J Mol Model (2012) 18:4865–4884
DOI 10.1007/s00894-012-1489-x

ORIGINAL PAPER

Structural insights into interacting mechanism of ID1 protein with an antagonist ID1/3-PA7 and agonist ETS-1 in treatment of ovarian cancer: molecular docking and dynamics studies

Kannan Muthu · Manivel Panneerselvam · Muthukumaran Jayaraman · Nishith Saurav Topno · Arindam Atanu Das · Krishna Ramadas

Received: 29 August 2011 / Accepted: 28 May 2012 / Published online: 20 June 2012

Abstract Among the many abnormally expressed proteins in ovarian cancer, the prominent cancer in women, ID1 (inhibitors of DNA binding protein 1) is a potential one among other several targets. Interaction of ID1 with ETS-1 (transcriptional activator of $p16^{INK4a}$) suppresses the transcription of $p16^{INK4a}$ and causes abnormal cell proliferation. A peptide aptamer (ID1/3-PA7) has been designed to prevent this interaction and thereby leading to the transcription of $p16^{INK4a}$. However, the structural basis behind the molecular interaction of ID1 with ETS-1 (agonist) and ID1/3-PA7 (antagonist) is poorly understood. In order to understand this structural recognition and their interaction mechanism, in silico methods were used. From this interaction analysis, the residues of ETS-1 involved in interaction with the $p16^{INK4a}$ promoter were found to be targeted by ID1. Subsequently, ETS-1 binding residues of ID1 were found to be targeted by its aptamer- ID1/3-PA7. These results suggest that both ETS-1 and ID1/3-PA7 binds at the same region harbored by the residues-H97, D100, R103, D104, L107, A144, C145, D149, D150 and C154 of ID1. All these observations correlate with the experimental reports, suggesting that the identified residues might play a crucial role in promulgating the oncogenic effects of ID1. In silico alanine scanning mutagenesis also confirms the role of identified hot spot residues in $p16^{INK4a}$ regulation. Finally, the molecular dynamic simulation studies reveal the prolonged stability of the aforementioned interacting complexes. The obtained results throw light on the structure and residues of ID1 involved in transcriptional regulation of $p16^{INK4a}$.

Keywords ETS-1 · ID1 · ID1/3-PA7 · Molecular docking and simulation · Ovarian cancer · $p16^{INK4a}$

Introduction

Ovarian cancer is a highly lethal disease and leading cause for death among the women population [1]. Though, the molecular origin for the development of ovarian cancer is not clear, yet the previous studies showed the involvement of several potential targets in the regulation of tumor suppressor genes. Tumorous conditions result due to varied reasons, of which the transcriptional shut down of certain essential genes, are commonly observed in many cases. In the molecular events mediated in ovarian cancer, suppression of the $p16^{INK4a}$, a cyclin dependent kinase inhibitor 2A (CDKN2A), results in the cell cycle progression and associated abnormal cell proliferation. The transcriptional suppression of $p16^{INK4a}$ was achieved through the inhibition of ETS-1 and ETS-2 by ID1 [2].

ID1 belongs to the family of DNA binding inhibitor proteins (ID1-4), which were identified as the new members of basic helix-loop-helix (HLH) family of transcriptional regulators [3–7]. In general, HLH family proteins have a characteristic HLH dimerization domain, composed of amphipathic helices and a DNA binding domain [8–11]. Members of HLH family are grouped into two classes: class A bHLH, also known as E proteins (E12, E47, HEB, and E2-2) [12–17] and class B bHLH proteins (MyoD, myogenin,

Electronic supplementary material The online version of this article (doi:10.1007/s00894-012-1489-x) contains supplementary material, which is available to authorized users.

K. Muthu · M. Panneerselvam · M. Jayaraman · N. S. Topno · A. A. Das · K. Ramadas (✉)
Centre for Bioinformatics, School of Life sciences,
Pondicherry University,
Puducherry 605014, India
e-mail: ramadaskr@gmail.com

NeuroD/BETA2, MASH, HAND and TAL) [14, 18]. Heterodimerization of these HLH proteins with other transcriptional regulators are found to mediate cell proliferation and are essential for progression of cell lineages [19–21]. The HLH domain in these proteins is essential for heterodimerization and is followed by a DNA binding domain, for interaction with the promoter regions of DNA. ID proteins are found to lack this DNA binding domain [3], yet are found to heterodimerize with both members of HLH family and prevents them from association with their target genes. By doing so, ID proteins inhibit various cellular processes including apoptosis and also are found to inhibit cell differentiation. The heterodimerization of ID1 with its partners is harbored by the hydrophobic residues (M69, Y73, L76, V89, V92, I94, L95, V98, I99, Y101, I102 and L107) and positive charged residues (N70, K88 and K91) of HLH motif which are highly conserved in HLH protein family. The mutagenesis study on ID protein against E47 and MyoD shows that these residues play an important role in HLH motif stability and hetero-dimerization with other HLH motif proteins [22]. Moreover residues such as M69, N70, C72, Y73, S74, L76, I79, V80, P81, P84, S90, E93, I94, L95, Q96, V98, I99, D100, Y101, I102, D104, L105 and Q106 are found to be conserved among the ID protein family [23].

The heterodimerization of ID1 with the ETS-1, a member of E26 transcription factor family, found to switch off the expression of a tumor suppressor gene $p16^{INK4a}$ [2, 3, 24]. ETS-1 comprise ETS domain (331–415) with a winged helix-turn-helix (368–397) element, which is responsible to interact with promoter region of their respective target genes. The members of ETS family, binds with the specific promoter E-boxes (GGAA/T core motif) of their target genes [25] and regulates their transcription which in turn mediates various cellular events leading to the uncontrolled cellular growth. Similarly, ETS-1 binds with $p16^{INK4a}$ promoter region of -101 AGGAAG -106 and expresses the tumor suppressor protein $p16^{INK4a}$. By direct association with ETS-1 winged helix-turn-helix element, ID prevents it from binding with the E-box of $p16^{INK4a}$ promoter region and consequently inhibits the expression of $p16^{INK4a}$ [26]. This leads to the functional activation of CDKs and as a result, it propagates the cell progression and cell differentiation. Under abnormal expression of ID1, this function was found to result in tumorous conditions in the cell. The functional inactivation of ID1 was observed upon its association with a novel 20 amino acid containing peptide aptamer (ID1/3-PA7). This interaction has prevented ID1 from binding with ETS-1 and subsequently, the cell cycle arrest was observed to take place [27, 28]. ID1 protein is the most potent one among ID proteins, frequently found in many types of human cancer such as breast [29], cervical [30], head and neck [31], pancreas [32] and prostate cancer [33]. Ovarian cancer samples are found to express over 70 % of ID1 protein, which is associated with cancer cell proliferation and resistance against apoptosis through upregulation of EGFR [34]. All these observations have made ID1 a potential tumor marker and an attractive drug target.

The precise understanding of the structural mechanism behind the interactions involved by ID1 might lead to the design of specific inhibitors. The availability of a three-dimensional structure is essential for understanding the structural mechanism behind the interaction of ID1 protein, yet its structure was still unidentified. In order to investigate the structural mechanism of interaction between ID1 proteins with its partner ETS-1, we have predicted its three-dimensional structure using several in silico protocols and have performed in silico interaction analysis between them to identify the potential hot spot residues involved in their interactions. Initially, the in silico interaction between ETS-1 with the promoter region of $p16^{INK4a}$ was performed to identify the residues of ETS-1 involved in DNA binding, later through the docking studies between ID1 and ETS-1, the potential residues of ETS-1 were identified. Finally, the interaction studies between ID1 and its peptide aptamer was used to identify the potential hot spot residues of ID1. All these identified residues were then mutated to alanine and their active role in interactions were confirmed. The obtained results identify the potential residues of ID1 involved in interaction of ETS-1 and also provide details on the structural parameters essential for its interaction with its peptide aptamer, which might be used in the design of specific inhibitors in future.

Materials and methods

Tertiary structure prediction

The protein sequence of ID1 (155aa, UniProtKB accession number: P41134) was retrieved from UniProt (http://www.uniprot.org). A Protein BLAST search was carried out against the PDB database (PDB, http://www.rscb.org/pdb/) to detect the most suitable template for homology modeling. This result revealed the absence of homologous proteins of known atomic structure for the sequence of ID1. However, the HLH motif of ID1 showed 31 % sequence similarity with MyoD (PDB ID: 1MDY) atomic structure, which also comes under HLH motif family. In order to determine the N and C terminal regions of ID1, which contributes to the functional orientation of HLH motif for interacting with its interacting partner, the full length of the ID1 protein is modeled using I-TASSER server [35, 36]. In this method, the ID1 sequence is threaded along all available PDB structures to look for the possible folds by using four simple variants of PPA methods, such analysis showed two hits, E47 (PDB ID: 2QL2) and MyoD for HLH domain of ID1. Hence their structural coordinates were used for the prediction of HLH region (66–106), while the unaligned

regions (N-ter 1–50 and C-ter 116–155) of ID1 are built by ab initio method. Finally, the model is refined by replica-exchange Monte Carlo simulation. To verify the modeled ID1, the secondary structure was also predicted using PSIPRED [37] online tool and compared with three-dimensional structure of ID1 and analyzed (supplementary Fig. 1).

Peptide aptamer ID1/3-PA7 [27] was built using Discovery studio 3.0 module build and edit protein, in which build action is set to create and grow chain, and conformation is set to right-handed α-helix. The initial structure of the peptide is set to right-handed α-helix based on their secondary structural prediction using PSIPRED (Supplementary Fig. 2). Then the structure was minimized using CHARMm force field using electrostatics spherical cutoff and the smart minimizer algorithms with max steps of 200.

