RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Structural transition of ETS1 from auto-inhibited to functional state upon association with p16INK4a native and mutated promoter region

Kannan Muthu, Manivel Panneerselvam, Nishith Saurav Topno and Krishna Ramadas*

Centre for Bioinformatics, Pondicherry University, Puducherry, India-605014

*Corresponding Author

Dr. R. Krishna,

Assistant Professor,

Centre for Bioinformatics,

Pondicherry University

Puducherry-605014

Abstract:

Detailed elucidation of structural changes invoked on transcriptional factors and their target genes upon their association is pivotal for understanding the genetic level regulations imposed in several diseases including ovarian cancer. Present study reveals the structural dynamics of ETS1, a member of E26 transcription factor family, exhibited upon binding with wild and mutant forms of p16^{INK4a}. The results highlight the presence of specific type of interactions that controls the transition of ETS domain from its auto-inhibited triangle shaped to relaxed (oval) conformation. Precisely, the interactions between residues of triangle basic patch of HI2-H1 loop, triplet residues at H3 helix of ETS domain with major groove of p16^{INK4a} displaces inhibitory helices HI2 from H4, and transforms ETS1 from auto-inhibited to active form. Further, these interactions induce HI1 helix folding and induce mode I and II distortion in p16^{INK4a} promoter region. But, mutations in the E-box motif of p16^{INK4a} promoter region and at the region preceding E-box motif distorts DNA mode, changes interacting pattern with ETS1 and renders ETS domain in a partial and exact triangle shape. Results emphasize that the mutations in promoter region impose structural changes in ETS domain and subsequently affects the expression of tumor suppressor genes in ovarian cancer.

Key words:

ETS1, p16^{INK4a}, structural transition, molecular docking and simulation, DNA distortion, ovarian cancer

Introduction

p16^{INK4a}, a well-studied tumor-suppressor protein accomplishes anti-tumor activity by arresting tumor cells in G1 phase of cell cycle through inhibition of the cell-cycle dependent kinases (CDK4 and CDK6) [1, 2]. Controlled regulation of p16^{INK4a}, thus remains instrumental in deciding the fate of tumor cells. The expression of p16^{INK4a} is largely affected by different mechanisms such as mutations in the DNA, promoter methylation, or by virtual deletion in all tumor cells [3-6]. The transcription of p16^{INK4a} is described to be down-regulated by its binding partners such as the transcription factor ETS1. Interaction of ID1 with ETS1 and ETS2 results in the suppression of p16^{INK4a} and is responsible for inducing abnormal cell growth (tumorigenesis) and metastasis of ovarian cancer cells[7-9].

ETS1 and ETS2 are members of E26 transcription factor family which binds with the promoter region of p16^{INK4a} by recognizing the ETS-binding site (EBS) that encompasses a 5'-GGAA/T-3' core motif [7]. ETS1 transcription factor plays a key role in embryonic development, angiogenesis, proliferation, and apoptosis, and is up-regulated and rearranged in various cancers[10]. ETS1 has an ETS domain flanked by two inhibitory regions at N-terminal (constituted by helices HI1 and HI2) and C-terminal (H4 and H5) and an inhibitory serine rich region. Specific orientation and interaction of inhibitory region and serine rich region with ETS domain affects the DNA binding abilities of ETS transcription factor[11]. Initially, the ETS domain remains in an auto-inhibited form prior to its binding with the promoter region and is characterized by a triangular shaped ETS domain[12, 13]. The triangular shaped ETS domain is effected by the intramolecular contacts between the inhibitory helices HI2 (N-terminal region) and H4 (C-terminal region). The parallel arrangement of helices HI1 and HI2 and helical nature of HI1 also acts as the key structural features characterizing the auto-inhibited state of ETS domain. Upon binding with promoter region, the intramolecular contacts are disturbed, thereby relaxing the ETS domain to adopt an oval shape [14-17]. Concomitantly, HI1 also moves away from HI2 and loses its helical nature due to disruption of hydrophobic interaction formulated by the helical orientation. The co regulators of ETS1 and the nucleotide sequence of ETS target gene regulate the structural transition in ETS domain of ETS transcription factor. Clear picture of these interactions are pivotal for understanding ETS1 mediated gene regulation, particularly in conditions such as in ovarian cancer, where although ETS1 is up-regulated, the expression of its target gene p16^{INK4a} is not

observed, which remains as one of the key factor contributing to the oncogenic transformations in ovarian cells. The presence of mutations at the promoter region near the ETS binding site is postulated to be responsible for altering the binding abilities between ETS1 and p16[INK4a] gene[7]. In general, the folding pattern of ETS domain differs in response to different ETS target genes[18]. Hence, the detailed picture explaining how it binds p16 gene and how the presence of mutations in p16[INK4a] affects this binding can sufficiently explain why p16[INK4a] tumor suppressor remains un- expressed in ovarian cancer.

In the present work, we explored the binding and associated structural events that mediate the transition of ETS1 binding with the native and mutated forms of p16[INK4a]. Also, the significance of several intra and inter molecular associations that govern the transition of ETS1 from auto-inhibited to functional form is being reported through molecular dynamics approach.

Materials and methods

The p16[INK4a] promoter region

The three dimensional structure of p16[INK4a] promoter region (5'-CACCGG**AGGAAG**AAAGA-3') containing E-box motif (5'-GGAA-3'), was built as a B-type DNA-duplex using the *build and edit nucleic acid* module of Discovery studio. In order to understand the effect of mutations on p16[INK4a] promoter-ETS1 interaction, three mutant forms of p16[INK4a] promoter region, M1 (5'-CACC*CC***AGGAAG**AAAGA-3'), M2 (5'-CACCGG***ATT*A**AG**AAAGA-3') and M3 (5'-CACC*CC***A*TT*A**AG**AAAGA-3')[7] were built as like wild type form. All these DNA constructs were minimized using smart minimizer protocol of Discovery studio V3.1 for 1000 steps, which combines both the steepest descent and conjugate gradient methods. CHARMm force field was used for energy minimization steps and the minimized structures were used for further analysis.

The protein-DNA docking

The molecular interaction of ETS1 protein with the native and mutated forms (M1, M2 and M3) of p16[INK4a] promoter region, was predicted using High Ambiguity Driven Biomolecular Docking (HADDOCK v2.0) [19, 20] in combination with CNS V1.2. Crystallographic structure of ETS1 in auto inhibited dimeric form (PDB ID: 1GVJ) was retrieved prepared using the protein preparation Wizard of the Schrödinger 2010 suite (Schrödinger LLC, New York, USA). The

bHLH/ ETS DNA binding motif of ETS1 (331–415)[14] and E-Box (AGGAAG) motif [7] of p16[INK4a] promoter region were considered as active site residues and neighboring residues were identified as passive residues for docking. The docking protocol includes three imperative stages; the first involves complex formation and orientational optimization followed by semi-flexible docking and finally refinement of the complex in explicit solvent. The first stage generated 1000 docked complexes based on the rigid-body energy minimization. These structures were further scrutinized based on their intermolecular energy such as the electrostatic, Van der Waals (VdW) and also the AIR energy terms which narrowed down the count to 200 complexes. The chosen complexes were then subjected to semi-flexible simulated annealing (SA) within the torsional space. Finally, the lowest energy complexes were refined in an 8 Å explicit water model (TIP3) and the best complexes were obtained based on their root mean square deviation (RMSD) lying within a cut-off range of 7.5 Å.   A clear and detailed methodology for generating protein-protein and protein-DNA complexes has already been discussed in our previous studies [21-23].

Molecular dynamics simulation

The docked complexes of ETS1 with native and mutated p16[INK4a] promoter were subjected to molecular dynamic simulation for the stability and structural transition analysis using GROMACS 4.5.3 [24, 25] suite. Hydrogen atoms were added to the complexes and the topology was generated by assigning proper geometrical parameters according to Amber99SB-ILDN force field [26]. Then, the complexes were settled in cubic box type where the edge of the box from the molecule was set to 1.0 nm in all directions. SPC216 water model was used to solvate the box based on Periodic boundary conditions, total charge was neutralized and system was minimized by steepest descent algorithm up to a maximum of 50,000 steps and a convergence tolerance of 1000 kJ mol^{-1} nm^{-1}, following which, conjugate gradient algorithm was used with the same steps and convergence tolerance. Before the production molecular dynamics run, two different methods for position restrain: NVT (constant number of particles, volume and temperature) and NPT (constant number of particles, pressure and temperature) were used to equilibrate the system for 100 ps. In both the cases LINCS [27] holonomic constraints were used for bonded parameters and SETTLE [28] was used for constraining the water geometry. Particle Mesh Ewald (PME) [29] coulomb type was used for long-range electrostatics with a PME order of 4 and

maintaining the Fourier spacing by 0.16. In addition, the V-rescale temperature coupling was used to retain the temperature at 310 K for both protein and non-protein coupling groups. In NPT equilibration step, Parrinello-Rahman pressure [30] coupling was introduced which includes the isotropic coupling type to maintain a uniform scaling of box vectors with 1.0 bar as the reference pressure. Production MD run was carried out for 50 ns time scale for native and mutated $p16^{INK4a}$ promoter-ETS1 complexes using the above mentioned protocol. A total of 200 ns time scale of MD simulation was performed using Tesla Server with 2 x Intel Xeon Quad-Core processor running at 2.4 GHz on a CentOS Linux-based operating system. The parameters used for molecular simulation can also be referred from our previous work [31-33]. In order to explore the structural transition and conformational orientation of both bio-molecular complexes as an influence of each other, T-pad, PCA and FEL analyses were carried out for the 50 ns large simulation data of native and mutated $p16^{INK4a}$ promoter- ETS1 complexes.

Principal component analysis and Free-energy landscape

Principal component analysis (PCA) describes the high-amplitude concerted motion in a trajectory based on their eigenvectors of the mass-weighted covariance matrix of protein atomic fluctuations [34]. Here, the protein-DNA complexes of native and mutated $p16^{INK4a}$ promoter region were used to define the cosine content (c_i) of the principal component (p_i) of covariance matrix, which explains whether the time interval of MD simulation used to extract the sampling of trajectory is sufficient to represent the free energy landscape defined from PCA [35, 36]. It is an absolute and sensitive measure of trajectory convergence and has been tested to be efficient enough in generating the free energy landscape of the chosen principal components (PCs).

This cosine content value varies between 0 (no cosine) and 1 (perfect cosine) in the total simulation time (T):

$$C_i = \frac{2}{T} \left(\int \cos(i\pi t) p_i(t) dt \right)^2 \left(\int p_i^2(t) dt \right)^{-1}$$

Theoretically the first eigenvector's cosine contribution is the most efficient one to define a protein's characteristic nature in terms of its structural transitions. More frequently the first eigenvector is observed to have a cosine distribution closer to 1 which depicts the large scale motions in protein dynamics and hence cannot be used to interpret the protein behavior in terms of free energy landscape (FEL). Previous studies have indicated that FEL obtained from different

trajectories with cosine content below 0.2 or 0.5 produce qualitatively similar and smoother results with a single basin [36]. Hence, the cosine content was calculated on the first 20 PCs (projection eigenvectors) of each protein-DNA complex and the first two PCs with cosine content equal to or below 0.2 were selected as PC1 and PC2 to define the FEL. The FEL defines the energy minima on the energy landscape obtained based on the probability of the given combination of MD data points. This concept was used to map the minimum energy configuration of protein-DNA complexes during the simulation period and eventually connect the structural transitions between them. The g_sham program was used to calculate the FEL using the two PCs selected based on cosine contents.

