. Erg protein is essential for definitive hematopoiesis, adult hematopoietic stem cell function and maintenance of normal peripheral blood platelet numbers [4]. The TMPRSS2-Erg fusion oncogene transcripts observed in prostate cancer cells are significantly associated with aggressive cancer, metastatic spread and increased probability of death [5].

The Erg gene encodes five proteins, Erg-1, Erg-2, Ergp55, Ergp49 and Ergp38 as a result of different splicing, polyadenylation or initiation codon. The Ergp55 isoform contains four functional domains, which are involved in DNA binding, transcriptional activation and negative regulation of transactivation [6]. The Ergp55 protein forms dimer with itself and with two other isoforms Ergp49 and Ergp38, via PNT and Ets domain [7].

The central domain of Ergp55 behaves as inhibitory domain on dimerization and its removal enhances the transactivation property [7]. The critical residues of Ets domain of Ergp55, which mediate Ergp55-jun/fos-DNA ternary complex formation, have been identified and characterized [8].

So far, tertiary structure of any full-length Ets protein is not determined. However, structures of DNA-binding domains of several Ets proteins have been determined using X-ray crystallography and nuclear magnetic resonance techniques [9–16]. The genome-wide analysis of Ets-family DNA-binding *in vitro* and *in vivo* has been studied recently [17].

The precise mechanism by which, Ergp55 protein acts on transcription is not understood. To understand the structure and DNA binding autoinhibition mechanism of Ergp55, we performed circular dichroism, molecular modeling and theoretical structural prediction analysis on Ergp55 polypeptides. To understand the DNA binding autoinhibition mechanism, the binding studies of Ergp55 polypeptides with DNA sequences of E74 and cfos promoters were carried out. Our results indicated that (i) Ergp55 polypeptides contains high percentage of α -helix and random coil structures (ii) full length Ergp55 is a flexible and elongated molecule (iii) longer fragments beyond the canonical Ets domain of Ergp55 showed the evidence of autoinhibition.

Materials and Methods

Construction of Ergp55 plasmids

The full length Erg₁₋₄₇₉ and Erg₁₁₂₋₃₉₉ genes were cloned in pET28a (+) vector. The Erg₁₋₃₉₉ and Erg₃₀₇₋₃₉₉ genes were cloned in pRSETA and pET21a (+) expression vector. All deletion mutant genes were amplified from full-length Erg₁₋₄₇₉ plasmid using polymerase chain reaction.

Purification of Ergp55 polypeptides

The Erg₁₋₄₇₉, Erg₁₋₃₉₉, Erg₁₁₂₋₃₉₉ and Erg₃₀₇₋₃₉₉ plasmids were transformed into *E. coli*. BL21 (DE3) cells. The cells were grown in 3 liters luria bertani media containing appropriate antibiotics at 37°C, until OD₆₀₀ reached to 0.6. The 0.125 mM IPTG was induced at 37°C and culture further shaken for 4 h at 220 rpm. The cells were harvested by centrifugation at 6000x *g* for 10 min at 4°C. The cells were suspended in 50 ml of lysis buffer containing (25 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM benzamidine-HCl, 0.1% triton X-100, 5% glycerol, 3 mM 2-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride and

0.5 mg/ml lysozyme) and disrupted by sonication at 4°C. The crude lysate was centrifuged at 25000x *g* for 20 min at 4°C. The supernatant was loaded on Ni-NTA column, pre-equilibrated with binding buffer (25 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM benzamidine-HCl, 5% glycerol, 2 mM 2-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride and 10 mM imidazole). The proteins were eluted from column with elution buffer (25 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM benzamidine-HCl, 5% glycerol, 3 mM 2-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride and 250 mM imidazole). The eluted fractions were pooled, concentrated and loaded on Sephacryl S-200 HR gel-filtration column pre-equilibrated in buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 3 mM 2-mercaptoethanol and 5% glycerol). The purified proteins were concentrated to 10 mg/ml using amicon ultra centrifugal filter device (Mw cutoff~ 10 kD). The protein concentration was measured by UV radiation at 280 nm by using extinction coefficient calculated with EXPASY software (<http://us.expasy.org/tools/protparam.html>).

N-terminal protein sequencing and ion spray mass spectrometry confirmed identity and purity of Ergp55 polypeptides. Protein