Molecular dynamics simulation (MDS) of modeled ID1 and aptamer ID1/3-PA7

The GROMACS 4.5.1 molecular dynamics package [38, 39] and Amber99sb-ILDN force field [40] was used to analyze the modeled ID1 and aptamer ID1/3-PA7 stability. The protein and peptide models were solvated with SPC216 water model that extend to 1.00 and 1.50 nm cubic box respectively from the molecule and the edge of the box. Periodic boundary conditions were applied in all direction. In both the systems the total charge is zero. The maximum of 50,000 energy minimization steps was carried out for the constructed protein and peptide models using a steepest descent algorithm with a tolerance of 1000 kJ mol^{-1} nm^{-1}. Consequently, 50,000 steps of a conjugate gradient algorithm are also used to minimize the protein and peptide models with a tolerance of 1000 kJ mol^{-1} nm^{-1}. For long-range interactions, the PME [41] method was used with a 1.2 nm cut-off and a Fourier spacing of 0.16 nm. The solvated and minimized system were considered a reasonable one in terms of geometry and solvent orientation and used for further simulation steps. All bond angles were constrained with LINCS algorithm [42], while the geometry of the water molecules was constrained with SETTLE algorithm [43]. V-rescale weak coupling method was used to regulate the temperature (310 K), while the Parrinello-Rahman method [44] was used to set the pressure (1 atm) of the system. The position restrains (PR) MD for both NVT (constant number of particles, volume and temperature) and NPT (constant number of particles, pressure and temperature) were carried out for 100 ps. The results of such a PR method show that the temperature, pressure, density and total energy of the system were well equilibrated. This pre-equilibrated system was later used in the 50000 ps (50 ns) production MDS with a time-step of 2 fs. Structural coordinates were saved every 2 ps and analyzed using the analytical tool in the GROMACS package. The above mentioned protocol is referred from our previous work with some modifications [45]. The computation 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 refined models were validated using the structural analysis and verification server (SAVES) which include PROCHECK [46] and ERRAT [47].

Molecular docking of ID1-ETS-1 and ID1-ID1/3-PA7 complex

In order to find the hotspot residues which are responsible for the agonist activity of ID1 against ETS1 and peptide aptamer ID1/3-PA7 against ID1, the protein-protein docking analysis was performed in High Ambiguity Driven Biomolecular Docking (HADDOCK v2.0) [48, 49] in combination with CNS v1.2. The ensemble docking was carried out by providing the structural ensemble of ID1 conformations taken at 2 ns interval from the equilibrated 10th–50th ns trajectory of ID1. Similarly, for ID1/3-PA7, the ensemble holding conformations taken at 2 ns interval of total 50 ns simulation was used as an input for performing ID1- ID1/3-PA7 docking studies. These ensemble conformations of ID1 and peptide aptamer ID1/3-PA7 from the MDS and ETS1 structure (PDB ID: 1GVJ) from RCSB were selected and prepared by using protein preparation Wizard of the Schrödinger 2010 suite (Schrödinger LLC, New York, USA) which was subsequently used for docking studies. Because of this, the HLH (66–106) motif of ID1, which is proven to interact with ETS1, the full length of peptide aptamer ID1/3-PA7 and ETS DNA binding site (331–415) of ETS1 [50, 51] were defined as an active site and their surface neighbors as passive residues. The standard and default protein-protein docking protocol comprises three stages, complex generation and orientational optimization, semi flexible docking and refinement in explicit solvent. The amino acids residues at the interface were considered as flexible segments and defined using the ambiguous interaction restraints (AIRs) ±2 sequential residues for flexible docking. This program employs 1000 independent rigid-body minimization leading to 1000 docked complexes. Based on their intermolecular energy (sum of the Van der Waals [VdW], electrostatic and AIR energy terms), the best of 200 were then subjected to semi-flexible simulated annealing (SA) in torsional space. Finally, each of these best structures obtained from the semi-flexible simulated annealing were refined in an explicit (TIP3) 8 Å water layer. The ensembles of docking conformation complexes were obtained using backbone root mean square deviation (RMSD) cut-off of 7.5 Å. Then, the best lowest intermolecular energy clusters were selected for manual analysis, in which the non-bonded (hydrogen bond and VdW contacts) intermolecular interaction was analyzed with DIMPLOT software, which is part of the LIGPLOT package

[52], using the default parameters, including the heavy-atom distance cut-off for non-bonded contacts: 3.9 Å; proton–acceptor and donor–acceptor distance cut-offs: 2.7 and 3.35 Å, respectively; minimum angle [D–H–A, H–A–AA, D–A–AA] for hydrogen bonds: 90°. All parameters used in the above mentioned protocol are referred from our previous work [45, 53].

Alanine scanning mutagenesis

In order to find the crucial role played by the hotspot residues of ID1 interacting with ETS1 and peptide aptamer ID1/3-PA7 the best structural complex of ID1-ETS1 and ID1-ID1/3-PA7 is submitted to DrugScore^{PPI} web server. This server automatically scans for the interface residues of given bio-molecular complexes. First it computes ΔG^{WT} (wild type) and mutates one of the interface residues to alanine then calculates the ΔG^{MUT} (mutant type) which allows subsequent calculation of $\Delta\Delta G$ (change in binding free energy) by subtracting the ΔG^{WT} from the ΔG^{MUT}. This procedure will be continued until all the interface residues $\Delta\Delta G$ are calculated [54]. From the results, the hot spot residues will be identified.

Molecular dynamics simulation of ID1-ETS-1 and ID1-ID1/3-PA7 complex

The best and lowest energy complexes of ID1-ETS1 and ID1-ID1/3-PA7 obtained from protein-protein interacting studies were subjected to molecular dynamics simulation using the earlier defined parameters for 50 ns (50,000 ps) production MD run. The lowest potential energy conformations of the complexes were selected from the 50 ns MDS trajectory. The analysis part was done using the programs build within the GROMACS package. The abovementioned MD protocol was used and all calculations were performed in a TESLA server.

Results and discussion

Molecular structure of ID1

ID1 has an HLH domain (66–106) which is conserved throughout the ID super family proteins (ID1 to ID4) and a motif for nuclear export signal (98–111). Identifying the potential candidate protein for homology modeling is not feasible in the case of full length structure predicition of ID1 protein, due to the absence of homologous proteins. However, the HLH motif of ID1 was found to have 31 % similarity with MyoD HLH protein [10]. In order to obtain the reliable full-length model of ID1, the I-TASSER fold recognition method was used.

Amidst the various obtained models, one with the TM-score of 0.37±0.12 (nearly equivalent to the expected score of >0.5, indicating the model of reliable topology) was selected for further analysis.

The predicted model of ID1 contains five helices α1 to α5 with four long coils connecting the helices. The regions of α-helices are α1 (3–12), α2 (28–41), α3 (54–79), α4 (91–115) and α5 (131–143) are shown in Fig. 1a. The PTLP motif at the beginning of a variable loop region (80–90), which separates the α3 and α4 initiates the dimerization specificity of ID1 [23, 55]. The phosphorylation of ID1 at Ser111 by casein kinase II is not found to affect the ability of the ID1 hetero-dimerization with other HLH factors [56]. Hence, in our study, we used un-phosphorylated ID1 for consequent analysis. In general, ID proteins are very short-lived (20–60 minutes depending on the cell type) and are stabilized by formation of dimers with other HLH factors [57, 58].

HLH motif

The three-dimensional structure of ID1 HLH motif is more significant for its dimerization capability with ETS-1 and ETS-2, which in turn inhibits the expression of p16^{INK4a} leading to ovarian cancer. Hence, the HLH motif of ID1 is predicted by using the structural information of HLH motif containing E47 and MyoD proteins through, I-TASSER modeling server. The predicted HLH motif of ID1 has about 0.638 and 0.719 Å RMSD with crystallographic structure of E47 and MyoD respectively (Fig. 1b).

The HLH motif of ID1 has two α helices (α3 and α4) connected by a short loop (80–90). In general, the sequence similarity searches reveal the presence of HLH motif in the region 66–106 of ID1, whereas, in ID1, the helices were observed to be formed (54–115) even beyond these regions. Hence, this region (54–115) was termed as the extended helical region and this observation was confirmed through experimental analysis. The helical nature of α4 is enhanced by the presence of the loop and α3 which is proven by experimental analysis [23]. In accordance to that, three-dimensional structure of α4 is stabilized by forming three hydrogen bonds and VdW interactions with α3 and variable loop region. Accordingly, the K91, E93 and H97 of α4 make hydrogen bond interactions with N70 of α3 and S90 and T82 of a variable loop region with a distance of 3.12, 2.96 and 2.79 Å respectively. In addition to that, the K91 CD of α4 makes C-H-π interaction with phenyl ring (Cg: CD1, CE1, CZ, CE2, CD2 and CG) of Y73 with a distance of 3.31 Å also found to be stabilized α4 helical segment. Moreover, the residues M69, Y73, L76, K77, V80, T82, V89 and S90 of α3 and variable loop region makes strong VdW interactions with K91, E93, I94, L95, H97, V98 and Y101 of α4. All these interactions play a significant role in

the stabilization of α4 helical segment, which is described in Fig. 1c.