Structural analysis of DNA

The structural analysis of DNA was carried out using 3DNA software [37, 38]. Here, the docked and most populated minimum free energy representative structures generated from the PCA and FEL analyses were used for calculation. The interior parameters such as *Roll (ρ)*, *Slide (D_y)* and *Twist (Ω)* of each dinucleotide base pair and the exterior parameters such as minor and major groove width of the DNA were considered to understand the bending, single degree of freedom and conformational transition[39-41]. The geometrical parameters of four DNA duplexes (A, B, CAP repressor contacted and TBP contacted DNA) from EI Hassan et.al. was used as reference [41]. The authors in their work used average values for comparison, although here we have considered the successive of each single nucleotide base pair and its conformational transition upon interacting with ETS1 protein. Based on their comparative notes, we observed two types of DNA distortions named mode I and II related to the *Twist (Ω)* vs *Roll (ρ)* and *Twist (Ω)* vs *Slide (D_y)* parameters and also the minor and major groove width for outlining the DNA distortion as obtained from their interior parameters. The mode I distortion is concerned with moderate high *Roll (ρ)* values and narrowing of the major groove width whereas high *Roll (ρ)* values and without penalty in terms of narrowing the major groove corresponds with mode II distortion. Information retrieved in order to describe the DNA distortions upon the influence of ETS1 were used based on the EI Hassan et. al. reference [41].

Binding energy calculation using MMPBSA

The binding free energy of the native and mutated $p16^{INK4a}$ – ETS1 complexes were calculated using the following formula. Here, the $G_{complex}$ defines the binding energy of protein-DNA complex whereas $G_{protein}$ and G_{ligand} represent the protein and DNA alone respectively.

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand})$$

The last 15 ns MD trajectory of all the complexes were extracted and subjected to g_mmpbsa calculation, a tool developed specially for GROMACS MD trajectory analysis [42].

Results and Discussion

Molecular interaction of ETS1 with native and mutated $p16^{INK4a}$ promoter region

The interacting conformation of native $p16^{INK4a}$ promoter with ETS1 shows good internal energy and binding energy in comparison with M1_ $p16^{INK4a}$, M2_ $p16^{INK4a}$ and M3_ $p16^{INK4a}$ ETS1 complexes (Table 1). The interacting conformation of native $p16^{INK4a}$ promoter with ETS1 was noticed to have fair internal and binding energy in comparison to that of M1_ $p16^{INK4a}$, M2_ $p16^{INK4a}$ and M3_ $p16^{INK4a}$ ETS1 complexes (Table 1). The PDBePISA interface [43, 44] analysis was carried out to investigate the binding free energy of the docked complexes. The analysis showed the native complex of $p16^{INK4a}$ – ETS1 to have solvation energy of -22.4 Kcal/mol whereas the complexes of M1_ $p16^{INK4a}$, M2_ $p16^{INK4a}$ and M3_ $p16^{INK4a}$ ETS1 complexes were observed to have lesser energies of -17.6, 18.0 and -18.6 Kcal/mol, respectively. The interface P-value of solvation free energy explains the nature of interaction-specific hydrophobicity towards complex formation. When the interface P-value is greater than 0.5, the interface of complex is less hydrophobic than the native state, whereas, a value equal to or lesser than 0.5 indicates the interaction-specific hydrophobicity. In this case the native $p16^{INK4a}$ – ETS1 complex had a P-value of 0.526, whereas, the mutated $p16^{INK4a}$ – ETS1 complexes (M1: 0.707, M2: 0.711 and M3: 0.639) was noticed to have a P-value more than 0.6 thereby clearly describing the reduced interaction-specific hydrophobic interface. The interactions of native and mutated $p16^{INK4a}$ – ETS1 complexes have been analyzed and are discussed clearly in the forthcoming sections.

Interaction of ETS1 with native $p16^{INK4a}$ promoter region

As evident from few other ETS1-DNA complexes, we observed the EBS core motif (**AGGAAG**) of the native $p16^{INK4a}$ promoter region flanked by and holding interaction with the triplet residues with R391, R394 and Y395 of H3 helix of ETS1 just as protein-DNA interaction observed in the

(Ets-1)2-S-EBS complex[14]. The NH1 atom of R394 points towards the N7 atom of DA13 establishing a hydrogen bond of 2.92 Å in distance and an N-H…π interaction with the imidazole moiety of DG12 (EBS) with a distance of 2.70 Å. The OH atom of Y395 forms a hydrogen bond with the N7 atom of DG9 (EBS) at a distance of 3.03 Å and also an N-π (imidazole) and N-H…π (pyrimidine) interaction with DA10 (EBS) is formed at a distance of 3.4 and 4.4 Å, respectively. The residue R391-NH1 was also observed to form a hydrogen bond (3.18 Å) with O1P atom of DG12 (Figure 1A). Though, the triplet residues interactions with DNA were observed to be similar to that of already explored crystallographic structure of ETS1-DNA complexes[12, 14, 15, 45], the unavailability of atomic level interactions is clearly discoursed in this study. The transformation of the auto inhibitory mechanism to functional form of ETS1 upon binding with DNA is aided by two conditions – interactions that accommodate for binding and interactions that affect the conformation of the inhibitory module[15]. Representing the previous state, the loop of bHLH motif and the loop connecting B3 and B4 shift was observed through NMR studies and proven to facilitate for the proper nucleic acid contacts subsequently initiates functional form of ETS domain. Additionally previous studies have reported the helix H1 (Amide group of L337) to be positioned precisely to form hydrogen bond with DNA phosphodiester backbone[12, 15, 46]. Accordingly, along with the triplet residues Q336 (beginning of H1 helix), Y396, K399, H403, G407 and R409 residues of the ETS domain are noted to make hydrogen bonds with DG9, DA10, DC20 and DC22 of p16[INK4a] whereas residues L337, W375, K379, M384, K388, Y397, K404 and K408 of ETS domain connect via a VdW interaction to DA10, DA11, DG12, DA13, DG9, DT20, DT21 and DT22 of p16[INK4a] (S1A). In contrast to the other ETS1-DNA complex, L337 is involved for VdW interaction with p16[INK4a] promoter region whereas polar charged amino acid Q336 forms hydrogen bond with phosphodiester backbone of DG9 by a contact distance of 2.96 Å. All these interacting residues are highly conserved in complexes formed by ETS1 with the promoter region of other targets such as ETS1$_2$-S-EBS complex. Moreover, the experimental studies of different ETS1-DNA complexes show that the alternative conformations of contact residues play a crucial role for the target site selection by ETS domain of ETS1 protein[46, 47]. Accordingly, there are different kinds of interaction and its subsequent transition leading to the conversion of auto inhibitory state to functional form was observed in this study.

Interaction of ETS1 with mutated p16^{INK4a} promoter region

A comparison of the interacting complex of ETS1 – M1_p16^{INK4a} with ETS1 – Native_p16^{INK4a} complex shows a structural RMSD of 1.046 Å as noted through superimposition of the two complexes and the DG12 base is observed to displace about 2.2 Å and 26.5° with respect to native complex. The displacement can be reasoned with a mutation in the region preceding the E-box motif of p16^{INK4a} promoter (5'-CACC*CC*AGGAAGAAAGA-3'). The minor and major groove constituted by mutated *CC*AGGAAG was seen to have little variation in their groove width when compared with the native complex (displayed in Figure S2). Such DNA displacement induces changes in its interaction with the triplet residues of ETS1. As a result, the phenyl ring of Y395 residue flips and the CB atom is displaced by about 0.622 Å forming C=O…π (3.3 Å) and N…π (3.6 Å) interactions with pyrimidine and imidazole moieties of DG9, respectively. The interaction of Y395 with DA10 of p16^{INK4a} which was observed in the native complex was lost in the mutated complex thus citing the already proven importance of the triplet residue towards DNA binding and its subsequent transition of the ETS domain into its functional form[12, 14, 15]. Likewise, R394 tilted by about 1.186 Å with respect to NE atom and establishes a hydrogen bond between its NH1 atom and O6 of DG12 (contact distance 3.43 Å) and an N-H….π interaction (contact distance 4.7 Å) with imidazole ring of DA11. Though R391 has tilted 1.470 Å, it maintains the hydrogen bond with DA11 strongly and the contact distance becomes closer (2.85 Å) than the native complex (Figure 1B). In addition to it, other residues Q336, L337, W375, K379, K388, Y396, Y397, K399, K404, G407, K408 and R409 also form hydrogen bond and non-bonded contacts with DG8, DG9, DA10, DA11, DG12, DT20, DT21, DT22, DC23 and DT25 (S1B). Although, similar interactions were observed in E-box motif of native and mutant complexes, the buried surface area, interaction energy and DNA base displacements renders the mutants less effective than the native complexes.

The interaction profile of ETS1 with M2_p16^{INK4a} promoter region shows minor variation in the minor groove width due to mutation at the E box motif (GG to TT) and shows an RMSD of 0.545 Å upon superimposition with native complex. One of the triplet residues, Y395, is tilted about 0.498 Å (with respect to OH) and forms N-H…π and N…π interaction with DA11 at a contact distance of 3.0 and 4.2 Å, respectively. R391 moves about 1.076 Å apart and forms a hydrogen bond (2.94 Å) with O1P atom of DA13 as already reported in ETS1-DNA

10

complexes[14, 15] in contrast to the hydrogen bond formed with DA11 in native complex. Also, the R394 moves by about 1.571 Å when compared with the native complex and loses its potential hydrogen bonding capability and N-H…π interaction with E-box motif (Figure 1C). In spite of these differences, the M2_p16^{INK4a} – ETS1 complex possesses better interface area (1086.3 Å^2) well stabilized by hydrogen bond and non-bonded interactions between I335, Q336, L337, W375, K379, K381, K388, Y391, Y397, K399, H403, K404, R409 and DA10, DA11, DG12, DA13, DA15, DT18, DC19, DT20, DT21, DA27. This complex also shows lesser solvation free energy at the interface when compared with the native complex (S1C). As a result, the effect of this mutant lies in the reduced interaction capability and subsequent effects in ETS1 binding.

The double mutant M3_p16^{INK4a} promoter interaction with ETS1 does not show significant deviation in both major and minor groove width as observed in native promoter ETS1 complex. Superimposition of the native and M3_p16^{INK4a} –ETS1 complexes shows an RMSD of 0.566 Å, which is lesser than other two mutant complexes and signifies a similar interaction as observed in native complex. The Y395 phenyl ring moves about 0.550 Å away with respect to NE atom in comparison with native complex and makes N-H…π interaction with pyrimidine moiety of DA10 (2.7 Å). As seen in the native complex, the M3_p16^{INK4a} – ETS1 R394 also forms N-H…π interaction with imidazole ring of DG12 and is titled from its native position at about 0.634 Å of RMSD. Similarly, R391 also maintains its hydrogen bond with O1P of DG12 at a distance of 3.05 Å (Figure 1D). Though, the interacting base pair is different from that of native complex, the triplet residues contribution was observed to similar with other experimentally determined ETS1-DNA complexes[14, 15]. Additionally, the residues Q336, L337, K379, K381, K388, Y396, Y397, K399, D398, H403, K404, G407 and R409 makes hydrogen bond and non-bonded contacts with DT9, DA10, DA11, DG12, DA13, DC19, DT20, DT21, DT22, DC23 and DT24 (S1D). Though, the interacting residues are well conserved in comparison with native complexes, the solvation free energy and interface P-value of solvation free energy shows less significant interaction between ETS1 and p16^{INK4a}.