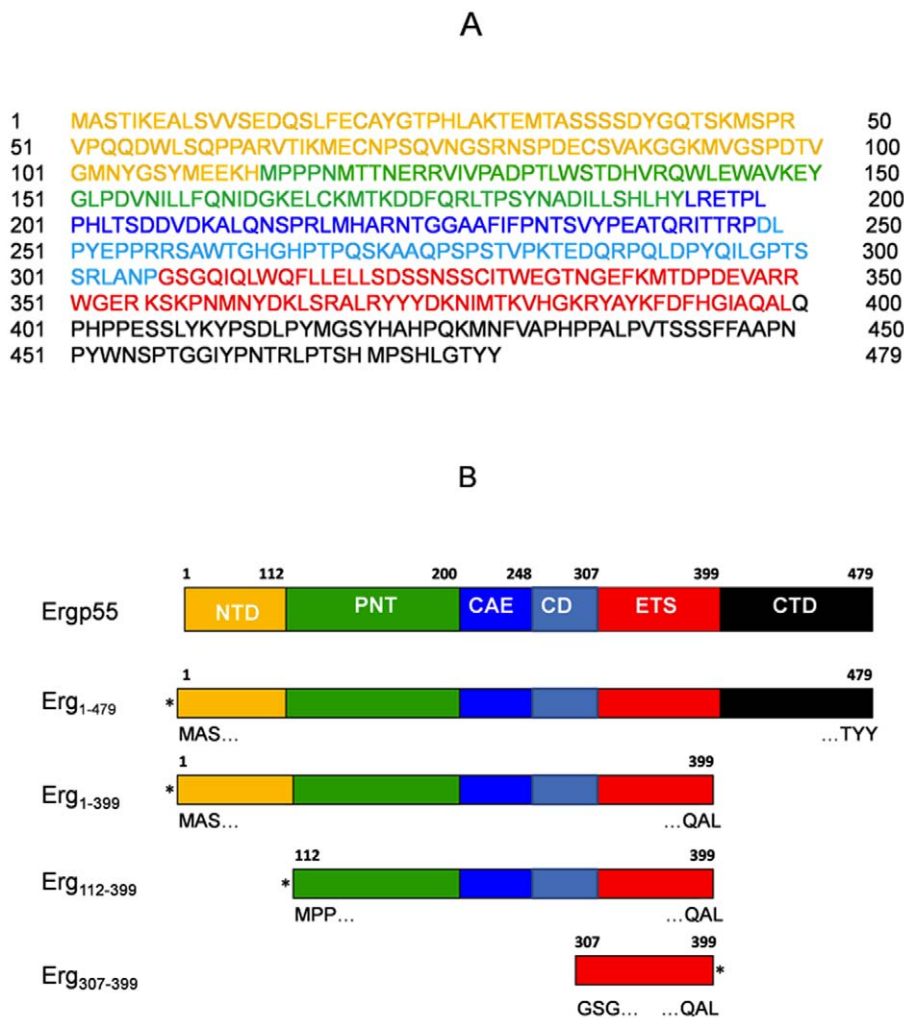


Figure 1. The human Ergp55 protein sequence. (A) Shown here the amino acid sequences full length Ergp55 without 6xHis tag and cleavage site. The residues were highlighted according to the size of Ergp55 domains, NTD (yellow), PNT (green), CAE/CD (blue/cyan), Ets (red) and CTD (black). (B) Shown here the Ergp55 constructs used in the experiments (coloring scheme same as in Fig. 1A). The labeled residues defining the sequence of beginning and end of constructs. Asterisks denote the position of 6xHis tag on various Ergp55 constructs.
doi:10.1371/journal.pone.0039850.g001

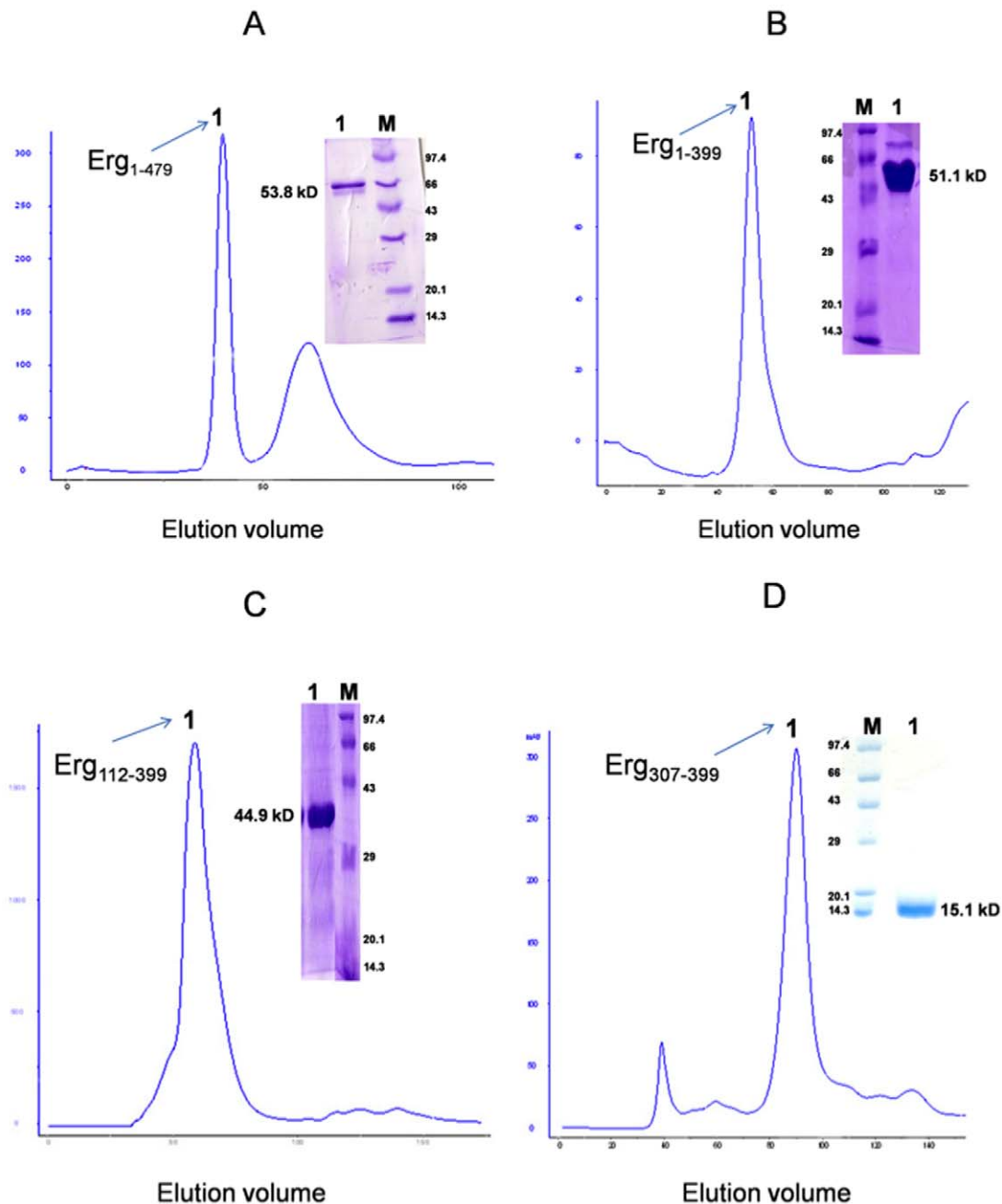


Figure 2. Size exclusion chromatography of Ergp55 polypeptides. The chromatogram of (A) full length Erg₁₋₄₇₉ (B) Erg₁₋₃₉₉ (C) Erg₁₁₂₋₃₉₉ and (D) Erg₃₀₇₋₃₉₉ polypeptides are shown. The SDS-PAGE analysis and calculated molecular weight of each protein are denoted on each chromatogram. doi:10.1371/journal.pone.0039850.g002

concentration was determined using absorbance at 280 nm. Coomassie brilliant blue stained SDS-PAGE analysis indicated that all Ergp55 polypeptides were purified greater than 95% purity. All proteins were stored at -20°C .

Surface plasmon resonance experiment

Biosensor studies were performed using BIAcore 2000 (Biacore Pharmacia Biosensor AB, Uppsala Sweden) equipment [18]. All experiments were performed at 25°C in HBS-buffer containing [10 mM HEPES buffer pH 7.4, 150 mM NaCl, 50 mM EDTA, 0.005% P20 as surfactant]. Since all Ergp55 polypeptides contain 6xHis tag either on N- or C-terminal, the sensor chip containing

high density of immobilized Ni-NTA is an ideal tag to immobilize these proteins. Initially, the flow cells of the chip were activated by passing nickel chloride solution over it. 40 μl of each Ergp55 polypeptide *e.g.*, Erg₁₋₄₇₉ (0.14 $\mu\text{g}/\mu\text{l}$), Erg₁₋₃₉₉ (0.19 $\mu\text{g}/\mu\text{l}$), Erg₁₁₂₋₃₉₉ (0.92 $\mu\text{g}/\mu\text{l}$) and Erg₃₀₇₋₃₉₉ (0.17 $\mu\text{g}/\mu\text{l}$) were injected in different flow cells of Ni-NTA chip at flow rate of 10 $\mu\text{l}/\text{min}$.