N-terminal α1

The N-terminal α1 is formed by the residues (Nter-VASG-STATAA-Cter) V3 through A12 and have three turns in length. In contrast with the secondary structure prediction by PSIRPED (supplementary Fig. 1), this region adopts an alpha helical conformation in three-dimensional space by using necessary classical hydrogen bonding pattern. The hydrogen bonding, VdW and hydrophobic forces formed within the alpha helices and between the helices, contribute to the protein stability and act as a principal determinant of protein conformation. Accordingly, α1 hydrophobic residue V3 makes C-H-O interaction with D68 of α3 by a contact distance of 3.52 Å, which is referred to as soft-acid…soft-base hydrogen bond [59] and expected to contribute significantly to the stabilization of α1. Moreover, the residues V3, S5, A9 and T10 of α1 make strong VdW interaction with α3 (D68 and R75) and loop region, which connects the α1 and α2 (K23 and T24) also stabilizes the α1 helical nature (Fig. 1d).

N-terminal α2 and C-terminal α5

The N-terminal α2 contains a major (Nter-AGEVVRCLS-Cter) and minor (Nter-EQSVA-Cter) helical segment and C-terminal α5 (Nter-TLNGEISALTAEA-Cter) have two to three turns in length. The minor helical segment of

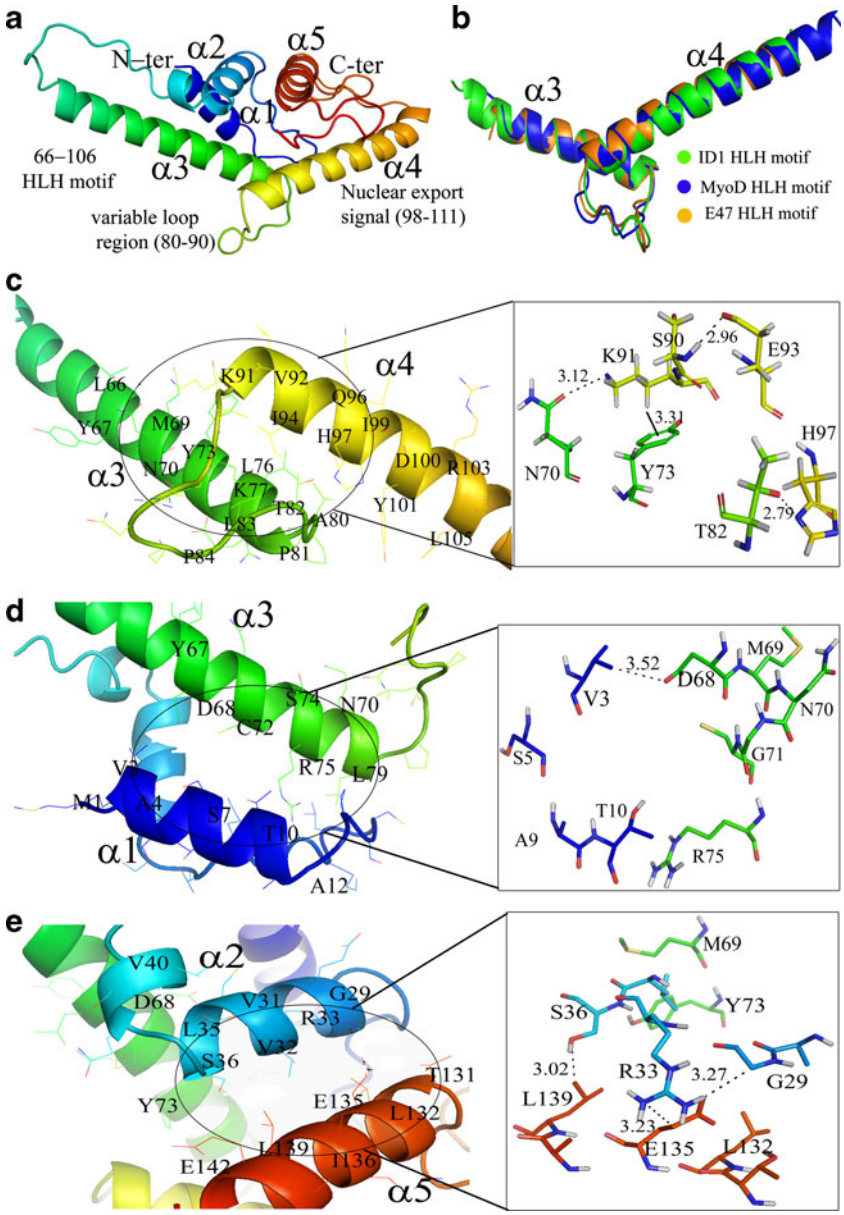


Fig. 1 The predicted three dimensional structure of ID1. (**a**) ID1 structure with five helixes named α1 to α5, which contains HLH motif by the residues of 66–106 (half of α3 and α 4 and variable loop region), variable loop region (80–90) responsible for dimerization and a nuclear export signal (98–111). (**b**) The superimposition of ID1 HLH motif with MyoD and E47 HLH motif, which shows less RMSD deviation in the HLH motif (color codes: green-ID1, blue-MyoD and orange-E47). (**c**) Displays intra molecular interaction of α4 with α3 and variable loop region, which explains the stability of the α4 in three-dimensional space. (**d**) Folding nature of α1 seized by making interaction with α3 and loop connecting α1 and α2. (**e**) The α2 and α5 have strong hydrophobic interaction with each other and retain their helicity in three-dimensional space

α2 is not stabilized throughout the 50 ns MD simulation studies, accordingly this segment is ignored for the further helicity analysis. The helical nature of α5 is in good agreement with the results of secondary structure prediction (supplementary Fig. 1) done by PSIPRED. In the case of α2, only QSVA region forms a helix in secondary structure prediction, whereas in, I-TASSER modeled structure of α2 helices formed by the residues of A28-S36 described in Fig. 1e. This helical nature of α2 is stabilized by making strong interaction with parallel α5 and adjacent α3. The G29 of α2 makes C-H…O interaction with E135 of α5 by a contact distance of 3.27 Å and in another hand, L139 CD2 of α5 makes C-H…O interaction with OG S36 of α2 (3.02 Å). In addition, the CG group of E135 makes C-H…N interaction with R33 with a distance of 3.22 Å. All these soft-acid…soft-base interactions formed through dispersion and polarization energy [59] rose between the α2 and α5, which contributes to the stabilization of each other. Moreover, the hydrophobic patch formed by V31, V32 and L35 in α2 makes strong hydrophobic and VdW interaction with α3 and α5. In view of that, the L35 packed with Y73 of α3, like, V32 also makes hydrophobic contacts with L132 of α5. Then, the hydrophobic V31 and M69 make VdW interaction with D68 and E38 respectively. All these interactions contribute for the stabilization of α2 and α5 helicity.

Molecular structure of ETS1

The crystallographic structure of ETS protein (PDB ID: 1GVJ) (Fig. 2a) was used to study the interacting mechanism with p16^{INK4a} promoter region and ID1. The structure of ETS1 contains two regions (280–330 and 416–441) that flank DNA-binding ETS domain (331–415) and inhibit DNA binding. These two flanking regions are also called two inhibitory α-helices and reside in the N-terminal (HI-1 and HI-2) and C-terminal (H4 and H5) region respectively. The N-terminal inhibitory region HI-1(304–314) unfolds when the ETS-1 binds to DNA, which is proven by protease sensitivity and circular dichroism spectroscopy studies [50, 51], and HI-2 (323–330) retains its structure, but the orientation alters from inhibitory ETS-1 to active form. The C-terminal inhibitory region contains a short H4 (418–422) followed by a long H5 (426–433) helices pack against N-terminal inhibitory region and vice versa and also pack against with H1 of ETS domain, which collectively forms an inhibitory module of ETS-1. The auto-inhibition of ETS-1 might be relieved by interaction with heterologous transcription factors. The ETS domain of ETS-1 has four α-helices (H1, H1', H2 and H3) and β-sheets (B1, B2, B3 and B4) which holds a winged helix-turn-helix element responsible for DNA binding [50].

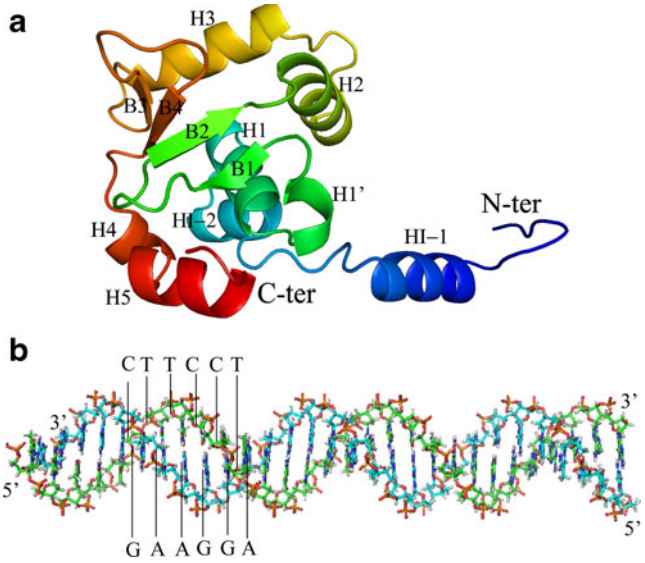


Fig. 2 Three dimensional structures of ETS-1 and p16^{INK4a}. (**a**) ETS-1 structure contains ETS DNA binding domain (331–415 contains H1, H1', H2, H3, B1, B2, B3 and B4) with winged HLH motif (H2, loop and H3) and flanked by N (H4 and H5) and C-terminal (HI-1 and HI-2) auto inhibitory region. (**b**) The structure of p16^{INK4a} promoter as DNA duplex, which contains 5'-AGGAAG-3' motif responsible for interaction with ETS-1 protein

Structure of p16 ^{INK4a}

The promoter sequence of p16 ^{INK4a} was referred [2], and the three-dimensional structure is modeled using Discovery studio using *build and edit nucleic acid modules* using nucleic acid type: DNA-duplex and conformation: B-helix. The mutation introduced in the sequence of -101 to -106 (AGGAAG mutated GG to TT) overlapping putative ETS-binding site in the human p16 ^{INK4a} promoter showed a reduced response to ETS2. Hence, the region -101 to -106 is the essential promoter region for the binding of ETS1 and ETS2. Accordingly, the promoter sequence was built and used for consequent analysis to understand the interacting mechanism with ETS-1. The three-dimensional structure of p16 ^{INK4a} promoter sequence is displayed in Fig. 2b with the notation of 5' and 3' end.