Structural stability of ETS1 with native and mutated p16^{INK4a} promoter region

The stability of ETS1 was analyzed by comparing the backbone RMSD of ETS1 in complex with native and mutated p16^{INK4a} promoter. In the native p16^{INK4a} – ETS1 complex, the RMSD of ETS1 is stabilized at 0.5 to 0.6 nm after 10 ns of the 50 ns production MD run. In case of

M1_p16^{INK4a} – ETS1 complex, ETS1 fluctuates gradually from 0.5 to 0.9 nm and stabilizes only after 40 ns. However, the RMSD of M2 and M3_p16^{INK4a} – ETS1 complexes are equilibrated at 0.6 and 0.75 nm after 25th and 20th ns, respectively (Figure 2A). These RMSD profiles describe the equilibrium strategies of native and mutant complexes where the native p16^{INK4a} – ETS1 complex is well stabilized than the mutants. This indirectly signifies the dynamics nature of ETS1 protein upon binding with p16^{INK4a} promoter. The RMS fluctuation shows that in the native p16^{INK4a} – ETS1 complex, the loop connecting HI2 and H1 of ETS1 express fluctuations than the other mutated complexes. Other regions of ETS1 do not deviate much in both native and mutated p16^{INK4a} complexes with an exception of M2_p16^{INK4a} – ETS1 complex. In this M2_p16^{INK4a} – ETS1 complex, the H1' helix, the loop connecting β1-β2 and the loop connecting H2-H3 (HLH motif) of ETS1 show fluctuations upon binding with p16^{INK4a} promoter. The loop connecting H2-H3 helix was proven to fluctuate more in the auto-inhibited form of ETS1 protein (PDB ID: 1R36) as solved through NMR studies[15], having an RMS fluctuation up to ~4 Å in the M2_p16^{INK4a} – ETS1 complex hence showing the indirect preservation of auto-inhibitory state of the ETS domain. In all the complexes, the HI2 helix and the loop connecting HI2 and H1 shows variation which indirectly explains the structural dynamics of N-terminal helix (Figure 2B). In order to understand the structural transition upon mutation at the p16^{INK4a} promoter region, Principal component analysis and Free energy landscape analysis was carried out.

Structural transition of ETS1 using PCA, FEL and T-pad analysis

The N-terminal HI1 inhibitory helix of ETS1 folds and makes a series of hydrophobic interactions with another N-terminal inhibitory helix (HI2) and C-terminal (H4 and H5) region[12, 15]. These interactions are crucial for maintaining the ETS domain in its compact, triangle shaped, and auto-inhibited state. Upon binding with p16^{INK4a}, the helix HI1 unfolds and the consequent relaxation of tight hydrophobic bridges with other inhibitory helices initiates the transition from triangular ETS domain to an oval shaped ETS domain. The helical nature of HI1 is lost in DNA bound conformation and its functional form[12, 14, 15]. The model used in this study is retrieved from the dimeric form of ETS1 where the helical nature of HI1 remains unaltered and is retained during the docking study with promoter of p16^{INK4a}. However, this helix showed major variations during simulations and the results are summarized below.

The principal components extracted with a cosine content of 0.2 for all ETS1 complexes (with native and mutated p16^{INK4a} promoter) were used to construct the FEL contour map. The native, M1, M2 and M3 p16^{INK4a} – ETS1 complex shows four, seven, six and four lowest free energy representative structures, respectively, and the helical orientation of the corresponding ETS domains are shown in Figure 3. In the native p16^{INK4a} – ETS1 complex, HI1 is not located parallel against HI2 as like other ETS1 structure (auto-inhibited form)[12, 15] although it is placed towards H2 and subsequently loses its helical nature due to the absence of interaction with other inhibitory helices[46]. But, in M1 and M2 states, the helical nature of HI1 remains unaffected although it is not parallel to HI2. Surprisingly, in M3 double mutant state, HI1 is placed exactly parallel to HI2 and its helical nature is also well maintained which is one of the sole criteria for auto-inhibited form of ETS1 verified to be formed upon interacting with mutated p16^{INK4a} promoter region[15, 46]. This evidently supports that double mutation of the p16^{INK4a} promoter region maintains the ETS domain's auto-inhibitory state.

The angular dispersion plot was generated from the 50 ns simulation data for each ETS1 p16^{INK4a} complex using the T-pad analysis (for methodology refer supplementary information). Accordingly, the residual fluctuation, transition and short transition of each ETS1 protein complex are displayed in Figure 4. In the Native p16^{INK4a} – ETS1 complex, the N-terminal HI1 helix residues (Y307, R309 and D310), the loop connecting HI1 and HI2 (K316 and K318) and HI2 helix residue (T330) attains full transition during the 50 ns MD simulation. Literature studies have indicated residues K305-R309 and D306-D310 forming unfavorable (i, i+4) charge-charge interaction in the functional form as well as in the free form due to the alignment along the HI1helix[15, 48, 49]. This may lead to the unfolding mechanism of HI1 helix upon binding with DNA. Our analysis on the PAD degrees revealed residues attain full transition ranging from 60-110 °, thus exploring the poor arrangement of the residues that lead to the unfolding of HI1 helix. Also, the polar residues K383, M384 and N385 of the loop connecting H2 and H3 (HLH motif) accomplishes full transition and K381 holds only short transition based on their PAD degree (Figure 4A). The residues of loop connecting HI2 and H1 (S332 and G333), shows higher fluctuation with higher PAD degree ranging from 70-130°. The S332 was proven to form hydrogen bond with Q339 and the disruption of this bond helps in the shift of HI2 and HI2/H1 loop subsequently leading to DNA bound state (functional form) of ETS domain[15]. The mode

vector graph clearly explains that the loop connecting H2 and H3 of ETS domain and H2-H3 move towards each other. From the angular dispersion plot and mode vector analysis it becomes evident that the loop region of HLH motif plays a crucial role in interaction with p16[INK4a] promoter region, which subsequently influences the structural transition of HI1 and ETS domain from auto-inhibited to functional form, similar to that of a DNA bound ETS1 crystallographic structure[14].

In case of M1_p16[INK4a] – ETS1 complex, full transition was observed for the residues R309, D310, R311, K318, V320, I321, T330, G331 and S332 belonging to HI1, loop connecting HI1 and HI2 and loop connecting HI2 and H1. Similarly, the residues K381, K383, and M384 (loop connecting H2-H3) and K394 (H3) of HLH motif achieves full transition although there is a split in the H3 helix which is observed due to transition in K394 (Figure 4B). Also, a short transition was observed in the residues K399 and N400 of loop connecting the H3 and β3. The mode vector analysis denotes that the loop of HLH motif moves away from the direction of movement of H2 and H3, which in turn move towards the major groove of p16[INK4a] promoter region. Hence, it can be concluded from angular dispersion analysis that, less number of transitions is observed in the HI1 region and there is no specific factor (loop of HLH motif) which influences the ETS domain movement. The M2_ p16[INK4a] – ETS1 complex shows fluctuations of about 40-80 PAD degree only in the HI1 helix and the loop connecting H2 and H3. Since the mode vector of loop (HLH motif) moves away from the direction of H2 and H3, the consequent transition of ETS domain is lost (Figure 4C). The M3_ p16[INK4a] – ETS1 complex describes less number of transition in HI1 (V308, R311 and L314) showing no structural deviation and K318 (loop connecting HI1 and HI2) attains a short transition. In H3 helix, the loop connecting H2-H3 and H3-β3 and β3 sheet acquired more number of transitions by its residues such as K381, K383, G392, R394, Y397, K399, N400, I401 and K404 (Figure 4D). Among these, the fluctuation observed in the loop connecting the H2-H3 showed higher PAD degree of 130°. The high degree of fluctuation showing residual changes in the ETS domain which was not observed in native and M1-M2 complexes. Consequently, the mode vectors also show that the loop of HLH motif moves towards H3 helix and H2 and H3 helices move towards the direction of loop. Also, the β3 and β4 was noticed to move towards the direction of H3 which subsequently makes the ETS domain to be in a compact state as in the auto-inhibited state[15, 46].

14

Structural transition of native p16^{INK4a} – ETS1complex

The representative structure of Native p16^{INK4a} – ETS1 complex extracted from FEL upon superimposition with the docked conformation shows a displacement in p16^{INK4a} promoter region (indicated in block arrows) and the factor (loop connecting HI1-H1 and H2-H3, H3 helix) which influences this displacement by strong interaction is denoted by orange arrows (Figure 5A). Additionally, the tight interaction between HLH motif of ETS domain and the minor and major grooves of DNA induces structural transition. Interestingly, helix HI1 is placed in between HI2 and H2, unlike its auto-inhibited state as observed in crystallographic structure[14], indicating the occurrence of transition. The following structural changes attribute to the key differences in the structures of ETS1 before and after simulation and several plausible mechanisms for conformational transition can be elucidated and compared with auto-inhibited and functional form of ETS1[14, 15, 46]. Mechanism I (Formation of clamp): The residue P334 in the HI2-H1 loop forms hydrophobic interaction with DT28 and DG8 of p16^{INK4a} promoter and this defines the formation of a primary hydrophobic clamp. The crystallographic and NMR structure of ETS1 with DNA speaks of hydrophobic interaction observed between the HI1 helix which was aligned parallel to H1 helix thus making the different state of ETS domain by unfolding the HI1 helix[12]. Accordingly, the simulation result of native p16^{INK4a} - ETS1 show a hydrophobic clam which helps to disrupt these hydrophobic patches. Additionally, the residues K379, K381 and K388 of H2-H3 loop proven to contact with DNA surface which forms a triangle basic patch and express three hydrogen bonds with O2P atoms of DA10, DT28 and DG12 with a contact distance of 2.69, 2.93 and 2.54 Å, respectively. These interactions hold the intermolecular contacts and are proven to stabilize the surface contacts[45] which subsequently help for the stability of primary hydrophobic clamp. Besides, the triplet residues R391, R394 and Y395 of ETS1 orients towards the major groove of DNA E-box motif by establishing three hydrogen bonds with DG12 (O1P), DG12 (O6) and DG9 (O1P) with a distance of 2.65, 3.00 and 2.86 Å respectively. Also, K399 and Y396 form hydrogen bond with DG9 and DA10 respectively, which collectively helps in holding the major groove of p16^{INK4a}. All these interactions together constitute a specific interacting pattern and are maintained throughout the simulation. This interaction network could establish a pulling force via the HI2-H1 loop and provides a way for the folding of HI1 helix towards H1 and H3 and thus be stabilized by a series of intra hydrophobic interactions[14].