In different flow cells of Ni-NTA chip, following RU unit of each Ergp55 polypeptide was immobilized *e.g.*, Erg₁₋₄₇₉ (2904 RU), Erg₁₋₃₉₉ (2848 RU), Erg₁₁₂₋₃₉₉ (2825 RU) and Erg₃₀₇₋₃₉₉ (247 RU), where 1 RU corresponds to immobilized protein of concentration ~ 1 $\mu\text{g}/\text{mm}^2$. Binding experiments were performed with three different concentration [1, 2, 3 μM] of DNA sequence

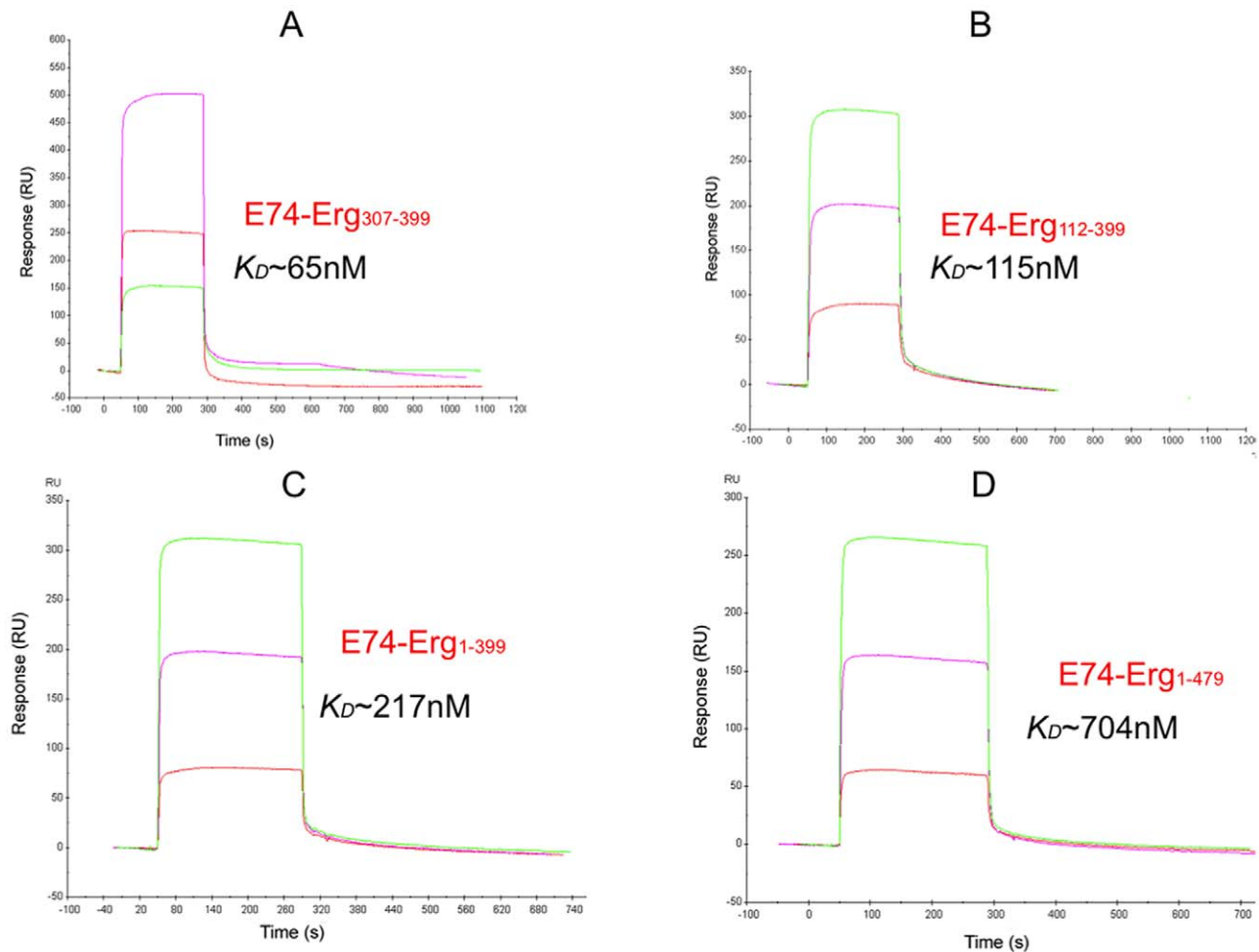


Figure 3. Binding analysis of DNA sequence of E74 promoter to immobilized Ergp55 polypeptides. In figures (A) Erg₁₋₄₇₉ (B) Erg₁₋₃₉₉ (C) Erg₁₁₂₋₃₉₉ and (D) Erg₃₀₇₋₃₉₉ polypeptides were immobilized on Ni-NTA chip. Three concentration (1, 2, 3 μ M) of DNA sequence of E74 promoter injected on each immobilized Ergp55 polypeptide. doi:10.1371/journal.pone.0039850.g003

of E74 promoter (5' TACCGGAAGT 3') in HBS-buffer and injected over immobilized Ergp55 polypeptides at flow rate of 10 μ l/min. Similar experiment was performed using DNA sequence of cfos promoter sequence (5' GACAGGATGTG 3'). The sensogram allowed to run for another 4 min. The regeneration of biosensor surface was done using 30 s pulse of 1 mM NaOH at flow rate of 10 μ l/min. Associate and dissociation kinetic constants were calculated by BIAevaluation 3.0 software using simple 1:1 Langmuir model with the assumption, that density of Ergp55 polypeptides on the sensor chip were not high enough to support bivalent DNA binding.

Circular dichorism

CD measurements were recorded using ChirascanTM CD spectropolarimeter (Applied Photophysics) with a water bath to maintain the constant temperature. The Ergp55 polypeptides were diluted to 0.2 mg/ml in 10 mM sodium phosphate buffer, pH 8.0 and loaded on 0.1 cm quartz cuvette. The blank of all experiments was 10 mM sodium phosphate buffer, pH 8.0. The final spectrum was an average of three sequential scan.

For thermal denaturation study of full length Ergp55, the CD spectra were recorded at 10°C increment starting from 10°C to 90°C. Before measurement, the sample cuvette was equilibrated at

each temperature. Temperature readings were taken within cuvette holder agreed with temperature of water bath. All CD data were converted to mean residue ellipticity (deg. cm²/dmol). The Dichroweb server [19] was used to estimate the amount of secondary structure in Ergp55 polypeptides from CD spectra.

Secondary structure prediction

The SOPMA [20], GOR [21] and PSIPRED [22] algorithms were used to predict the secondary structure contents in full length Ergp55, which showed ~21% α -helix, 12–15% β -sheet and 65–69% random coil structures.