Stability of the modeled ID1 and aptamer peptide ID1/3-PA7 by MDS

To assess the quality of the modeled ID1 and peptide aptamer ID1/3-PA7, molecular dynamics calculation was carried out for 50 ns. The parameters, energy minimization, production MD and validation results are listed in Table 1, which shows that the steepest descent energy minimization of ID1 and ID1/3-PA7 is converged in the steps of 1124 and 759 respectively. The minimum and maximum potential energies of ID1 and peptide ID1/3-PA7 illustrates that the

Table 1 The results of molecular dynamics simulation and structural validation. The parameters, energy minimization steps, potential energy, RMSD profile, Ramachandran plot and ERRAT scores are listed for both modeled ID1 and ID1/3-PA7

Protein and peptide	ID1	ID1/3-PA7
Parameters and energy minimization		
Grid cell (nm)	20×20×20	16×16×16
Number of SOL molecules	21734	10621
Volume (nm3)	679.337	323.267
Steps of steepest descents EM converged (Fmax <1000)	1124	759
Conjugate gradients: potential energy	-1.1900560e+06	-5.6824750e+05
Molecular dynamics simulation results of 50 ns		
Potential energy (kJ/mol)		
Maximum potential energy	-1072753.87	-488697.87
Minimum potential energy	-1107915.25	-494860.90
Equilibration period (ps)	~10000 ps	~ 2000 ps
RMSD (nm)	0.50	0.65
Validation of lowest energy conformer: ID1 and ID1/3-PA7		
Ramachandran plot assessment		
Residues in favored region (%	84.7	70.6
Residues in additionally allowed region (%)	14.5	29.4
Residues in generously allowed region (%)	0.0	0.0
Residues in disallowed region (%)	0.8	0.0
ERRAT: overall quality factor	88.43	100.00

models were energetically stable during the production MD simulation.

Helical nature of ID1 protein

In order to evaluate the helical nature of the I-TASSER modeled ID1, the φ and ψ values are retrieved from the 50 ns MD simulation for each helix and plotted in Fig. 3a to e. In general, protein residues are classified as helical when the backbone dihedral angles φ and ψ lie in a constricted region of the Ramachandran plot, with φ between -95º and -35º, and ψ between -15º and -70º. The sum of φ and ψ values of helices also determines the type of helicity, when the sum of φ and ψ angles on adjacent residues is ~-105º adopts alpha-helical nature. For π-helix and 3_{10}-helix, the sum of above mentioned angles is -75º and -130º respectively [60, 61]. Accordingly, the helical α1, α2, α3, α4 and α5 of ID1 lies in the generally allowed alpha-helix region of Ramachandran diagram with the distribution of φ=[-110, -35º] and ψ=[-70, 10º]. Figure 3a to e shows the majority of each helical φ and ψ trajectory lies in the normal range mentioned above. Then, the sum of the φ and ψ angles of each helix through 50 ns simulation were also calculated. The average sum values of φ and ψ for each alpha-helix α1, α2, α3, α4 and α5 have -108.34, -98.94, -103.67, -104.96 and -105.59º respectively, which is closer to the value of an alpha-helicity and confirmed through 50 ns simulation. In addition to that, the residues of each helix of ID1 percentage propensity to retain their helicity through 50 ns simulation is calculated and is given in Fig. 4a, c, e, g and i for α1, α2, α3, α4 and α5 respectively. Then, the backbone hydrogen bond profile includes acceptor C = O to H-N donor pairings for each helix, which is the determinant of the alpha-helices that are also calculated through 50 ns MD simulation. The normal hydrogen bond distance for an alpha helicity requires a minimum donor and acceptor distance of 3.5 Å, which is described in Fig. 4b, d, f, h and j for α1, α2, α3, α4 and α5 respectively. The inset picture of Fig. 4b, d, f, h and j shows superimpose of α1, α2, α3, α4 and α5 at each 2 ns MD simulation from starting conformation at 10 ns up to 50 ns (totally 20 conformation) respectively. From the residues percentage propensity, it becomes evident that the residues V3, A4, G6, S7, T8 and A9 of α1, maintain their helical nature throughout their time scale, and their corresponding hydrogen bond profile too, adds to the validation of their structural propensity to form α-helix (Fig. 4a and b). Meanwhile, the helical percentage propensity and the average classical hydrogen bonding pattern (n-n+4) of α2 (major segment A28-S36) shows V31, V32 and R33 only stabilize their helical nature whereas the hydrogen bonding pattern shows that it varies from 3.5 to 4.5 Å and gets stabilized after 40 ns simulation (Fig. 4c and d). In order to understand this deviation, we have prepared an individual backbone hydrogen bonding profile for each n-n+4 residue (A28-V32, G29-R33, E30-C34, V31-L35, V32-S36 and R33-E37) and they are shown in supplementary Fig. 3. This indicates that, the first three n-n+4 hydrogen bonding interactions formed between A28-V32, G29-R33 and E30-C34 fluctuate more in their hydrogen bonding profile, this may be the reason that average n-n+4