Mechanism II: Here, the residues I335, F340, L341 and L344 of H1, are the central part of the ETS domain proposed for the allosteric mechanism of auto-inhibition[15, 50] which forms series of hydrophobic contacts with W375, R378, R374 and V371 of H2 whereas E343 of H1 forms two hydrogen bonds with R378 and disrupts its structural state. Furthermore, hydrophobic contacts were formed between A327, A323 and I321 of HI2 and W338, L326 and L342 of H1 helix (mechanism III) and are shown to have higher relaxation in the NMR studies of auto-inhibited ETS1 structure[15]. All these together makes the pulling and folding of HI1 helix towards DNA and makes series of basic patch (R309, R378 and R374) to acidic patch (D347 and E343) interaction with residues of HI2 and H1 helix (mechanism IV).

These interactions also induce displacements in the structure of p16^{INK4a} promoter DNA and are described based on the external (minor and major groove width) and internal (*Roll (ρ)*, *Slide (D_y)* and *Twist (Ω)*) aspects of DNA helical structure[39, 41, 51]. The *Roll (ρ)* and *Twist (Ω)* of each dinucleotide along with the reference DNA is compared in Figure 5B and the variable step parameters are listed in Table 2, where high *Roll (ρ)* angles relating to particular dinucleotide base parameters are highlighted in bold due to its impact on DNA distortion (Supplementary Table 1). A high *Roll (ρ)* angle is endorsed by 4th and 11th single dinucleotide bases (CG/CG and AG/CT) with an angle of 12.92 and 16.19°, respectively. Also, no clear correlation was observed with reference data[41] in order to classify them as mode I and II distortion due to its peculiar distribution in the plot. In general, the type of DNA distortion cannot be classified based on single dinucleotide step alone but successive bases should also be considered. The *Slide (D_y)* and *Twist (Ω)* plot (Figure 5C) clearly depicts that 5-GG/CC and 6-GA/TC lie in the clear conformational channel[39, 52] of mode I DNA distortion which is also represented by 4-CG/CG (moderately high *Roll (ρ)* angle). The 10-AA/TT comes under the mode II distortion as represented by 11th AG/CT (high *Roll (ρ)* angle), though the 12-GA/TC lies in the conformational channel of mode I. Previous studies have put forward that, the CA/TG and AA/TT[39, 53, 54] single dinucleotide step induces a severely bent DNA conformation upon binding with protein and is classified as mode II distortion[41]. Accordingly, the 10th AA/TT is also observed to attain mode II distortion by the influence of ETS1 protein binding which were found to be rigid in naked DNA[54]. An indirect relationship was also found to define the DNA distortion using the exterior parameters (minor and major groove width) that was calculated for

un-simulated and simulated free energy representative complex and plotted in Figure 5D. As described by EI Hassan et.al. a high *Roll (ρ)* angle defines mode I distortion and is accompanied by a narrowing of the major groove, whereas in mode II distortion, the *Roll (ρ)* angle increases without any influence on the major groove. Similarly, the free energy representative complex shows narrowing of major groove width (20.7 Å) and widening of minor groove width (13.6 Å) of 5-GG/CC than other single dinucleotide step. Such an observation has already been reported, wherein the GG/CC influences the opening of minor groove in PUT3-DNA complex based on the hydrophobic seal[55]. The mode II distortion attributed by AG/CT in the major groove of ETS domain does not show any effective changes in the groove width as compared with native complex. All these results clearly signifies that, the mode I distortion is attributed in the minor groove of DNA preceding the E-box motif and influenced by the loop connecting the HI2-H1 and H2-H3, whereas, the mode II distortion at the E-box motif and successive bases is due to the effective H3 helix hydrogen bonding. This observation is noted to be the expected mode of DNA distortion upon binding with ETS1 protein which subsequently facilitates its transition from auto-inhibited to functional form which will be discussed further.

Structural transition of M1_p16^{INK4a} – ETS1complex

Superimposition of docked M1_p16^{INK4a} – ETS1complex with free energy representative structure clearly shows the displacement (distance between P334 and DG is 7.2 Å) of HI2 - H1 loop and distortion in minor groove of mutated DNA leading to the loss of pulling force between them (Figure 6A). In addition, the triangle basic patch interaction formed by K381, K388 and K379 with DNA is disrupted and loses the hydrogen bond formed by K381 in its native free energy representative structure. However the residues K379 and K388 were shown to be highly flexible residues in the NMR structure of ETS1[15] forming two hydrogen bonds with DA10 and DA11 with a contact distance of 2.85 and 2.96 Å respectively (mechanism I). Among the triplet residues, only R394 is able to maintain the hydrogen bond with DA11 (2.74 Å), whereas, R391 and Y395 is involved only in non-bonded interactions while losing their hydrogen bonds observed in the native representative structure. The Y396, K399 and K404 form additional hydrogen bonds with major groove of DNA though the contribution of triplet residues to hold the H3 helix in major groove is absent. Absence of the hydrophobic interactions by P334, perfect

triangle basic patch and triplet residue interactions results in weak interactions between other helices of ETS1 domain. Therefore, the helical orientation of H2 (W375, R378 and R374) and H1 (P334, I335, F340 and E343) maintains weak hydrophobic interactions due to their helical displacement of about 10.9 Å (mechanism II) which is 10.7 Å in the native free energy representative structure. Further, this displacement along with H5 induces a displacement in the orientation of HI2 and H1 to 10.1 Å (native HI2 and H1 helical distance is 7.1 Å) and results in less effective hydrophobic interaction by the residues of HI2 (L326 and I321), H1 (E343 and T346) and H5 (L418, L421 and L422) (mechanism III). Altogether, the displacement of H1-H2 and HI2-H1 and the subsequent effect in hydrophobic interactions renders the HI1 helix unfolded persistently during the 50 ns MD simulation (mechanism IV). The displacement of loop connecting HI2 and H1 is attributed to the mutation (GG to CC) preceding the E-box motif which induces the hydrophobic interaction of P334 in native state. The variable step parameters such as *Roll (ρ)*, *Slide (D_y)* and *Twist (Ω)* and their relationship plot for single dinucleotide base and minor and major groove widths of DNA structure were used to define the DNA distortion (Table 2). Three high *Roll (ρ)* values were observed at 3rd, 6th and 12th dinucleotide positions (Supplementary Table 1) and the relationship plot with *Twist (Ω)* shows that these values are distributed and hence it is not possible to classify their mode of distortion (Figure 6B). The *Slide (D_y)* and *Twist (Ω)* relationship plot describes that the 3rd single dinucleotide base CC/GG with high *Roll (ρ)* angle of 16.16° (Figure 6C) comes under mode I distortion. In spite of its high *Roll (ρ)* value, the exterior aspects like minor (9.4 Å) and major (21.3 Å) groove width has less effective changes in contrast to a regular mode I distortion (Figure 6D). The next moderate *Roll (ρ)* value (angle: 8.80°) corresponds to 6-CA/TG and the *Slide (D_y)* and *Twist (Ω)* relationship plot and denotes a mode II distortion due to its position related with Tc[53]. However, this distortion is classified as mode I in this study due to (i) the absence of protein impact on DNA, (ii) the position of preceding single dinucleotide bases (parallel to mode I distortion range) (iii) the narrowing of major groove (5 Å) and (iv) widening of minor groove (3 Å) width (Supplementary Table 1). This is also supported from the literature that the CA/TG in naked DNA might attain mode I distortion[41]. Also, the 6-CA/TG single dinucleotide expresses vast deviation in the *Twist (Ω)* values (about 11°) and this distortion makes a gap which consequently makes the HI2-H1 loop inaccessible. The third high *Roll (ρ)* value (18.77°) is attributed to 12-

GA/TC which clearly shows mode I distortion near to A-DNA[39] (Figure 6C), but is inconsistent due to minor deviations in the groove width values. Though, it shows very high *Roll (ρ)* angle and comes near to Cc (mode II distortion) in *Roll (ρ)* and *Twist (Ω)* plot, the influence of protein interactions make the distortion controversial. In conclusion, mutations preceding the E-box motif creates mode I distortion in DNA without any direct influence at the protein interfere, whereas, induced mode I distortion at the E-box motif influenced by protein interactions.

Structural transition of M2_p16[INK4a] – ETS1complex

The free energy representative M2_p16[INK4a] – ETS1complex shows the folding of HI1 helix towards HI2 due to the pulling force generated by hydrophobic interaction between HI2 and H1 loop and minor groove of DNA. The P334 and Q336 make hydrophobic contacts with DT28 and DA10 at a distance of 4.8 and 4.3 Å, respectively, which pulls the HI2 helix towards DNA minor groove (S3A). The triangle basic patch (K381, K379 and K388) interaction observed in native free energy representative structure is absent due to mutation in E-box motif of p16[INK4a] promoter and only K381 forms one hydrogen bond with DA27 (2.72 Å). The triplet residues R391, R394 and Y395 makes strong hydrogen bond towards DA13, DG16 and DA10 with a distance of 2.81, 2.85 and 2.66 Å, respectively (mechanism I). In addition to the triplet residues, the Y396 and K404 also form hydrogen bonds with DA11 (2.67 Å) and DT18 (2.86 Å), respectively. These above mentioned interactions consequently creates a series of hydrophobic contacts to maintain the helical orientation of H1 (I335, E343, L344 and F340) with respect to H2 (R378, W375, R374 and V371) at a distance of 11.7 Å (mechanism II). As a result the HI2 (L326, A323 and T330) and H1 (Q339, L342, F340 and E343) interaction initiates the folding of HI1 helix towards HI2 (mechanism III). This folding is stabilized by basic-acidic patch (R311, R378 and R374 with D317, D347) and hydrophobic interactions (Y307, F304, K377 and K379) for the entire 50 ns MD run (mechanism IV). Though HI1 attains a proper fold as seen in the native free energy representative structure, the inability of forming a perfect triangle basic patch interaction and loss of clamp hydrophobic interaction affects the structural transition. The distortion of p16[INK4a] DNA upon binding with ETS1 was analyzed based on *Roll (ρ)-Twist (Ω)* (S3B) and *Slide (D_y) - Twist (Ω)* plots (S3C). From both plots it can be observed that the 5-GG/CC with moderately high *Roll (ρ)* angle (11.06°) is placed parallel to A-DNA (S3B plot) and

close to B-DNA distortion (S3C plot). Also, about 2 Å narrowing of the major groove width and about 0.5 Å (Supplementary Table 1) widening of minor groove together classifies the 5-GG/CC as mode I distortion. The dinucleotide pair 7-AT/AT (15.90°) and 8-TT/AA (7.88°) also shows high and moderate *Roll (ρ)* angle due to mutation in the E-box motif. 7-AT/AT is classified as a mode I distortion due to its placement in conformational channel of S3C plot and about 4 Å narrowing of the major groove (S3D) with a high *Roll* angle. The 8-TT/AA single dinucleotide step accomplishes mode II distortion due to its separation from the mode I distortion area in both plots and a stable major groove width in spite of the moderate *Roll* angle. Collectively, these two (7-AT/AT and 8-TT/AA) single dinucleotide base step forms mode II DNA distortion due to its strong interaction with ETS1 triplet residues but least effective when compare with the native free energy representative structure.