Modeling of Ergp55 polypeptides

Phyre server [23] was used to obtain the structure of full length Ergp55 protein (1–479 residues). The server yielded the structure of PNT domain (95–221 residues) of Ergp55 using template PDB-2YTU (solution structure of SAM-PNT domain of human friend leukemia integration factor-1 transcription factor, not yet published). The NMR solution structure of PNT domain (108–201 residues) was obtained from protein data bank (PDB-1SVO) [24]. The Phyre server also yielded the structure of Ets domain (284–412 residues) of Ergp55 using the template PDB-2NNY (Regulation of transcription factor Ets-1 by DNA mediated

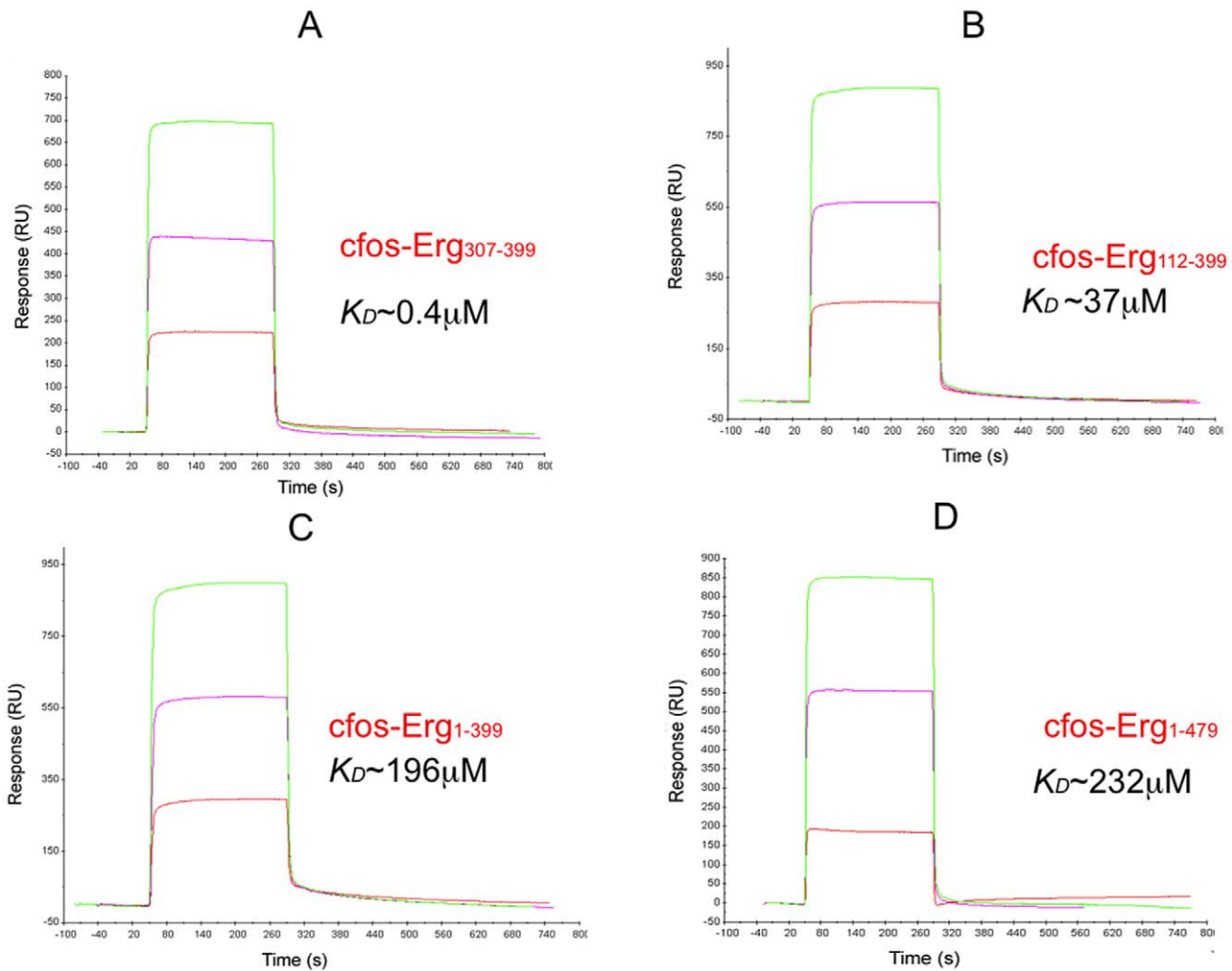


Figure 4. Binding analysis of DNA sequence of cfos promoter to immobilized Ergp55 polypeptides. In figures (A) Erg₁₋₄₇₉ (B) Erg₁₁₂₋₃₉₉ (C) Erg₁₁₂₋₃₉₉ and (D) Erg₃₀₇₋₃₉₉ polypeptides are immobilized on Ni-NTA chip. Three different concentration (1, 2, 3 μ M) of DNA sequence of cfos promoter injected on each immobilized Ergp55 polypeptide. doi:10.1371/journal.pone.0039850.g004

homodimerization) [13]. The NMR structure of Ets domain of Fli1 (306–403 residues) was obtained from protein data bank (PDB-1FliA) [25]. The structural modeling of N- terminal (1–94 residues), central domain (222–283 residues) and C- terminal (413–479 residues) of Ergp55 were not attempted due to lack of input model.

The Modeler [26] and LOMETS threading [27] programs were used to build the structure of full length Ergp55 using following inputs (i) structural model of PNT domain (95–221 residues) of Ergp55 (ii) structural model of Ets domain (284–412 residues) of Ergp55 (iii) NMR structure of PNT domain (108–201 residues) of Ergp55 and (iv) NMR structure of Ets domain (306–403 residues) of Fli1. Energy minimization was performed on modeled Ergp55 using Gromacs program version 4.0.5 [28]. 100 steps of steepest descent and 500 steps of conjugated gradient algorithms were used in energy minimization calculation.

Molecular dynamics simulations

The 10 ns molecular dynamics (MD) simulation was performed on minimized Ergp55 model using GROMACS program (version 4.0.5) with Gromacs43a2 force field [28]. The Ergp55 model was immersed in a cubic box extending 0.5 nm from the protein

surface and solvated with explicit SPC water molecules. Chloride and sodium ions were added to neutralize the systems, which were then simulated with periodic boundary conditions. The solvated Ergp55 model consists of 4832 protein atoms surrounded by ~390,000 water molecules. Before running the simulation, whole system was energy minimized for 200 iterations of steepest descents and then equilibrated for 20 ps keeping protein atoms restrained. All restraints were removed from the protein and temperature was gradually increased in 10 distinct steps of 5 ps simulations each.