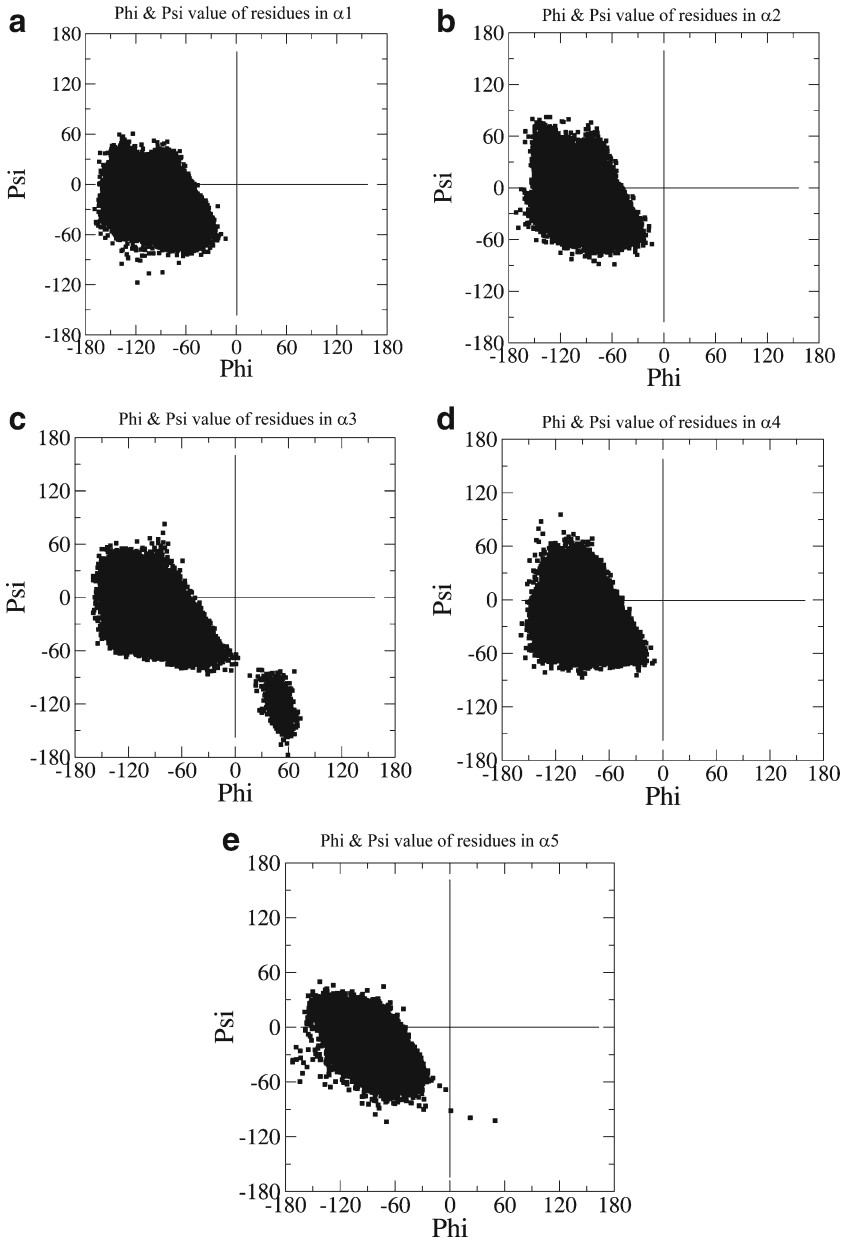


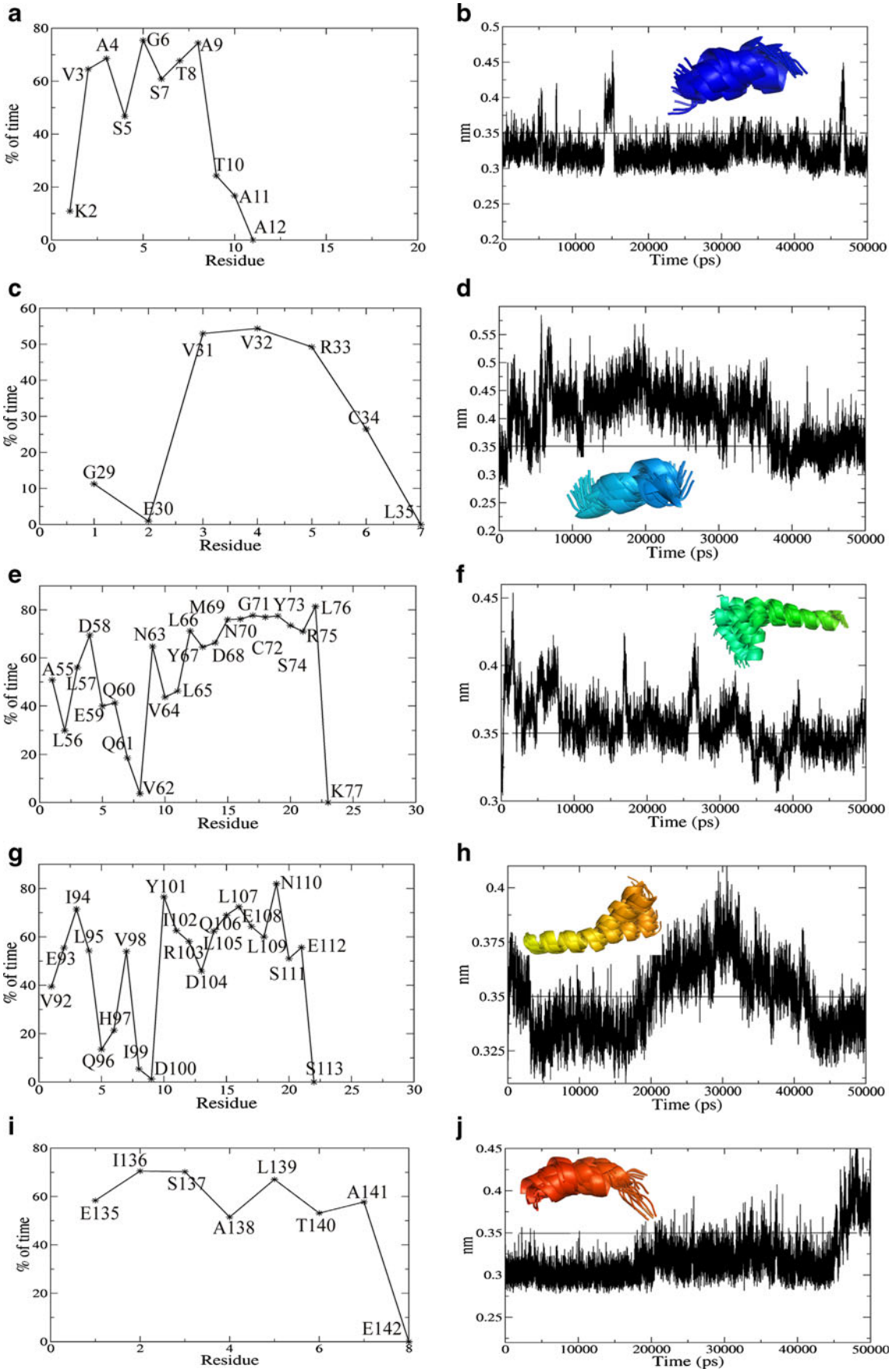
Fig. 3 The φ and ψ values of each helix throughout 50 ns MD simulation. The φ and ψ values of alpha (**a**) helix- α1, (**b**) helix- α2, (**c**) helix- α3, (**d**) helix- α4 and (**e**) helix- α5. The plot explains that, all the helices are within the constricted region of Ramachandran plot

classical hydrogen bonding pattern are shown as less effective. Nevertheless, the major segment of α2 is stabilized in their backbone hydrogen pattern through V31-L35, V32-S36 and R33-E37 and retain their helicity with these residues. Moreover, the helical nature of α3, α4 and α5 (α3: A55-L76; α4: E93- N110; α5: N133- A141) are found to be maintained throughout the time scale. Furthermore, the residues maintain per residue helical propensity values for more than half period of the simulation time (Fig. 4e–j). These results reveal the reliability of the modeled ID1 structure.

Structure of peptide aptamer ID1/3-PA7

The small peptide aptamer ID1/3-PA7 of 20aa length can in general adopt any possible conformations in the actual cellular conditions. However, upon interaction with its partner, it has to adopt a permanent three-dimensional conformation, which can even be considered as functional conformation of the peptide. However, in respect to the in silico analysis of exploration of such small peptides through molecular dynamic studies, the requirement for the starting 3D structure can not be ruled out. So by any means, the starting conformation for peptides has to be generated.

Fig. 4 The residue percentage propensity and average n-n+4 hydrogen bond length for alpha helix of ID1 protein calculated through 50 ns MD simulation. The helicity per residues for ID1 of (**a**) alpha helix- α1, (**c**) alpha helix- α2, (**e**) alpha helix- α3, (**g**) alpha helix- α4 and (**i**) alpha helix- α5. The average n-n+4 hydrogen bond length for ID1 of (**b**) alpha helix- α1, (**d**) alpha helix- α2, (**f**) alpha helix- α3, h) alpha helix- α4 and (**j**) alpha helix- α5. The inset picture shows superimposition of helices at each 2 ns MD simulation up to 50 ns for corresponding picture


a
% of time
Residue
K2
V3
A4
S5
G6
S7
T8
A9
T10
A11
A12
b
nm
Time (ps)
c
% of time
Residue
G29
E30
V31
V32
R33
C34
L35
d
nm
Time (ps)
e
% of time
Residue
A55
L56
L57
D58
E59
Q60
Q61
N63
V64
L65
L66
Y67
D68
M69
N70
G71
C72
Y73
S74
L76
R75
V62
K77
f
nm
Time (ps)
g
% of time
Residue
V92
E93
I94
L95
Q96
H97
V98
I99
D100
Y101
I102
R103
D104
Q106
L105
L107
E108
L109
N110
S111
E112
S113
h
nm
Time (ps)
i
% of time
Residue
E135
I136
S137
A138
L139
T140
A141
E142
j
nm
Time (ps)

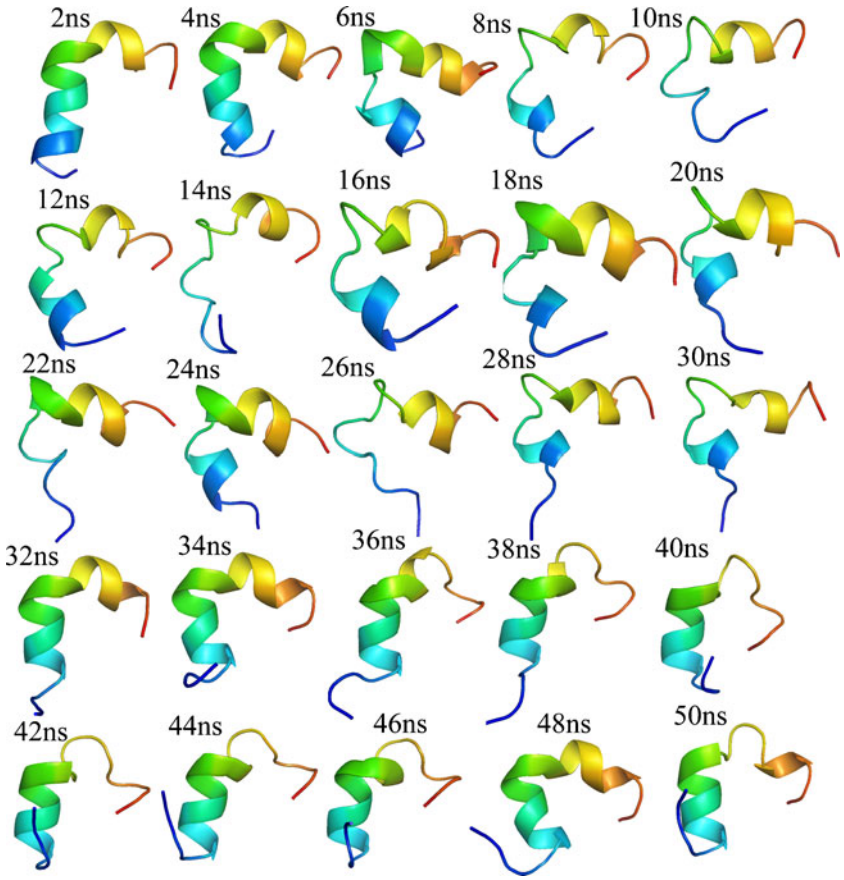


Fig. 5 Structural folding of peptide aptamer ID1/3-PA7 at each 2 ns interval up to 50 ns MD simulation

Amidst all available options, the idea of using the secondary structure of the concerned peptide might prove useful and reliable in predicting the starting conformation of small peptides.

Accordingly, the structure of peptide aptamer ID1/3-PA7 was predicted to form alpha helix based on their secondary structural prediction made by PSIPRED (supplementary Fig. 2) and was modeled using the build module of Discovery Studio. In order to explore the possible conformations, the 50 ns MD simulation was carried out. The results indicate the transition of helices (7–16) to coils in various time scales and toward the end of the simulations, the transition in reverse direction from coil to helices was also observed (Fig. 5). The distortions and loss of helices are caused due to the changes occurred in the backbone hydrogen bonding pattern in the respective regions, as implied by the free rotations in their φ and ψ angles. Obtained results indicate the wide possibility for the peptide aptamer to adopt any of the random coils to structured helical conformations in the given cellular conditions.