Structural transition of M3_p16[INK4a] – ETS1complex

The double mutant (GG to CC preceding the E-box motif and GG to TT at the E-box motif) p16[INK4a] promoter upon complex with ETS1 stabilizes HI1 folding towards HI2 and forms a perfect helical orientation in comparison with other complexes mimicking the auto-inhibited form of ETS1. The HI1 in M3 is folded as in native and M2_p16[INK4a] – ETS1complex through hydrophobic contacts between P334 and L337 of ETS1 and DT28 and DT9 of DNA which is in contrast to HI1 in M1_p16[INK4a] – ETS1complex (S4A). As compared to M1_p16[INK4a] – ETS1complex, the P334 hydrophobic pulling is absent due to mutation preceding the E-box. Whereas, the mutated E-box motif forms hydrophobic interaction with L337 and Q336 and subsequently, K379 and K381 forms two hydrogen bonds with DA11 and not forming a perfect triangle basic patch. In addition, among the triplet residues only R391 and Y395 forms hydrogen bond with DG12 and DT9 with a contact distance of 2.71 and 2.78 Å, respectively (mechanism I). All these interactions strengthen the protein-DNA complex similar to native and M2_p16[INK4a] – ETS1complexes and thus balances the effect of double mutant on HI1 folding mechanism. Consequently, the helical orientation of H1 (I335, E343, L344, L341 and F340) and H2 (R378, W375 and V371) is stabilized at a helical distance of 10.8 Å (closer to native p16[INK4a] – ETS1complex) by a series of hydrophobic contacts (mechanism II). This also influences the hydrophobic interactions between HI2 (A327, L326 and A323) and H1 (Q336, I335, F340, E343

and L342) which interact strongly at a helical distance of 7.9 Å (mechanism III) similar to native free energy representative structure. The HI1 (Y307, F304 and V308) is folded towards HI2 (P322 and Y329) with a perfect helical orientation distance of 9.3 Å stabilized only by hydrophobic interactions (mechanism IV). Although the folding mechanism of M3_p16[INK4a] – ETS1 and native p16[INK4a] – ETS1complexes are similar, the absence of Triplet residue and basic patch interactions due to double mutant affects the transition of ETS domain from auto-inhibited to functional form. The Figure S4B and S4C shows the relationship between *Roll (ρ)-Twist (Ω)* and *Slide (D_y)-Twist (Ω)* values describing the mode of DNA distortion. The first two moderate high Roll angle (Table 2) corresponds to 3-CC/GG (11.45°) and 5-CC/GG (11.20°) which are located near to Cc and A-DNA distortion, respectively, in both the plots. Since the distribution of *Roll (ρ)-Twist (Ω)* plot is not in correlation with the reference data, the *Slide (D_y)-Twist (Ω)* plot has to be considered along with the narrowing (18.9 Å) and stable (20.6) major groove width to classify 3rd and 5th under mode II and mode I distortions (S4D), respectively. Since protein binding does not impact 3-CC/GG, it can also be classified as mode I distortion. The next high Roll angle is attributed to 9-TA/TA (14.50°) which is clearly plotted near the A to B-DNA transition via mode I distortion in *Slide (D_y)-Twist (Ω)* plot and accordingly the major groove width is narrowed about 2.5 Å. The very high *Roll* angle of 12-GA/TC (16.06°) was classified as mode II distortion due to its placement towards Tc distortion and stable major groove width (20.2 Å). In general, the double mutant induces mode I, protein influenced mode I and mode II distortion.

Mechanism towards the structural transition of ETS domain from auto-inhibited to functional form

ETS1 protein upon binding with their target promoter region attains a structural transition from auto-inhibited to functional form observed as the conversion of ETS domain from triangle to oval shape. In the auto-inhibited form, the H1, H2, H3 and β1-β4 of ETS domain is tightly packed by hydrophobic interactions due to close helical orientation of HI2 and H4 inhibitory helices. This tight packing is influenced by inhibitory helices which subsequently determine the structural transition [56] and hence, the helical distance of HI2 and H4 and its contribution towards the transition of ETS domain was analyzed. In order to identify the shape of ETS domain the

distance between H1, H2, H3 and β1-β4 was calculated using the CA residues from each helix (H1: L337 and L344 or T346 or D347; H2: P368 and K379; H3: Y386 and D398) and anti-parallel sheet (β1-β4: I402 and I354) as reference points. The H3 helix was used to define the base of ETS domain. Figure 6 describes the transition of ETS domain in native and mutated $p16^{INK4a}$ – ETS1complexes. In native complex, the distance map clearly shows that, the ETS domain attains an oval shape which is confirmed by the displacement of helix H2 away from H3 about 5.7 Å. Also, H2 is twisted away from the center of ETS domain and the distance of β1-I354 and H1-L344 from β3-I402 is similar (15.5 and 15.6 Å) signifying an oval shape. In this relaxed form of ETS domain, the HI2 helix is displaced away from H4 inhibitory helix to about 13.0 Å (Supplementary Table 2) leading to better binding energy with promoter (-9318.508 kJ/mol) calculated using MMPBSA (Table 3). In case of $M1\_p16^{INK4a}$ – ETS1complex, the binding free energy is noted to be -7470.787 kJ/mol, lesser than the native complex and supported by the weak interaction reflected in the transition of ETS domain. The CA residues of H2 (Y386) and H3 (K379) are closer in $M1\_p16^{INK4a}$ – ETS1complex. The stable variation in the distance of β1-I354 (15.2 Å) and H1-T346 (16.2) to β3-I402 and the correlated motion of HI2 and H4 inhibitory helices (helical distance of 12.0) (ST 1) collectively defines a partial triangle shape as a result of mutation preceding the E-box motif weakening the complex. Similarly, $M2\_p16^{INK4a}$ – ETS1complex also forms partial triangle shape due to the stable variation in distance formulated with β3-I402 to β1-I354 (15.4 Å) and H1-D347 (19.8) and the orientation of HI2 towards H4 inhibitory helix. In contrast to the native complex, the $M2\_p16^{INK4a}$ – ETS1complex showed better binding energy of about -9857.330 kJ/mol, which might be due to the interaction of HI1 helix with the minor groove of DNA. However, the lack of basic patch interaction and triplet residues interaction with the major groove defines their partial auto-inhibited nature. The $M3\_p16^{INK4a}$ – ETS1 complex also shows a binding energy (-9515.791 kJ/mol) higher than the native complex and can be attributed to the interaction of HI1 helix with DNA. Despite of all these observation, the $M3\_p16^{INK4a}$ – ETS1 complex shows fully auto-inhibited form by forming an exact triangle shape as identified by the distance of β3-I402 to β1-I354 (15.1 Å) and H1-L345 (13.7 Å). The compact HI2 and H4 inhibitory helices with a helical distance of 12.3 Å also stabilizes the triangle shape as reinforced by the distance between K379 of H2, L337 of H1 and Y396 of H3. The H1 helix was extended two, three and one residue in

M1, M2 and M3 mutated complexes, respectively. This helix extension makes a vast difference in the shape of ETS domain along with their intramolecular (within the ETS domain) and intermolecular (basic patch and triplet residues interaction with p16^{INK4a} promoter region) interaction pattern. Comparing all the results, we hypothesize that in the native form, ETS domain attains the functional form, while in M1 and M2 it attains only a partial triangle shape. But, in M3 mutation the ETS domain retains completely auto-inhibited form, due to weak interactions with double mutant p16^{INK4a} promoter.

Conclusion

The transcriptional activation of certain genes is regulated by the interaction of transcriptional proteins with their respective promoter regions. The p16^{INK4a}, a tumor suppressor gene is regulated by the active interaction of ETS1 protein with its promoter. Two structural forms (functional form and auto-inhibited form) of ETS1 protein were already observed upon binding with other promoter region of target DNA, though clear elucidation of ETS domain arrangement and the role of inhibitory helices in structural transition is yet to be addressed specifically for p16^{INK4a} promoter induced ETS domain transition. Also, mutations in p16^{INK4a} promoter preceding and within the E-box motif results in improper interaction with ETS1 which subsequently inhibits the expression of gene (p16^{INK4a}) responsible for cancer suppression. Hence, the present study describes the molecular interaction of ETS1 with native and mutated p16^{INK4a} promoter regions along with their functional interaction structurally analyzed using protein-DNA docking protocol. The interaction energy of the docked complexes clearly shows that native p16^{INK4a} – ETS1 complex has better solvation free energy interface than the mutated complexes. Further, to understand the residual and structural transition of ETS1 protein upon binding with native and mutated p16^{INK4a}, molecular dynamics simulations were carried out for 50 ns production MD run and analyzed using t-pad, FEL and MMPBSA calculations. Accordingly, the native p16^{INK4a} – ETS1 complex representative structure clearly elucidates that, the hydrophobic interactions of P334, triangle basic patch interaction and triplet residues contribution towards the minor and major groove of DNA defines two types of structural transitions: first is the folding of HI1 helix towards HI2 and second describes the auto-inhibited to functional form transition of ETS domain via displacement of HI2 from H4 inhibitory helix

which is further confirmed based on the MMPBSA binding energy. Additionally, p16^{INK4a} DNA distortion analysis upon binding with ETS1 was performed using 3DNA software, that shows two types of DNA distortion (mode I and II) based on moderate and very high *Roll* angles, respectively. Considering the native p16^{INK4a} – ETS1 complex as a reference for understanding the structural transition, the mutant M1_p16^{INK4a} – ETS1 complex shows distortion of DNA in three single dinucleotide bases. Among these, the first two attributes mode I distortions were induced due to mutation in the DNA alone, whereas, the third mode I distortion occurs as an influence of ETS1 H3 helix interaction. This subsequently displaces the HI2 - H1 loop leading to unfolding of HI1 helix. Further, the movement of HI2 helix towards H4 inhibitory helix induces the ETS domain to attain a partial triangle shape. Similarly, the ETS domain of M2_p16^{INK4a} – ETS1 complex also attains a partial triangle shape due to the close proximity of its HI2 and H4 inhibitory helices. The folding of HI1 helix towards HI2 creates a tight hydrophobic interaction between the minor groove of DNA and ETS1. Though M2_p16^{INK4a} shows protein mode I and II DNA distortion in minor and major groove, respectively, the lack of triangle basic patch and triplet residues interaction makes this complex less effective.   In case of double mutant M3_p16^{INK4a} – ETS1 complex, the ETS domain maintains a perfect triangle shape even in complex with DNA due its double mutation and strongly interacting HI2 and H4 inhibitory helices. The M3_p16^{INK4a} mutant expresses four high *Roll* angles corresponding to mode I, protein induced mode I and mode II DNA distortion which subsequently affects both basic patch and triplet residues interactions. Altogether, the nature of ETS domain transition from auto-inhibited to functional form and subsequent effect in the DNA distortion clearly depicts that, the relaxed conformation of native complex makes perfect interactions with its target gene and activates the transcriptional machinery. Notably, the absence of abovementioned criteria in mutated complexes makes them ineffective to perform gene expression. Hence, in this study, mutations in the p16^{INK4a} promoter region were used as the main criteria to understand the ETS1 mediated transcriptional machinery.

Additional Information

Supplementary data: The supplementary methods, supplementary figures S1 to S4 and supplementary table ST1 and ST2 are given as a supporting material.