Berendsen coupling was employed to maintain a constant temperature of 300 K with a coupling constant τ of 0.1 ps. Van der Waals interactions were modeled using 6–12 Lennard-Jones potentials with 1 nm cutoff. The coulomb cut off was 1.0. The time step employed was 2 fs and coordinates were saved every 5 ps for analysis of MD trajectories.

The stereochemistry of simulated Ergp55 model was checked by PROCHECK program of CCP4 suite [29]. Secondary structure composition was measured by DSSP program [30] and structure visualization by PyMOL program [31].

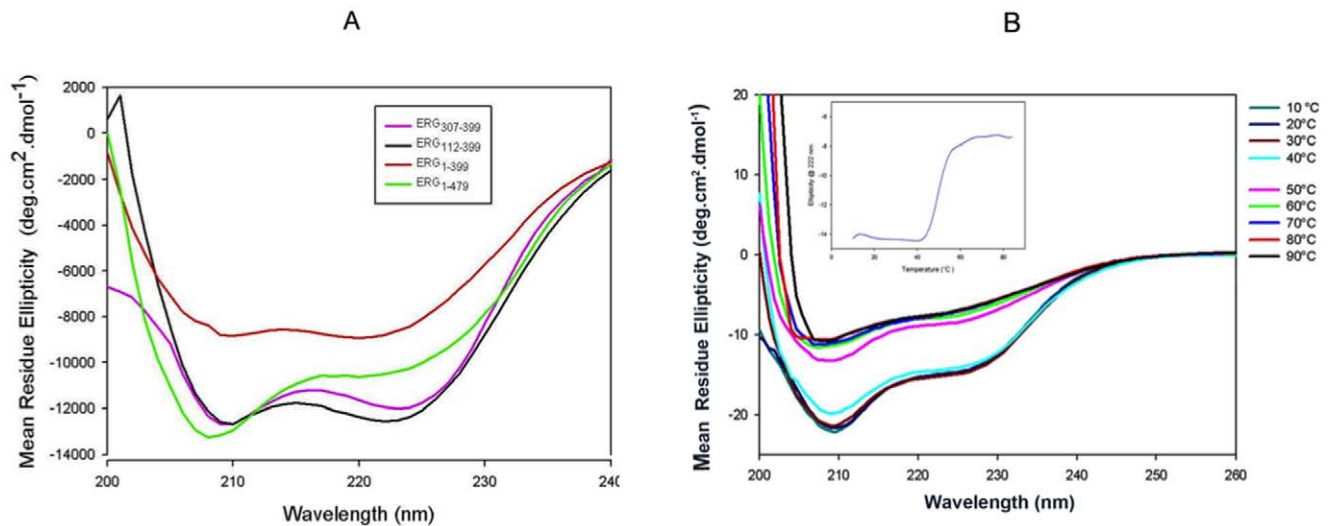


Figure 5. CD spectra of Ergp55 polypeptides. (A) CD spectra of Erg₁₋₄₇₉, Erg₁₋₃₉₉, Erg₁₁₂₋₃₉₉ and Erg₃₀₇₋₃₉₉ polypeptides recorded from 200 nm to 260 nm. (B) Temperature induced unfolding of full length Ergp55 protein. The plot (inner) containing mean residue ellipticity versus temperature indicates the denaturation of α -helices of folded full length Ergp55. doi:10.1371/journal.pone.0039850.g005

Results

Purification of recombinant Ergp55 polypeptides

We produced full length and smaller polypeptides (containing subset of predicted domains) of Ergp55 in *E. coli* and purified using standard chromatographic techniques (Fig. 1A–B). During size exclusion chromatography, the purified Erg₁₋₄₇₉, Erg₁₋₃₉₉, Erg₁₁₂₋₃₉₉ and Erg₃₀₇₋₃₉₉ polypeptides eluted at volumes corresponding to their molecular weights (Fig. 2A–D). The full length Erg₁₋₄₇₉ eluted at the size of 53.8 kD, Erg₁₋₃₉₉ polypeptide eluted at 51.1 kD, Erg₁₁₂₋₃₉₉ eluted at 44.9 kD and Erg₁₁₂₋₃₉₉ eluted at 15.1 kD. The Erg₁₋₄₇₉ and Erg₁₋₃₉₉ polypeptides have tendency to degrade if kept for 7 days at 4°C. The Erg₁₁₂₋₃₉₉ and Erg₃₀₇₋₃₉₉ polypeptides do not degrade over time and more stable than Erg₁₋₄₇₉ and Erg₁₋₃₉₉ polypeptides.

DNA binding studies using E74 and cfos promoter sequences

The functionality of purified Ergp55 polypeptides was assessed by DNA binding experiment using surface plasmon resonance technique. The observed K_D value of Ergp55 polypeptides with DNA sequence of E74 promoter were *e.g.*, Erg₁₋₄₇₉ ~ 704 nM, Erg₁₋₃₉₉ ~ 217 nM, Erg₁₁₂₋₃₉₉ ~ 115 nM and Erg₃₀₇₋₃₉₉ ~ 65 nM (Fig. 3A–D). These results indicated that Ets domain of Ergp55 (Erg₃₀₇₋₃₉₉) has the highest affinity for E74 promoter DNA sequence. The DNA binding affinity decreases ~ 2 fold for Erg₁₁₂₋₃₉₉ polypeptide, ~3 fold for Erg₁₋₃₉₉ polypeptide and ~10 fold for full-length Erg₁₋₄₇₉, when compared with Erg₃₀₇₋₃₉₉ polypeptide (Ets domain). These results indicate that N- as well as C- terminal domains with respect to Ets domain are involved in auto-inhibition of DNA binding to Ergp55 protein.

The observed K_D value of Ergp55 polypeptides with cfos promoter sequence were *e.g.*, Erg₁₋₄₇₉ ~ 232 μ M, Erg₁₋₃₉₉ ~ 196 μ M, Erg₁₁₂₋₃₉₉ ~ 38 μ M and Erg₃₀₇₋₃₉₉ ~ 0.45 μ M (Fig. 4A–D). These results also indicate that Ets domain has the highest affinity for DNA sequences of cfos promoter. Both N- and C-terminal domains with respect to Ets domain are involved in inhibition of cfos DNA binding to Ergp55 protein.

CD measurements of Ergp55 polypeptides

To identify the secondary structure contents in Ergp55 polypeptides, the far-UV CD spectroscopy was used (Fig. 5A). The CD data were de-convoluted using DICHROWEB web server [19] and percentage of α -helix, β -sheet and random coil structures were estimated. The CD spectra of full-length Erg₁₋₄₇₉ (Fig. 5A) showed two minima around 208 nm and 222 nm, a characteristic of α -helical structure. Deconvolution of data predicts ~35% α -helix, 15% β -sheet and 49% random coil structures in full length Ergp55. The CD data of Erg₁₋₃₉₉ polypeptide predicts ~25% α -helix, 17% β -sheet and 57% random coil structures, which shows less α -helix and β -sheet structure compare to full length Erg₁₋₄₇₉ structure.