RMSD profile and evaluation results

Figure 6a shows that the backbone RMSD profile of ID1 and ID1/3-PA7 attains equilibration after the 10,000 ps and 2000 ps and remains quite stable up to final production MD with the maximum RMSD of 0.50 and 0.65 nm respectively. The lowest potential energy conformation of ID1 and ID1/3-PA7 was selected and retrieved for the validation analysis, which confirms that these models had good quality factors and were reliable for subsequent analysis (Table 1). The RMS fluctuation of ID1 during 50 ns MD simulation also calculated and described in Fig. 6b, showed ID1 protein has four flexible regions. The extended helical region of α3 (53–56) and loop connecting α2 and α3 (49–52) are more flexible, and the RMSD deviates from 0.4 to 0.75 nm and is considered as I flexible region in ID1. The second region (II) 83–86 (variable loop region) has RMS fluctuation up to 0.38 nm, which also includes PTLP motif of L83 and P84. The extended α4 helix region and loop connecting α4 and α5 are the third flexible region (III) with RMSF ranges from 0.40 to 0.45 nm. The C-terminal loop region of α5 (IV) also fluctuates up to 0.43 nm.

Essential dynamics (ED)

The principal component analysis (PCA) for ID1 and ID1/3-PA7 was carried out to understand the conformity of simulation period and structural RMSD equilibration for their stability of the protein. For covariance matrix generation, the trajectory of 25,001 frames was used with covariance matrix

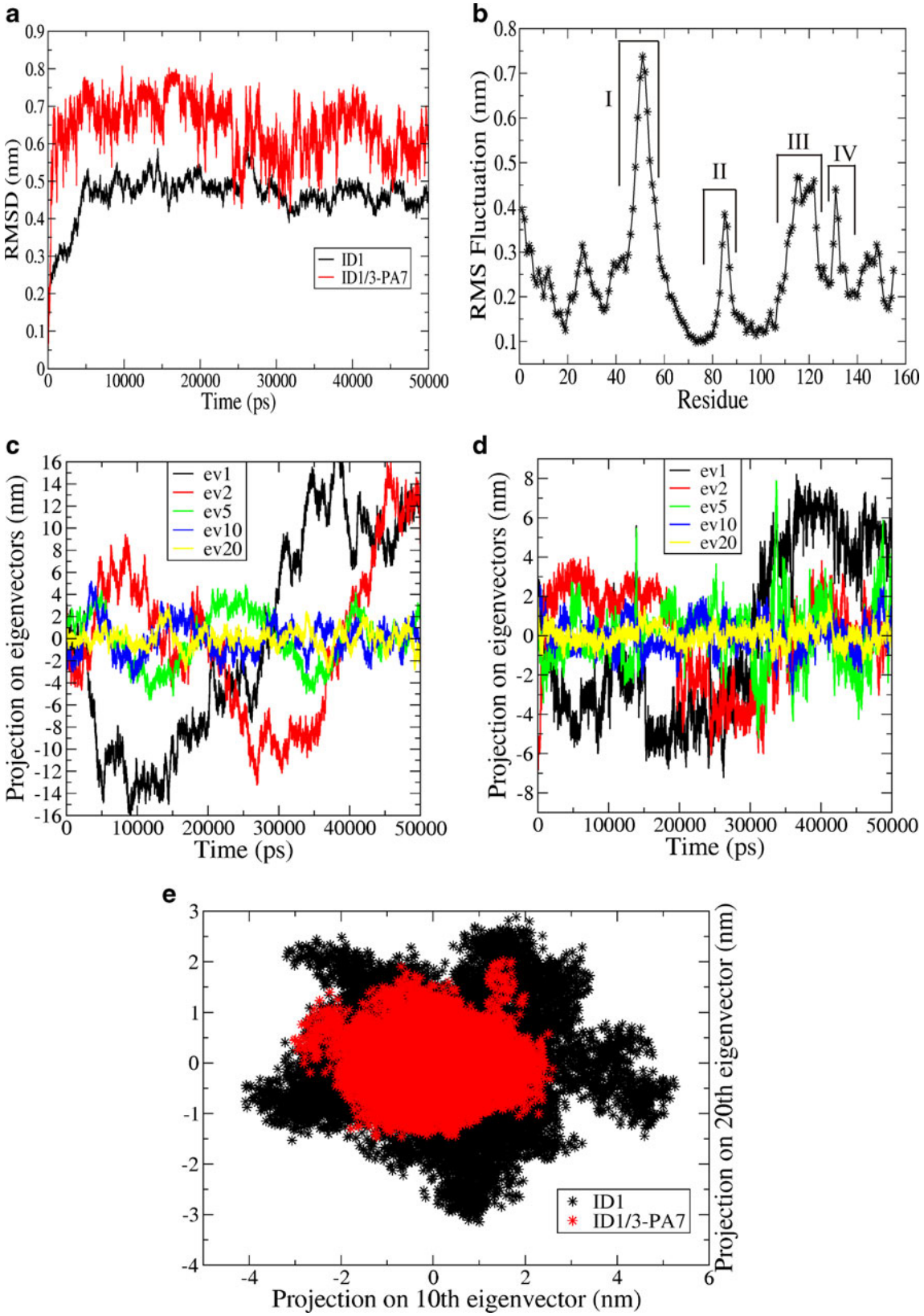


Fig. 6 The results of molecular dynamics simulation and essential dynamics. (**a**) The backbone RMSD of ID1 and ID1/3-PA7 for 50 ns MD simulation. (**b**) The backbone RMS fluctuation of ID1, which clearly indicates four region of ID1 fluctuate more. Motion and projection of trajectory by ED analysis are shown in (**c**) and (**d**), in which the motions along with the first two, fifth, tenth and 20th eigenvectors obtained from the protein coordinate covariance matrix for ID1 and ID1/3-PA7 respectively. (**e**) The projections of the trajectory onto the planes defined by the tenth and 20th eigenvectors from the protein coordinate covariance matrix for ID1 (black) and ID1/3-PA7 (red)

dimensions of 6849 for ID1 and 921 for ID1/3-PA7. These dimensions were given by 2283 of ID1 and 307 of ID1/3-PA7 protein elements with sum of the eigenvalues 234.228 and 48.3541 nm^2, respectively. The steep eigenvalue curve was obtained by plotting the eigenvalues against the eigenvectors of each ID1 and ID1/3-PA7 conformation, and it was noted that around 90 % of the protein motion is enclosed by the first 20 eigenvectors. From these results, we observed that, most of the internal motion of the ID1 and ID1/3-PA7 is limited to a subspace with very small dimensions. In accordance to the detailed analysis of the motion along with direction of eigenvector was gained by projecting the trajectory onto these individual eigenvectors. The first three, fifth, tenth and the 20th projections of the protein trajectory onto the eigenvectors obtained from the protein covariance matrix and plotted against time for each of ID1 and ID1/3-PA7 during the 50 ns simulation (Fig. 6c and d). Result of such analysis specifies that, the motions of the protein attain their equilibrium fluctuations in the first ten eigenvectors. Figure 6e demonstrates the trajectory projected onto the planes defined by two eigenvectors (the tenth and twentieth) from the protein coordinate covariance matrix for ID1 and ID1/3-PA7, which are strongly correlated and fill the expected ranges almost completely. These results point out that the sets are comparable and there is no high projection noticed far from the diagonal. From the results of RMSD profile and PCA analysis signifies that, the modeled structures have extended stability in the 50 ns MD simulation and sequentially favors the selection of these structures for further analysis.

Molecular docking analysis

In order to determine the antagonist activity of ID1 with ETS-1 and ID1/3-PA7, the HADDOCK docking simulation was performed to attain the structural insights into these complexes. The transcriptional regulatory role of ETS-1

Table 2 The HADDOCK docking results of ID1-ETS-1, ID1-ID1/3-PA7 and ETS-1-$p16^{INK4a}$ complexes

Bio-molecular Complexes	ID1 and ETS-1	ID1 and ID1/3-PA7	ETS-1 and $p16^{INK4a}$
Haddock scores and energies			
Internal energy complex (kcal mol^{-1})	-11079.3	-6160.3	-3023.58
Binding energy (kcal mol^{-1})	-12711.1	-10168.4	-11964
E_{vdw} (kcal mol^{-1})[A]	-67.78	-47.136	-64.74
E_{ele} (kcal mol^{-1})[A]	-664.15	-302.08	-671.38
Buried surface area ($Å^2$)	2188.97	1528.53	1797.43
Haddock analysis of best bio-molecular complexes			
Interacting residues	ID1	ID1	ETS-1
	T82, V92, E93, Q96, H97, D100, Y101, R103, D104, L107, E108, A144, C145, P147, A148, D149, D150, R155	T82, P84, R87, V92, Q96, H97, D100, Y101, R103, D104, L107, E108, D150, I152, R155	P334, Q336, L337, W375, K379, K388, R391, G392, R394, Y395, Y396, Y397, D398, G407, R409
	ETS-1	ID1/3-PA7	$p16^{INK4a}$ promoter
	S332, G333, P334, Q336, W375, K379, N380, K381, K383, E387, K388, L389,R391, R394, Y395, D398, K399	L1, A10, L12, C14, H15, R18, W19, M20	DA19, DG20, DG21, DA22, DA23, DG24, DT33, DT34, DT36, DT37, DG43
Validation for bio-molecular complexes			
Ramachandran plot assessment			
Residues in favored region (%)	86.6	84.5	87.0
Residues in additionally allowed region (%)	11.4	13.5	13.0
Residues in generously allowed region (%)	0.8	0.0	0.0
Residues in disallowed region (%)	1.2	2.0	0.0
ERRAT			
Overall quality factor	95.16	86.71	99.25

[A] The nonbonded energies were calculated with the OPLS parameters using an 8.5 Å cut-off

DA: deoxy adenine, DG: deoxy guanine, DT: deoxy thymine

with p16^{INK4a} was also determined by the docking analysis. Based on the intermolecular energies of the docked complexes, the best cluster was selected and showed low (<2 Å) RMSD deviations compared with the entire cluster, which signifies that the structure of each selected complex is very stable. The HADDOCK scores and energies, interacting residues and validation results of each bio-molecular complex are given in Table 2.