ACKNOWLEDGMENT

References:

1. M. Ruas and G. Peters, *Biochim Biophys Acta*, 1998, **1378**, F115-177.
2. C. J. Sherr and J. M. Roberts, *Genes & development*, 1999, **13**, 1501-1512.
3. M. Esteller, P. G. Corn, S. B. Baylin and J. G. Herman, *Cancer research*, 2001, **61**, 3225-3229.
4. S. Gonzalez and M. Serrano, *Cell cycle*, 2006, **5**, 1382-1384.
5. S. Ortega, M. Malumbres and M. Barbacid, *Biochim Biophys Acta*, 2002, **1602**, 73-87.
6. M. Ruas, S. Brookes, N. Q. McDonald and G. Peters, *Oncogene*, 1999, **18**, 5423-5434.
7. N. Ohtani, Z. Zebedee, T. J. Huot, J. A. Stinson, M. Sugimoto, Y. Ohashi, A. D. Sharrocks, G. Peters and E. Hara, *Nature*, 2001, **409**, 1067-1070.
8. W. Zheng, H. Wang, L. Xue, Z. Zhang and T. Tong, *J Biol Chem*, 2004, **279**, 31524-31532.
9. J. Li, M. J. Poi and M.-D. Tsai, *Biochemistry*, 2011, **50**, 5566-5582.
10. J. Dittmer, *Molecular cancer*, 2003, **2**, 29.
11. T. L. Goetz, T. L. Gu, N. A. Speck and B. J. Graves, *Molecular and cellular biology*, 2000, **20**, 81-90.
12. C. W. Garvie, M. A. Pufall, B. J. Graves and C. Wolberger, *The Journal of biological chemistry*, 2002, **277**, 45529-45536.
13. M. A. Pufall and B. J. Graves, *Annual review of cell and developmental biology*, 2002, **18**, 421-462.
14. E. P. Lamber, L. Vanhille, L. C. Textor, G. S. Kachalova, M. H. Sieweke and M. Wilmanns, *The EMBO journal*, 2008, **27**, 2006-2017.
15. G. M. Lee, L. W. Donaldson, M. A. Pufall, H. S. Kang, I. Pot, B. J. Graves and L. P. McIntosh, *J Biol Chem*, 2005, **280**, 7088-7099.
16. J. M. Petersen, J. J. Skalicky, L. W. Donaldson, L. P. McIntosh, T. Alber and B. J. Graves, *Science*, 1995, **269**, 1866-1869.
17. A. D. Sharrocks, *Nature reviews. Molecular cell biology*, 2001, **2**, 827-837.
18. D. O. Cowley and B. J. Graves, *Genes & development*, 2000, **14**, 366-376.

19.   S. J. de Vries, A. D. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar and A. M. Bonvin, *Proteins*, 2007, **69**, 726-733.
20.   C. Dominguez, R. Boelens and A. M. Bonvin, *J Am Chem Soc*, 2003, **125**, 1731-1737.
21.   K. Muthu, M. Panneerselvam, M. Jayaraman, N. S. Topno, A. A. Das and K. Ramadas, *Journal of molecular modeling*, 2012, **18**, 4865-4884.
22.   K. Muthu, M. Panneerselvam, N. S. Topno, M. Jayaraman and K. Ramadas, *Cell Signal*, 2015, **27**, 739-755.
23.   M. Panneerselvam, K. Muthu, M. Jayaraman, U. Sridharan, P. Jenardhanan and K. Ramadas, *Molecular bioSystems*, 2013, **9**, 1470-1488.
24.   B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *Journal of Chemical Theory and Computation*, 2008, **4**, 435-447.
25.   D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J Comput Chem*, 2005, **26**, 1701-1718.
26.   K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950-1958.
27.   B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *Journal of Computational Chemistry*, 1997, **18**, 1463-1472.
28.   S. Miyamoto and P. A. Kollman, *Journal of Computational Chemistry*, 1992, **13**, 952-962.
29.   M. Kawata and U. Nagashima, *Chemical Physics Letters*, 2001, **340**, 165-172.
30.   R. Martoňák, A. Laio and M. Parrinello, *Physical Review Letters*, 2003, **90**, 075503.
31.   W. Liang, S. Miao, B. Zhang, S. He, C. Shou, P. Manivel, R. Krishna, Y. Chen and Y. E. Shi, *Oncogene*, 2014, **0**.
32.   S. Miao, K. Wu, B. Zhang, Z. Weng, M. Zhu, Y. Lu, R. Krishna and Y. E. Shi, *Molecular cancer therapeutics*, 2014, **13**, 699-713.
33.   K. Muthu, M. Panneerselvam, N. S. Topno, M. Jayaraman and K. Ramadas, *Cellular signalling*, 2014, DOI: 10.1016/j.cellsig.2014.11.036.
34.   A. Amadei, A. B. Linssen and H. J. Berendsen, *Proteins*, 1993, **17**, 412-425.
35.   B. Hess, *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 2000, **62**, 8438-8448.
36.   G. G. Maisuradze and D. M. Leitner, *Proteins*, 2007, **67**, 569-578.
37.   X. J. Lu and W. K. Olson, *Nucleic acids research*, 2003, **31**, 5108-5121.
38.   X. J. Lu and W. K. Olson, *Nature protocols*, 2008, **3**, 1213-1227.
39.   M. A. El Hassan and C. R. Calladine, *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 1997, **355**, 43-100.
40.   *The EMBO journal*, 1989, **8**, 1-4.
41.   M. A. El Hassan and C. R. Calladine, *J Mol Biol*, 1998, **282**, 331-343.
42.   R. Kumari, R. Kumar and A. Lynn, *Journal of Chemical Information and Modeling*, 2014, **54**, 1951-1962.
43.   E. Krissinel, *J Comput Chem*, 2010, **31**, 133-143.
44.   E. Krissinel and K. Henrick, *J Mol Biol*, 2007, **372**, 774-797.
45.   G. M. Lee, M. A. Pufall, C. A. Meeker, H. S. Kang, B. J. Graves and L. P. McIntosh, *Journal of molecular biology*, 2008, **382**, 1014-1030.
46.   C. W. Garvie, J. Hagman and C. Wolberger, *Molecular cell*, 2001, **8**, 1267-1276.
47.   M. A. Pufall and B. J. Graves, *Structure*, 2002, **10**, 11-14.
48.   A. Chakrabartty and R. L. Baldwin, *Advances in protein chemistry*, 1995, **46**, 141-176.
49.   S. Marqusee and R. L. Baldwin, *Proceedings of the National Academy of Sciences of the United States of America*, 1987, **84**, 8898-8902.

50.   H. Wang, L. P. McIntosh and B. J. Graves, *The Journal of biological chemistry*, 2002, **277**, 2225-
       2233.
51.   M. A. el Hassan and C. R. Calladine, *Journal of molecular biology*, 1995, **251**, 648-664.
52.   K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, *Nature*, 1997, **389**, 251-
       260.
53.   J. L. Kim, D. B. Nikolov and S. K. Burley, *Nature*, 1993, **365**, 520-527.
54.   H. C. Nelson, J. T. Finch, B. F. Luisi and A. Klug, *Nature*, 1987, **330**, 221-226.
55.   K. Swaminathan, P. Flynn, R. J. Reece and R. Marmorstein, *Nature structural biology*, 1997, **4**,
       751-759.
56.   M. D. Jonsen, J. M. Petersen, Q. P. Xu and B. J. Graves, *Molecular and cellular biology*, 1996, **16**,
       2065-2073.

Figure legends:

Figure 1: Molecular interaction of native and mutated p16^{INK4a} – ETS1 complexes, specifically the contribution of triplet residues flanking the major groove of p16^{INK4a} promoter in A) native p16^{INK4a} – ETS1, B) M1_ p16^{INK4a} – ETS1, C) M2_ p16^{INK4a} – ETS1 and D) M3_ p16^{INK4a} – ETS1 complexes. The hydrogen bonds are represented as dotted line whereas the special types of interaction (N-H…π, N…π and C=O…π) are depicted in solid lines. The color codes ETS1 protein: green, native p16^{INK4a} promoter: cyan, M1_ p16^{INK4a}: Cornflower, M2_ p16^{INK4a}: Purple and M3_ p16^{INK4a}: Gray color. All atoms of protein and DNA were colored according to heteroatom type. Cg: represents the centroid of aromatic ring.

Figure 2: The stability analysis of native and mutated p16^{INK4a} – ETS1 complexes from 50 ns molecular dynamics simulation. A) Describes the backbone RMSD plot of ETS1 protein B) Displays the RMS fluctuation of each residues of ETS1 protein. Here, the secondary structure of ETS1 is represented as boxes and arrows for helices and sheets, respectively.

Figure 3: The free energy landscape of (A) native_ p16^{INK4a}, (B) M1_ p16^{INK4a}, (C) M2_ p16^{INK4a} and (D) M3_ p16^{INK4a} – ETS1 complexes as a function of first two principal components whose cosine content is less than 0.2. Based on the most populated free energy minimum cluster, representative structures along with the helical orientation of ETS domain is shown. Color code ETS1: green, native p16^{INK4a} promoter region: cyan, M1_ p16^{INK4a}: Cornflower blue, M2_ p16^{INK4a}: Purple and M3_ p16^{INK4a}: Gray color.

Figure 4: The residue fluctuations, Transitions and Short Transitions of ETS1 protein upon complex with A) native_p16^{INK4a}, B) M1_ p16^{INK4a}, C) M2_ p16^{INK4a} and D) M3_ p16^{INK4a} promoter calculated using t-pad tool and plotted based on the residue and their PAD degree. The insets show the mode vector analysis of ETS domain calculated for the 50 ns MD simulation. The residue fluctuations are represented as starred connected lines, full transition and short transitions are shown in plus (red) and rectangle (magenta) symbols.

Figure 5: The structural transition of ETS1 in complex with native_p16^{INK4a} promoter depicting the transition in ETS domain and DNA bending. A) Describes the four types of mechanisms in HI1 helix folding towards HI2 and their influence in DNA bending. The docked (protein: plum and DNA: tan) and free energy representative structure (protein: green and DNA: cyan) were superimposed to explain the changes in the protein (orange arrow) and DNA (black arrow). Mechanism I) the hydrophobic interaction formed by P334 (green spheres), triangle basic patch interaction (blue spheres) and triplet residues contribution (orange spheres) with the interacting distance. Mechanism II) the series of hydrophobic interaction (yellow spheres) and hydrogen bond (green spheres) between H2 and H1 helix with their distance profile. Mechanism III) the hydrophobic interaction formed between H1 and HI2 helix and its contact distances. Mechanism IV) folding of HI1 towards HI2 and basic (blue spheres) to acidic patch (red spheres) interaction. (B) *Roll (ρ) - Twist (Ω)* plot, (C) *Slide (D_y) - Twist (Ω)* plot and (D) major (straight line) - minor groove (dotted line) width parameters explain the mode of DNA distortion. The reference (B and C) / docked (D) and free energy representative (B, C, D) DNA is shown in black and red, respectively. The shaded band (in C) shows the conformational channel responsible for naked DNA mode I distortion.