In case of Erg₁₁₂₋₃₉₉ polypeptide, the CD data estimates ~29% α -helix, 15% β -sheet and 55% random coil structures. This polypeptide contains less α -helix, similar β -sheet structure compared to full length Erg₁₋₄₇₉. For Erg₃₀₇₋₃₉₉ polypeptide, the CD data estimates ~31% α -helix, 10% β -sheet and 59% random coil structures, which has less α -helix and β -sheet structure compare to full length Erg₁₋₄₇₉ structure. However, these values are close to secondary structure contents in crystal structure of Ets domain of Fli-1 (35.7% α -helix, 4.1% β -sheet, 60.2% random coil). The Ets domain of Fli-1 is the closest homologous to Ets domain of Ergp55 [32].

Thermostability of full length Ergp55

To assess the thermostability of full length Ergp55, a far UV-CD spectrum of protein was measured from 10 to 90°C (Fig. 5B). It is clear from spectra that secondary structure of Ergp55 denatures as temperature increased. When mean residue ellipticity at 222 nm is plotted against temperature, the inflection point of sigmoidal curve indicates the T_m of $45 \pm 2^\circ\text{C}$ of full length Ergp55.

Molecular modeling and dynamic simulation of full length Ergp55

The secondary structure prediction on full length Ergp55 using PSIPRED program is shown in (Fig. 6A). The Modeler and automatic threading LOMETS programs were used to construct full length Ergp55 model using (i) structure of 95–221 residues of

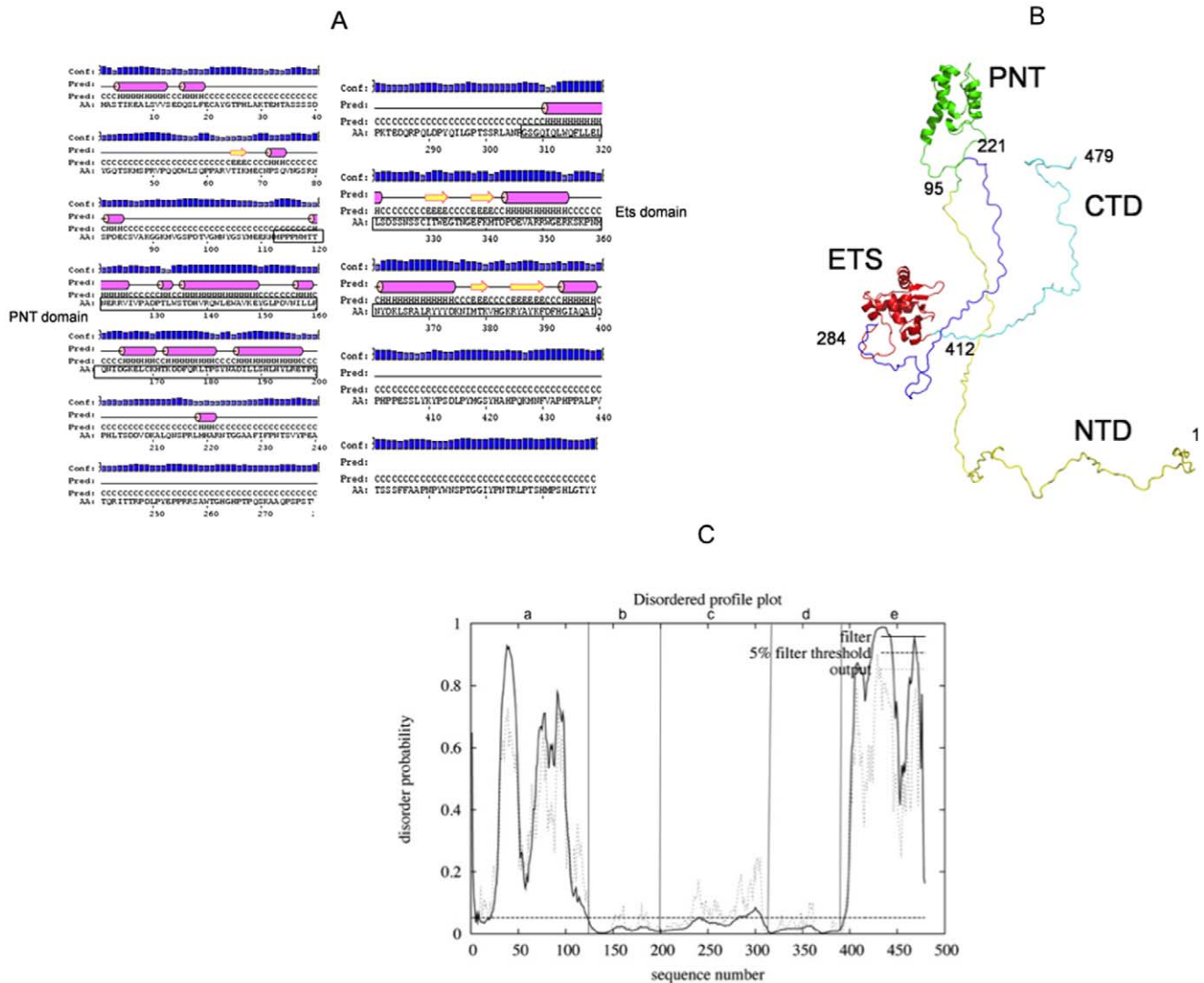


Figure 6. Model of Ergp55 polypeptides. (A) PSIPRED program analysis of full length Ergp55. (B) Structural model of full length Ergp55 after molecular modeling and dynamics simulation analysis using GROMACS program (C) The disordered prediction made by DISOPRED for (a), NTD (b), PNT (c), CAE/CD (d), ETS and (e), CTD regions of Ergp55. doi:10.1371/journal.pone.0039850.g006

Ergp55 using the template PDB-2YTU (not published) having 57% sequence identity (ii) the structure of 284–412 residues of Ergp55 using template PDB-2NNY [13] having 40% sequence identity and (iii) NMR structure of 108–201 residues of Ergp55 and (iv) NMR structure of 306–403 residues of Ets domain of Fli-1 having 90% sequence identity. Energy minimization and dynamics simulations analysis were performed on constructed Ergp55 model, which yielded a flexible and elongated structure (Fig. 6B). The Ergp55 structure remained very stable during whole simulation time, as confirmed by all the indicators commonly used to analyze MD simulation.