ID1 and ETS-1 interaction

The ID1 and ETS-1 complex with the lowest intermolecular energy (vdw and elec) of -731.93 kcal mol^{-1} was selected for interaction analysis. The DIMPLOT analysis of these complexes (Fig. 7a) exposes that the amino acid residues HLH motif: loop region (T82), α4 (V92, E93, Q96, H97, D100, Y101, R103, D104, L107 and E108) and C-terminal residue

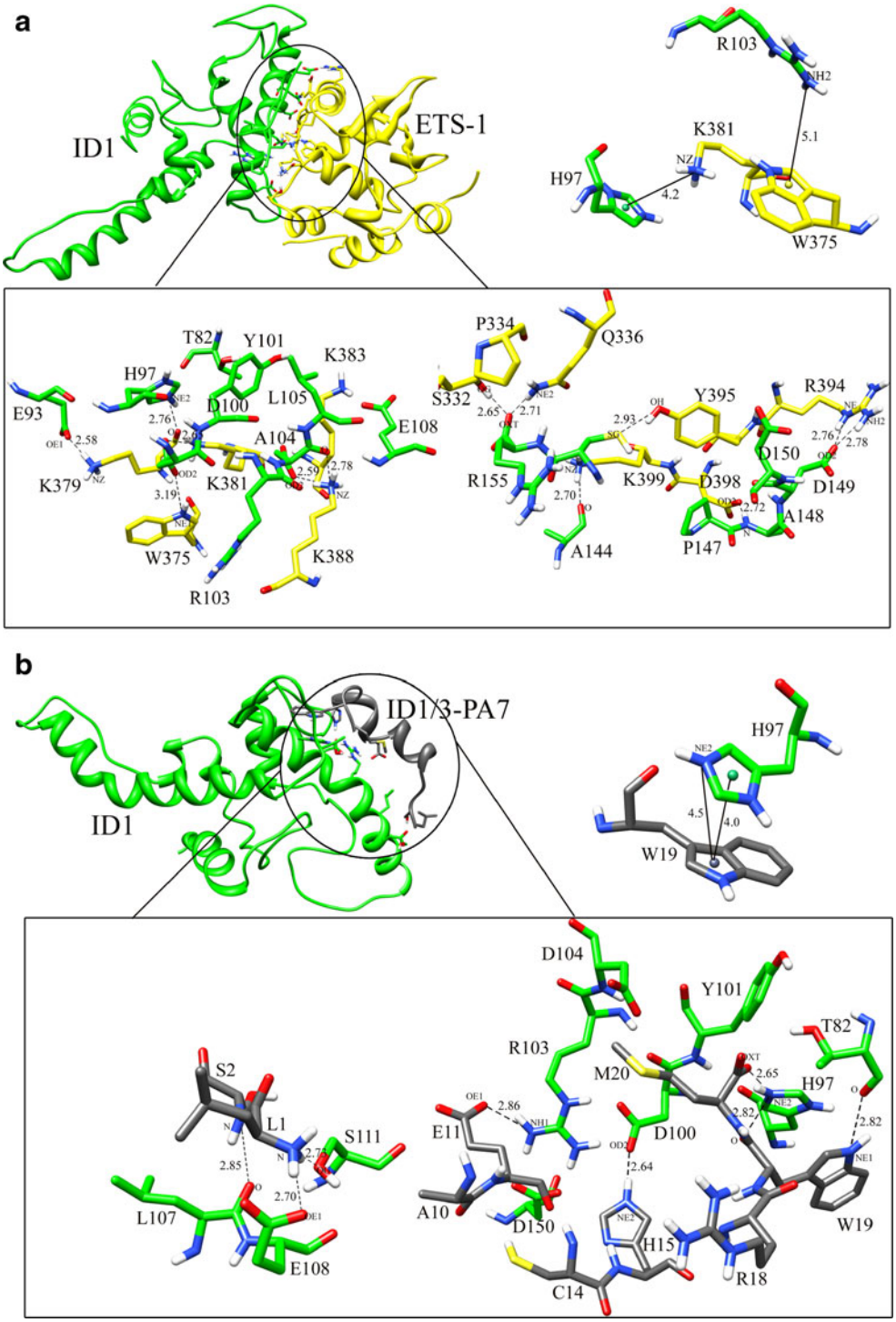


Fig. 7 The bio-molecular complexes obtained by HADDOCK docking simulation. (**a**) The HLH motif and C-terminal region of ID1 forms an interaction with ETS DNA binding domain containing winged helix-turn-helix motif of ETS-1 protein. (**b**) The interacting complex of ID1 and with ID1/3-PA7 displaying the interactions between HLH motif and C-terminal region of ID1 with ID1/3-PA7 peptide aptamer. The hydrogen bond and VdW interaction are shown in zoomed picture, color code green represents ID1, yellow for ETS1 and gray for ID1/3-PA7. Other inter-molecular interactions also shown in stick representation

A144, C145, P147, A148, D149, D150, C154 and R155 of ID1 is involved in HBs and VdW interactions with the residues in the Winged helix-turn-helix (H2:- W375; Loop:- K379, N380, K381 and K383; H3:- E387, K388, L389, R391, R394 and Y395) motif, the loop connecting HI-1 and H1(S332, G334, P334 and Q336) and D398 and K399 of ETS DNA binding domain containing ETS-1 protein (Table 2). In addition, the H97 of ID1 makes the strong π-cation interaction with K381 of ETS-1 by the distance of 4.19 Å whereas W375 of ETS-1 makes weak π-cation interaction [62] with R103 of ID1 with a distance of 5.14 Å (Fig. 7a). From the above mentioned interacting residues W375, K379, E387, K388, R391, R394, Y395 and K399 of ETS-1 are proven to be responsible for binding with the promoter region of DNA to carry out its transcriptional regulation activity [50], which is strongly restrained by ID1 interacting residues such as H97, D100, R103, D104, L107, A144, C145, D149, D150 and C154. The buried surface area of 2188 $Å^2$ provides more interface regions for the complex formation. The binding attraction for this process is extremely high, as predicted by docking scores, thus providing insight into the importance of ID1 in terms of antagonist activity against ETS-1. The validation results (Table 2) also expose the strength of the complex. These results indicate the clear enumeration of antagonist activity of ID1 against ETS-1 through its inhibition of DNA binding of ETS-1 with its partner.

ID1 and ID1/3-PA7 interaction

From the cluster of ID1 and ID1/3-PA7 interaction study, the best complex was selected in terms of lowest intermolecular energy (-349.21 kcal mol^{-1}) (Fig. 7b). In this complex, the HLH motif (loop:- T82, P84 and R87; α4:- V92, Q96, H97, D100, Y101, R103, D104, L107 and E108) and C-terminal loop (D150, I152 and R155) of ID1 makes strong HBs and VdW interaction with L1, A10, E11, L12, G13, C14, H15, R18, W19 and M20 of ID1/3-PA7 peptide aptamer, which is a strong inhibitor of ID1 [27] in ovarian cancer. In addition, H97 makes strong cation-π and π-π interaction with W19 with a distance of 4.5 and 4.0 Å respectively (Fig. 7b). The residues of T82, V92, Q96, H97, D100, Y101, R103, D104, L107, E108 and R155 of ID1 which makes strong interaction with ETS-1 is restrained by ID1/3-PA7 peptide aptamer. The buried surface area of 1528 $Å^2$ invokes for the high possibility of interaction between this complex, as shown by the HADDOCK docking solution and energy score. The quality of the bio-molecular complex was evaluated using the Ramachandran plot (>84 % of the residues are in the favored region) and ERRAT (86.71 % of an overall quality factor) score.

ETS-1 and p16 ^{INK4a} promoter interaction

The ETS DNA binding domain of ETS-1 binds to AGGAAG core binding motif of $p16^{INK4a}$ promoter elements with a lowest intermolecular energy of -736.12 kcal mol^{-1} (Fig. 8). The residues of ETS-1 resides in the loop region of HI-2 and H1 (P334, Q336 and L337), helix H2 (W375 and K379), helix H3 (K388, R391, G392, R394, Y395, Y396, Y397 and D398) and a β-hairpin loop formed by β3 and β4 (G407 and R409) makes HBs and VdW interaction with virtually identical bases (DA19, DG20, DG21, DA22, DA23 and DG24) and other bases (DT33, DT34, DT36, DT37 and DG43) in $p16^{INK4a}$ promoter region. As like ETS1-S-EBS complex (PDB id:2NNY) [50], the H3 triplet residue R391, R394 and Y395 of ETS-1 are flanked by a series of DG21, DA22 and DG24 of $p16^{INK4a}$ to make DNA sugar backbone interaction, and insert itself into the DNA major groove [51]. Moreover, one of the triplets R394 forms cation-π interaction [62] with imidazole ring of DA22 present in the $p16^{INK4a}$ promoter region with a distance of 5.79 Å respectively (Fig. 8). All these interactions and the high buried surface area (1797 $Å^2$) of these complexes, supports the strength of interaction, which might lead to the activation of tumor suppressor $p16^{INK4a}$ and in turn activates apoptosis. The validation result also indicates the quality of the complex structure (Table 2).

Though, the structural level evidence for the interaction of ID1 with ETS-1, ID1 with ID1/3-PA7 and ETS-1 with $p16^{INK4a}$ is not available. In our study, the interaction of ID1with its partner provides an insight into the molecular interactions and the key hot spot residues responsible for their antagonistic role played in $p16^{INK4a}$ transcriptional regulatory pathway.