Figure 6: The structural transition of ETS1 in complex with M1_p16^{INK4a} promoter. A) Describes the four types of mechanism governing unfolding of HI1 helix. The superimposition of docked (protein: plum and DNA: tan) and free energy representative structure (protein: green and DNA: cornflower blue) explains the changes in the protein (orange arrow) and DNA (black arrow). Mechanism I) the gap formed between P334 and (green spheres) DNA, triangle basic patch interaction (blue spheres) and triplet residues contribution (orange spheres) are shown. Mechanism II) the less effective hydrophobic interaction (yellow spheres) and hydrogen bond (green spheres) formed between H2 and H1 helix are displayed. Mechanism III) the hydrophobic interaction formed between H1 and HI2 helix. Mechanism IV) unfolding of HI1. (B) *Roll (ρ) - Twist (Ω)* plot, (C) *Slide (D_y) - Twist (Ω)* plot and (D) major (straight line) - minor groove (dotted line) width parameters explain the mode of DNA distortion. The reference (B and C) / docked (D) and free energy representative (B, C, D) DNA is shown in black and red, respectively. The

shaded band (in C) shows the conformational channel responsible for naked DNA mode I distortion.

Figure 7: The molecular orientation of ETS domain (H1-H3 and β1-β4) in complex with native_ p16^{INK4a}, M1_p16^{INK4a}, M2_p16^{INK4a} and M3_p16^{INK4a} to define the triangle auto-inhibited (native) to oval functional form (M3) through a partial triangle (M1 and M2).

Table legends:

Table 1: The binding energy calculated using HADDOCK, solvation free energy interface calculated using PDBePISA and the triplet interacting residues of ETS1 with p16^{INK4a} in native and mutated complexes.

Table 2: The parameters, *Roll*, *Slide*, *Twist*, H-twist, minor and major groove width calculated using 3DNA tool for the selected single nucleotide base pair steps which has high *Roll* values in each free energy representative DNA structure are listed.

Table 3: The van der Waal, Electrostatic, polar solvation, SASA and Binding energy calculated using MM-PBSA tool specifically designed for gromacs MD package.

Table 1:

Bio-molecular complexes	N_P16^{INK4a} – ETS1	M1_P16^{INK4a} – ETS1	M2_P16^{INK4a} – ETS1	M3_P16^{INK4a} – ETS1
Haddock energies				
Internal energy complex	-6039.21	-1777.16	2015.14	823.708
Binding energy	-1310.86	-1163.33	-1267.39	-1183.25
PDBePISA interface analysis				
Interface area (Å^2)	1009.8	1063.2	1086.3	1009.8
Solvation free energy interface (Δ^{i}G, Kcal/mol)	-22.4	-17.6	-18.0	-18.6
Interface P-value of solvation free energy (Δ^{i}G)	0.526	0.707	0.711	0.639
Interaction of bio molecular complexes				
Interacting residues	ETS1	ETS1	ETS1	ETS1
	R391,R394,Y395,	R391,R394,Y395,	R391,R394,Y395	R391,R394,Y395
	N_p16^{INK4a}	M1_p16^{INK4a}	M2_p16^{INK4a}	M3_p16^{INK4a}
	DG9,DA10, DG12,DA13	DG8,DG9, DG12,DA11	DA11,DG12, DA13	DT9,DA10,DA11,DG12

31

Table 2:

Representative structure	Base pair step		Roll ρ	Slide D_y	Twist Ω	H-twist	Minor Width	Major Width	DNA form
N_P16^{INK4a} – ETS1	4	CG/CG	12.92	-0.35	31.78	34.33	13.6	20.7	B
	11	AG/CT	16.19	0.26	35.82	39.52	12.0	18.8	B
	Avg		*2.17*	*-0.30*	*30.58*	*30.57*			
	S.D		*6.43*	*1.12*	*10.89*	*13.57*			
M1_P16^{INK4a} – ETS1	3	CC/GG	16.16	-1.36	34.34	38.70	9.4	21.3	B
	6	CA/TG	8.80	0.66	11.92	15.00	15.6	18.6	B
	12	GA/TC	18.77	-1.44	31.43	33.11	14.5	22.8	-
	Avg		*5.71*	*-0.43*	*29.95*	*32.00*			
	S.D		*8.46*	*0.97*	*8.20*	*7.25*			
	5	GG/CC	11.06	0.09	35.71	37.82	14.0	19.4	B
	7	AT/AT	15.90	-0.66	33.07	37.31	14.3	17.6	B
	8	TT/AA	7.88	-0.98	17.54	23.48	13.5	19.3	-
	Avg		*7.40*	*-0.68*	*26.81*	*27.37*			
	S.D		*8.74*	*0.76*	*16.40*	*20.63*			
M3_P16^{INK4a} – ETS1	3	CC/GG	11.45	0.00	30.54	32.59	13.7	18.9	B
	5	CC/GG	11.20	-1.91	31.79	33.93	13.3	20.6	B
	9	TA/TA	14.50	-0.48	32.91	36.38	11.1	19.1	B
	12	GA/TC	16.06	0.72	18.70	24.50	13.0	20.2	-
	Avg		*5.93*	*-0.44*	*31.60*	*33.17*			
	S.D		*6.58*	*0.80*	*5.28*	*4.47*			

Table 3:

MM-PBSA Complex	Van der Waal energy (kJ/mol)	Electrostatic Energy (kJ/mol)	Polar solvation energy (kJ/mol)	SASA energy (kJ/mol)	Binding Energy (kJ/mol)
N_P16^{INK4a} – ETS1	-321.144 (+/- 26.742)	-11485.987 (+/- 255.442)	2528.693 (+/- 197.601)	-40.069 (+/- 3.029)	-9318.508 (+/- 188.129)
M1_P16^{INK4a} – ETS1	-299.532 (+/- 24.484)	-9077.569 (+/- 283.368)	1940.861 (+/- 152.459)	-34.548 (+/- 3.005)	-7470.787 (+/- 192.240)
M2_P16^{INK4a} – ETS1	-356.570 (+/- 42.566)	-11932.882 (+/- 398.009)	2476.279 (+/- 270.504)	-44.158 (+/- 4.773)	-9857.330 (+/- 243.786)
M3_P16^{INK4a} – ETS1	-419.500 (+/- 27.161)	-11730.629 (+/- 260.166)	2681.243 (+/- 195.870)	-46.904 (+/- 2.689)	-9515.791 (+/- 162.890)
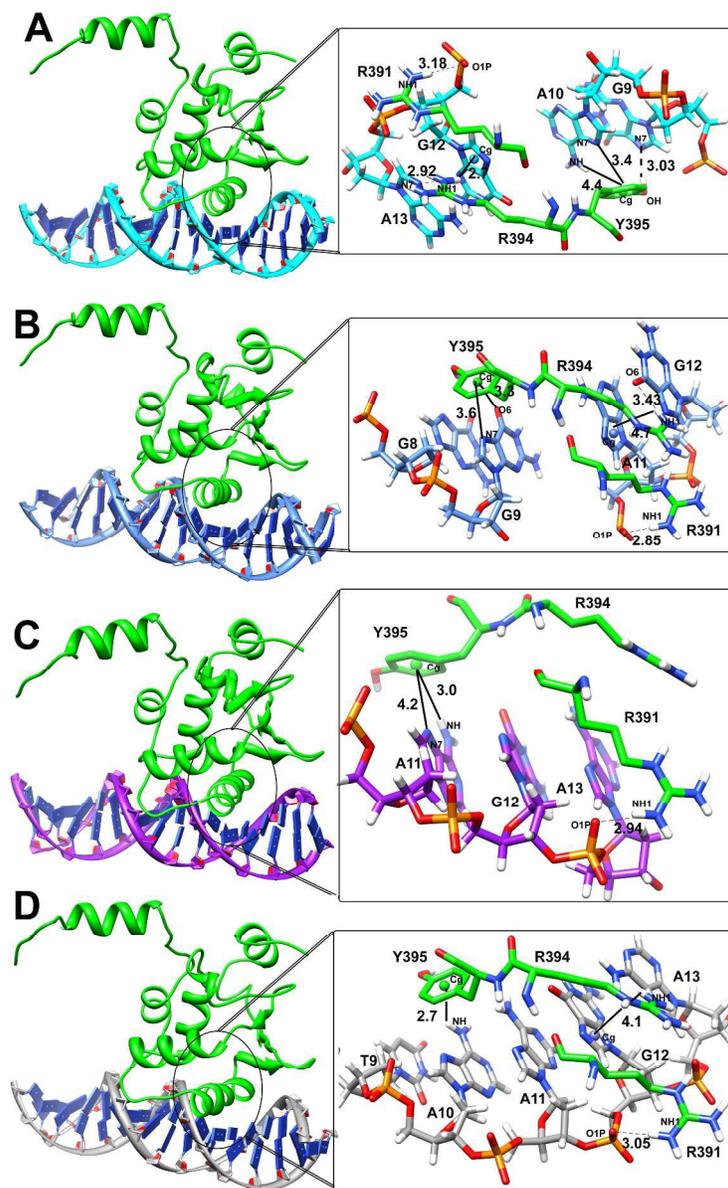
255x410mm (300 x 300 DPI)
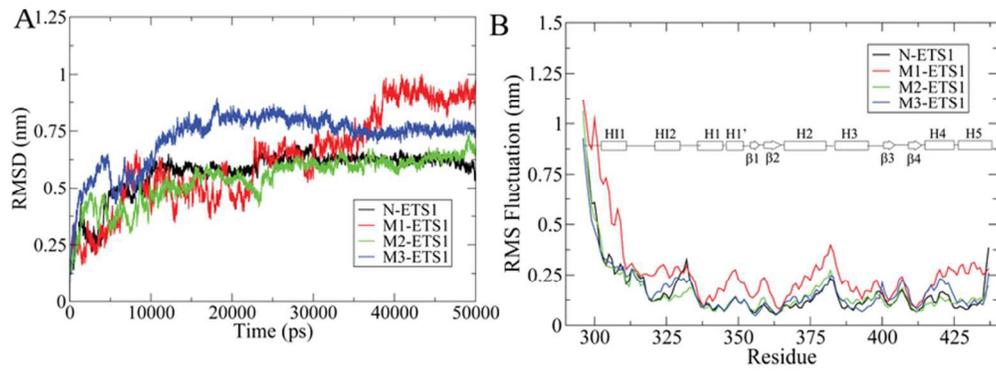
74x26mm (300 x 300 DPI)
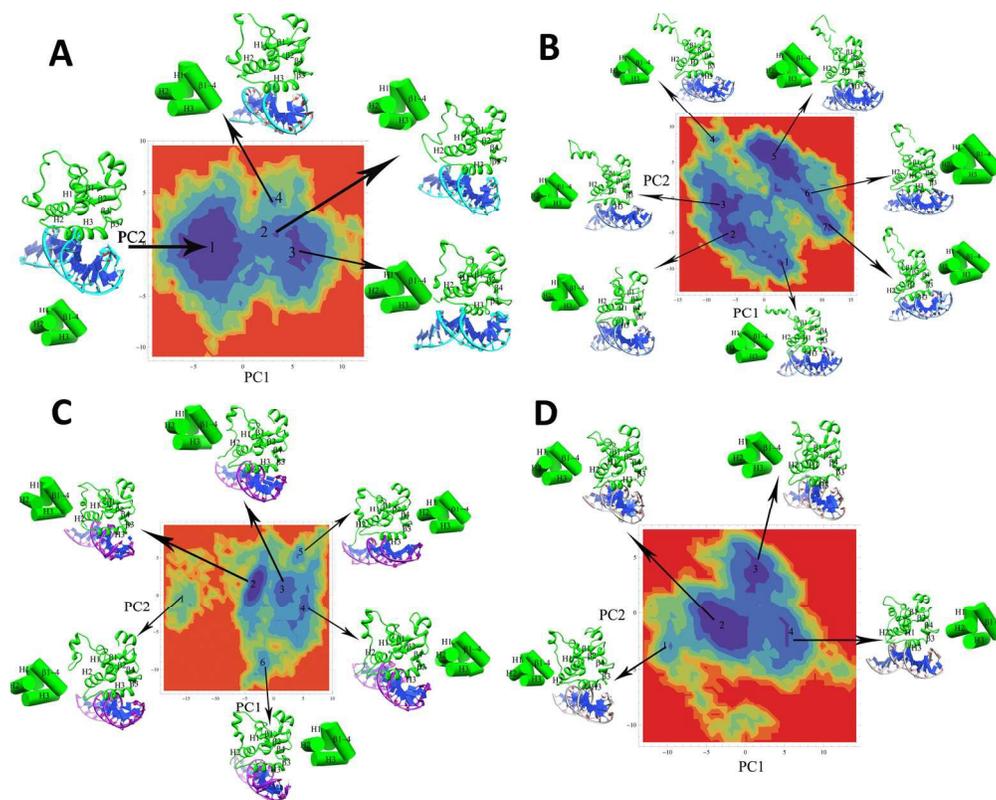
165x131mm (300 x 300 DPI)
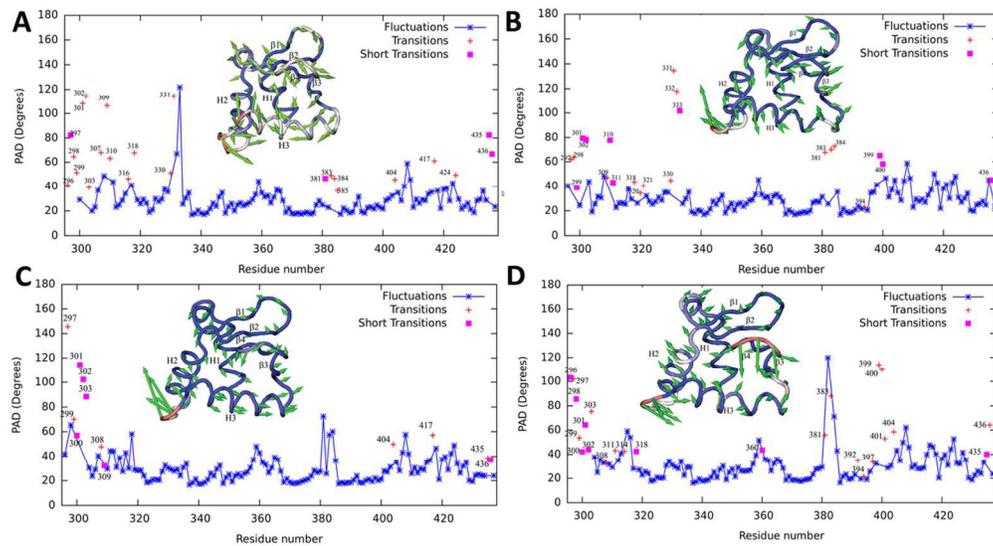
109x60mm (300 x 300 DPI)
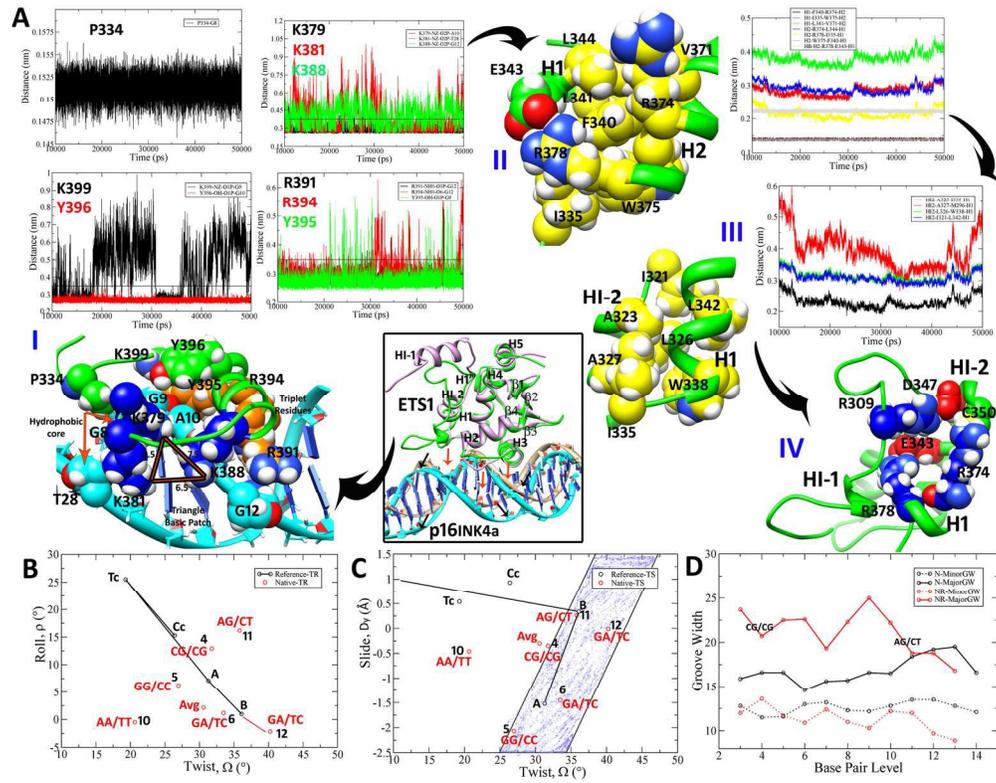
164x129mm (300 x 300 DPI)
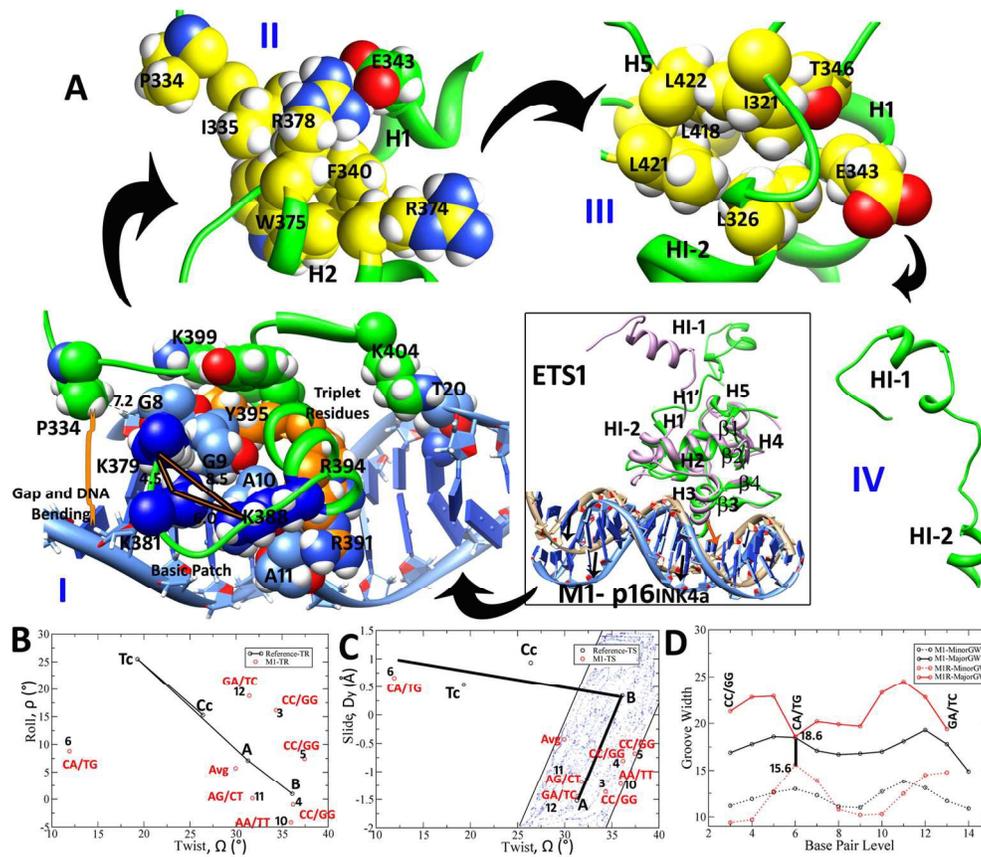
140x122mm (300 x 300 DPI)
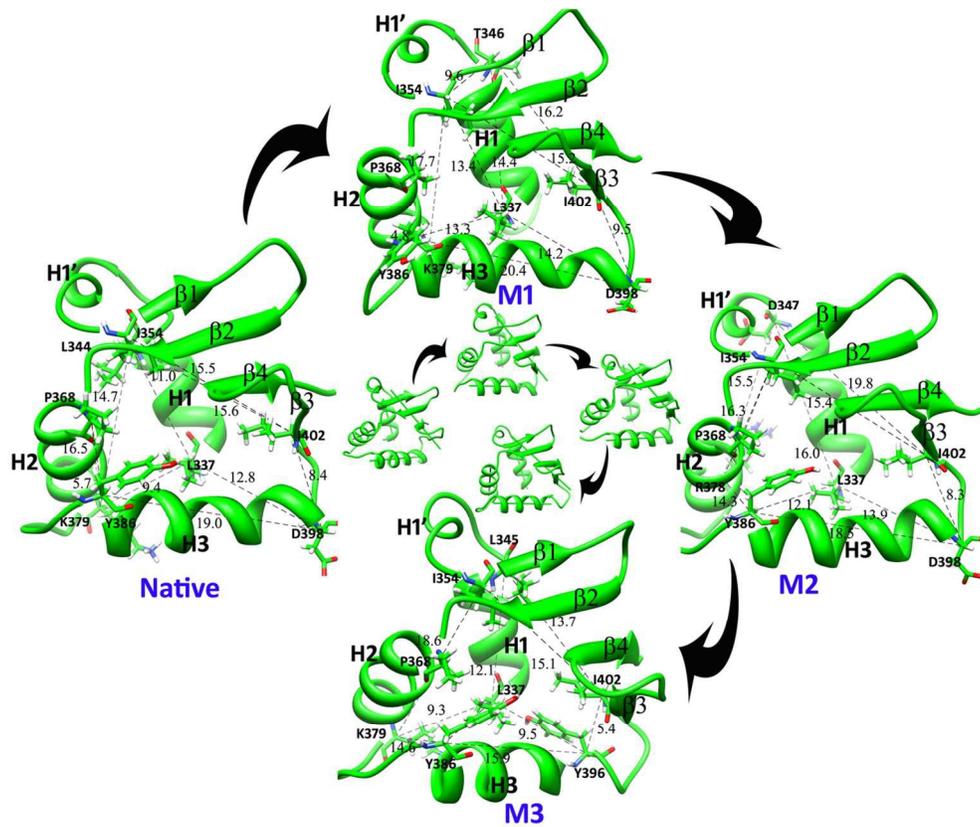
117x97mm (300 x 300 DPI)

58x33mm (300 x 300 DPI)