The 93% residues of Ergp55 model lie in most favored region of Ramachandran plot and a Prosa Z-score of -4.87 . DISOPRED [33] analysis on Ergp55 model indicated that N-terminal (1–118 residues) and C-terminal (397–479 residues) are largely disordered (except N-terminal 4–23 residues) and remaining Ergp55 structure (119–396 residues) were ordered (Fig. 6C). The structured PNT domain contains tertiary arrangement of four α -helices, characteristic of large group of SAM domain [34]. The Ets domain

consists of four α -helices and four β -sheets, a characteristic of Ets family proteins. In N-terminal domain, 15 residue stretch predicted to form α -helix and 3 residue long helix (219–221 residues) are observed in in CAE/CD domain of Ergp55. The stretches of residues in C-terminal domain of Ergp55 predicted to have only random coil structure.

The N-terminal and C-terminal domain of Ergp55 are positioned away from Ets domain. The DNA binding groove of Ets domain is exposed to solvent and free to bind promoter DNA sequences. The CAE/CD domain is positioned between PNT and Ets domain to regulate the activity of Ergp55. The N-terminal and C-terminal domains of Ergp55 are positioned in a region that do not prevent the DNA binding activity of Ets domain and play a role in transcriptional activation and localization of Ergp55.

Discussion

In current study, we have expressed full length and smaller polypeptides of Ergp55 in *E. coli*. The combinations of two chromatography steps (Ni-NTA affinity and size exclusion

chromatography) have yielded more than 95% pure Ergp55 polypeptides based on mass spectrometry and SDS-PAGE analysis. Prior to structural studies, the activity of purified Ergp55 polypeptides were checked by binding studies using DNA sequences of E74 and cfos promoters. The surface plasmon resonance technique was used for binding analysis. These results indicated that Ergp55 polypeptides produced in *E. coli* were in good conformation and bind specifically DNA sequences of E74 and cFos promoters with different affinities.

DNA binding autoinhibition of Ergp55. In case of E74 promoter sequence, following K_D values were observed for Ergp55 polypeptides (i) Erg_{307–399}, 65 nM (ii) Erg_{112–399}, 115 nM (iii) Erg_{1–399}, 217 nM and (iv) full length Erg_{1–479}, 704 nM. Comparison of (i) and (ii) indicated that N-terminal region (PNT+CAE/CD domains) preceding to Ets domain inhibit the E74 DNA binding to Ets domain. Comparison of (ii) and (iii) showed the evidence of increased DNA binding inhibition by having NTD domain in Erg_{1–399} polypeptide. Comparison of (iii) and (iv) indicate that adding CTD domain in Erg_{1–399} polypeptide showed enhanced inhibition in DNA binding to Ets domain. These results indicate that E74 DNA binding to Ergp55 is negatively influenced by CAE/CD, PNT and NTD domains located at N-terminal and CTD domain located C-terminal region in Ergp55.

With cfos promoter DNA sequence, following K_D values were obtained for Ergp55 polypeptides (i) Erg_{307–399}, 0.4 μ M (ii) Erg_{112–399}, 37 μ M (iii) Erg_{1–399}, 196 μ M and (iv) full length Erg_{1–479}, 232 μ M. These results indicate that cfos promoter sequence bind with different affinity to Ergp55 polypeptides than E74 promoter sequence, however mechanism of DNA binding inhibition was similar as observed in case of E74 promoter DNA sequence.

A cooperatively acting DNA inhibiting region (468–510 residues) was identified at C-terminal of Ets-1 transcription factor [34]. In case of ERM and PEA3 transcription factors, two main domains located at N- and C-terminal with respect to their ETS-domain inhibiting DNA binding affinity [35–38]. One domain corresponds to residue 280–360 residues of ERM transcription factor, involved in inhibition of ERM DNA binding capacity. These domains are rich in proline residues generally devoid α -helical structures. The mechanism by which two domains cooperatively inhibit ERM DNA binding is different than observed in case of Ets-1 transcription factor. The DNA binding activity of Ets domain is dependent on autoinhibitory module [39]. The binding affinity of Ets domain Ergp55 to E74 and cfos promoter DNA was consistent to the observation obtained in case of above transcription factors.

Circular dichorism analysis of Ergp55 polypeptides. The circular dichorism technique was used to

identify the secondary and tertiary structures of Ergp55 polypeptides. The full length and smaller Ergp55 polypeptides contain high α -helical and random coil structures. The CD data estimates 35% α -helix and 49% random coil structures in full length Ergp55 protein. Examinations of thermal stability and temperature effect on full length Ergp55 protein indicated that protein underwent an alteration of secondary structure upon heating. The secondary structure is regained after cooling the protein from 80°C to 20°C. Short change of temperature is unlikely to have any effect on secondary structure of Ergp55 protein.

Modeling and dynamics simulation of full length Ergp55. The molecular modeling and dynamics simulation analysis indicated that full length Ergp55 acquires a flexible and highly elongated structure. Only PNT and Ets domains are structured in protein and long flexible regions are observed at N- and C- terminus of Ergp55. The structure of PNT domain of Ergp55 consists of four-helix bundle Sam like structure [34]. The Ets domain of Ergp55 is structured in a winged helix-turn-helix with scheme $\alpha_1\beta_1\beta_2\alpha_2\alpha_3\beta_3\beta_4\alpha_4$ [40–41]. The central CAE/CD domains contain one small helix at position 220. The secondary and disordered prediction analysis on Ergp55 also supported the finding observed in modeling and dynamics simulation studies of Ergp55. All these observations supported the flexible, non-globularity and highly elongated structure of Ergp55 protein.

In conclusion, we have characterized the recombinant full length and smaller polypeptides of Ergp55 produced in *E. coli*. The Ergp55 polypeptides were purified greater than 95% purity as determined by mass spectrometry and SDS-PAGE analysis. The structural data presented here showed the evidence of flexible and highly elongated structure of full length Ergp55 protein. The binding analysis using DNA sequences of E74 and cfos promoters indicate that longer fragments of Ergp55 (beyond the canonical Ets domain) showed the evidence of autoinhibition.

Acknowledgments

The authors thank Sita R. Meena and Pratibha for their suggestions in current project. We acknowledge the help from advanced instrumentation research facility (AIRF) and central instrumentation facility (CIF) of Jawaharlal Nehru University for allowing us to conduct the CD experiment. The authors thank SPR equipment facility of AIIMS, Delhi for allowing us to conduct the DNA binding experiments.

Author Contributions

Conceived and designed the experiments: AKS. Performed the experiments: SPG SD AKS. Analyzed the data: AKS. Contributed reagents/materials/analysis tools: SD AKS. Wrote the paper: AKS.

References

1. Sharrocks AD (2001) The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol* 2: 827–837.
2. Shimizu K, Ichikawa H, Tojo A, Kaneko Y, Maseki N, et al. (1993) An ets-related gene, ERG, is rearranged in human myeloid leukemia with t(16; 21) chromosomal translocation. *Proc Natl Acad Sci USA* 90: 10280–10284.
3. Sorensen PH, Leessnick SL, Terrada DL, Liu XF, Triche, et al. (1994) A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor, ERG. *Nat Genet* 6: 146–151.
4. Loughran ST, Kruse EA, Hacking DF, de Graaf CA, Hyland CD, et al. (2008) The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. *Nat Immunol* 9: 810–819.
5. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, et al. (2005) Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in Prostate Cancer. *Science* 310: 644–648.
6. Siddique HR, Rao VN, Lee L, Reddy ES (1993) Characterization of the DNA binding and transcriptional activation domains of the erg protein *Oncogene*. 8: 1751–1755.
7. Carrere S, Verger A, Flourens A, Sthelin D, Duterque-Coquillaud M (1998) Erg proteins, transcription factors of the Ets family, form homo, heterodimers and ternary complexes via two distinct domain. *Oncogene* 16: 3261–3268.
8. Verger A, Buisine E, Carrere S, Winjens R, Flourens A, et al. (2001) Identification of amino acid residues in the Ets transcription factor Erg that mediate Erg-Jun/Fos DNA ternary complex formation. *J Biol Chem* 276: 17181–17189.
9. Agarkar VB, Babayeva ND, Wilder PJ, Rizzino A, Tahirov TH (2010) Crystal structure of mouse Ets3 C-terminal DNA-binding domain in complex with type II TGF-beta receptor promoter DNA. *J Mol Biol* 397: 278–289.
10. Batchelor AH, Piper DE, de la Brousse FC, McKnight SL, Wolberger C (1998) The structure of GABP alpha/beta: an ETS domain-ankyrin repeat heterodimer bound to DNA. *Science* 279:1037–1041.
11. Garvie CW, Hagman J, Wolberger C (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell* 8: 1267–1276.
12. Kodandapani R, Pio F, Ni CZ, Piccialli G, Klemsz M, et al. (1996) A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature* 380: 456–460.

13. Lamber EP, Vanhille L, Textor LC, Kachalova GS, Sieweke MH (2008) Regulation of the transcription factor Ets-1 by DNA-mediated homodimerization. *EMBO J* 27: 2006–2017.
14. Mo Y, Vaessen B, Johnston K, Marmorstein R (1998) Structures of SAP-1 bound to DNA targets from the E74 and c-fos promoters: insights into DNA sequence discrimination by Ets proteins. *Mol Cell* 2: 201–212.
15. Pufall MA, Lee GM, Nelson ML, Kang HS, Velyvis A, et al. (2005) Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* 309: 142–145.
16. Wang Y, Feng L, Said M, Balderman S, Fayazi Z, et al. (2005) Analysis of the 2.0 Å crystal structure of the protein-DNA complex of the human PDEF Ets domain bound to the prostate specific antigen regulatory site. *Biochemistry* 44: 7095–7106.
17. Wei GH, Badis G, Berger MF, Kivioja T, Palin K, et al. (2010) Genome-wide analysis of Ets-family DNA-binding in vitro and in vivo. *EMBO J* 29: 2147–2160.
18. Nieba I, Nieba-Axmann SE, Person A, Hamalainen M, Edebratt F, et al. (1997) BIACORE Analysis of Histidine-Tagged Proteins Using a Chelating NTA Sensor Chip. *Anal Biochem* 252: 217–228.
19. Whitmore L, Wallace B A (2004) DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucl Acid Res* 32: W668–673.
20. Geourjon C, Deleage G (1995) SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Bios* 11: 681–684.
21. Garnier J, Gibrat JF, Robson B (1996) GOR secondary structure prediction method version IV. *Meth Enzymol* 266: 540–553.
22. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292: 195–202.
23. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* 4: 363–371.
24. Gai D, Zhao R, Li D, Finkielstein CV, Chen XS (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen Cell. Cambridge Mass 119: 47–60.
25. Liang H, Mao X, Olejniczak ET, Nettesheim DG, Yu L (1994) Solution structure of the ets domain of Fli-1 when bound to DNA. *Nat Struct Biol* 1: 871–875.
26. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Science* 9: 1753–1773.
27. Wu S, Zhang Y (2007) LOMETS: LOMETS: A local meta-threading-server for protein structure prediction. *Nucl Acid Res* 35: 3375–3382.
28. Lindahl E, Hess B, Spoel DVD (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J Mol Mod* 7: 306–317.
29. Collaborative Computational Project, Number 4 (1994). *Acta Cryst D50*: 760–763.
30. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577–2637.
31. DeLano WL (2002) The PyMOL Molecular Graphics System DeLano Scientific. San Carlos CA USA.
32. Prasad DD, Rao VN, Reddy ES (1992) Structure, and expression of human Fli-1 gene. *Cancer Res* 52: 5833–5837.
33. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF and Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology* 337: 635–645.
34. Mackereth CD, Scharpf M, Gentile LN, MacIntosh SE, Slupsky CM, et al. (2004) Diversity in Structure and Function of the Ets Family PNT Domains. *J Mol Biol* 342: 1249–1264.
35. Laget MP, Defossez PA, Albagli O, Baert JL, Dewitte F (1996) Two functionally distinct domains responsible for transactivation by the Ets family member ERM. *Oncogene* 12: 1325–1336.
36. Greenall A, Willingham N, Cheung E, Boam DS, Sharrocks AD (2001) DNA binding by the ETS-domain transcription factor PEA3 is regulated by intramolecular protein-protein interactions *J Biol Chem* 276: 16207–16215.
37. Bojovic BB, Hassel JA. (2001) The PEA3 Ets transcription factor comprises multiple domains that regulate transactivation and DNA binding. *J Biol Chem* 276: 4509–45021.
38. Mauen S, Huvent I, Raussens V, Demonte D, Baert JL (2006) Expression, purification, and structural prediction of the Ets transcription factor ERM. *Biochem Biophys Acta* 1760: 1192–1201.
39. Brown LA, Mores A, Schilling TF, Jowett T, Baert JA, et al. (1998) *Oncogene* 17: 93–104.
40. Mo Y, Vaessen B, Johnston K, Marmorstein R (2000) Structure of the Elk-1–1 DNA complex reveals how DNA distal residues affect ETS domain recognition of DNA. *Nat Struct Biol* 7: 292–297.
41. Pio F, Kodandapani R, Ni CZ, Shepard W, Klemsz M, et al. (1996) New Insights on DNA Recognition by ets Proteins from the Crystal Structure of the PU.1 ETS Domain-DNA Complex. *J Biol Chem* 271: 23329–23337.