Alanine scanning mutagenesis

The hot spot residues of ID1, which are responsible for its strong interaction with ETS-1 and ID1/3-PA7, were identified through alanine scanning mutagenesis analysis. The ID1-ETS-1 and ID1-ID1/3-PA7 bio-molecular complexes are submitted to $DrugScore^{PPI}$ [54]. This tool search for interface residues of complexes and mutates them to alanine individually and finds the effect of this mutation on the binding free energy of the complexes (ΔΔG). The server produces the results with positive values of ΔΔG for potential hot spot residues [54]. Based on the results of such analysis, the mutant residues of ID1 such as V92A, E93A, D100A, Y101A, R103, D104A, L107A, D149A, D150 and

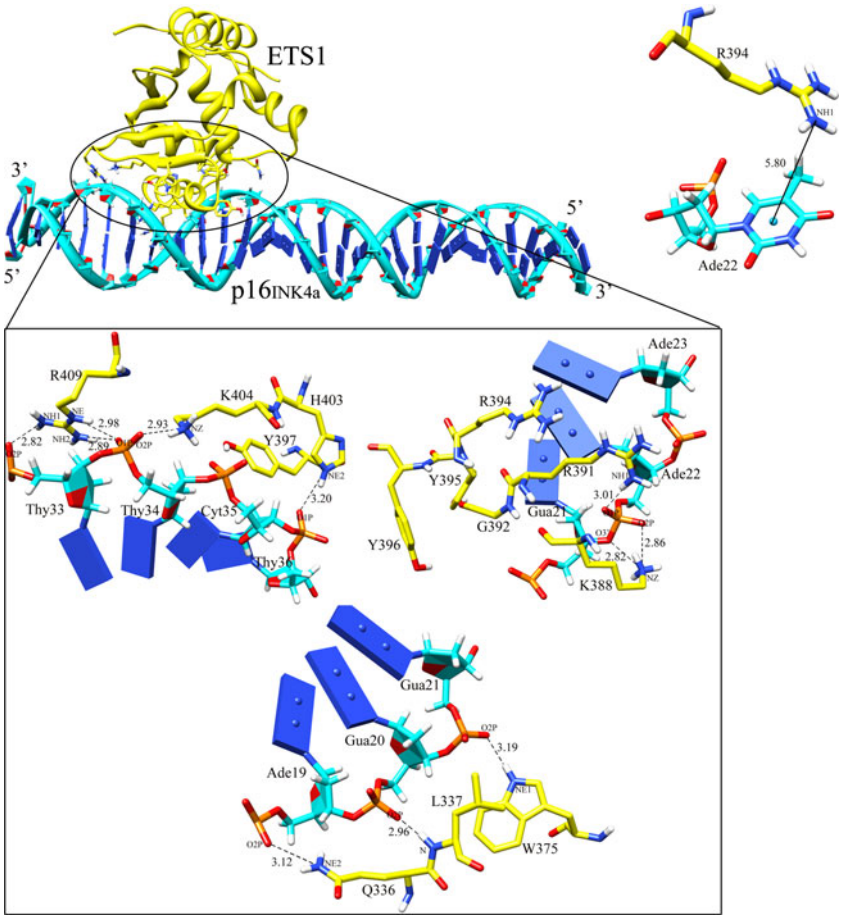


Fig. 8 The ETS DNA binding domain of ETS-1 forms molecular interaction with promoter region of $p16^{INK4a}$. The interactions are displayed in stick representation and the hydrogen bonds are drawn in black dotted line. Other inter-molecular interactions formed between these complexes are also given in stick representation

I152A are identified as potential hot spot residues based on their ΔΔG score (Fig. 9a). Similarly, the ID1-ID1/3-PA7 complex mutational analysis shows the residues D100A, Y101A, R103A, D104A, L107A, D150A and I152A as potential hot spot residues (Fig. 9b). From both the complex mutational studies, the residues D100, Y101, R103, D104, L107A, D150 and I152 of ID1 are identified as potential hot spot residues crucial for forming and stabilizing the biomolecular complexes. Hence, these residues are considered to be potential residues and play the role in ETS-1 mediated $p16^{INK4a}$ regulation.

Stability of ID1 with ETS-1 and ID1/3-PA7 complex

The docked ID1 with ETS-1 and ID1/3-PA7 complex grasp energetically stable conformation and it is confirmed by the molecular dynamics studies. The hot spot residues of ID1 responsible for its crucial role played with ETS-1 and ID1/3-PA7 were shown to be stable throughout the 50 ns MD simulation. The conformational changes between these complexes are monitored by the backbone RMSDs of each complex with respect to its initial structure. The RMSD profile of these complexes were shown in Fig. 10, which demonstrates the RMSD of the ID1 with ETS-1 complex attains their equilibrium after the 10 ns (10,000 ps) with the maximum RMSD of 0.60 nm, whereas ID1 with ID1/3-PA7 complex shows maximum RMSD of 0.40 nm after an equilibrium period of 10 ns (10,000 ps). The potential energy of the ID1 with ETS-1 and ID1/3-PA7 are -1.27×10^{6} and -7.02×10^{5} respectively, and shows the complex stability in terms of lowest energy.

Contact profile of ID1-ETS-1and ID1-ID1/3-PA7 simulated complex

The interactions harbored by the identified hot spot residues of ID1 are found to be stabilized throughout the time scale of simulations. In ID1-ETS-1 complex (Fig. 11a), the hydrogen bonds formed by E93, D100, D104 and D149 were stable, while the interaction played by the remaining ones was lost due to the changes in the interacting conformation. Similarly, the VdW contacts made by T82, V92, E93, H97, D100, Y101, R103, D104, L107, D149, D150 C154 and R155 of ID1 with P334, W375, K379, N380, K381, E387, K388, R394 and Y395 of ETS-1 are found to be in the normal range. Moreover, the interaction of ID1/3-PA7 (L1,

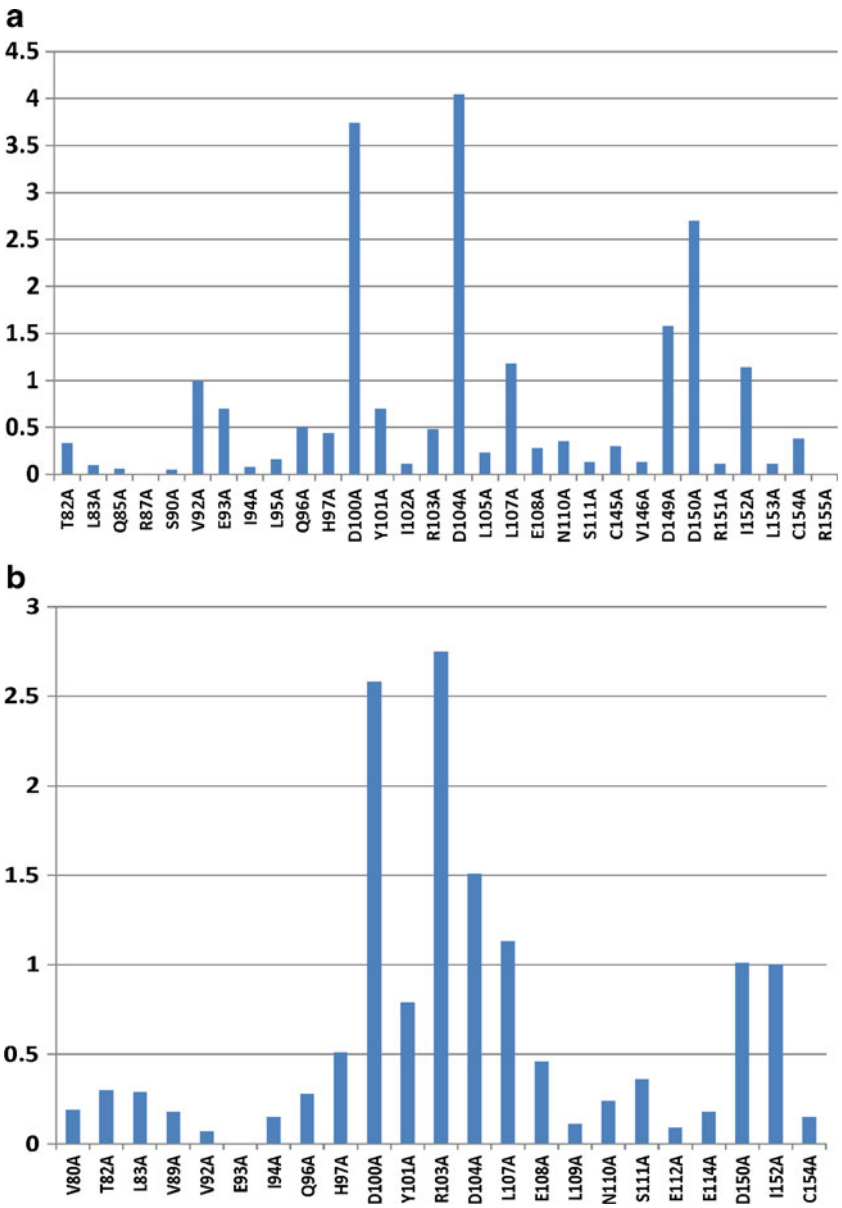


Fig. 9 The graph represents the ΔΔG value of each hot spot residue mutated to alanine and indicates the change in free energy binding in the given complexes. (**a**) ΔΔG of ID1-ETS1 complex and their hot spot residues. (**b**) ΔΔG of ID1-ID1/3-PA7 complex and their hot spot residues. The positive ΔΔG values represent the potential hot spot residues contribution in bio-molecular complexes

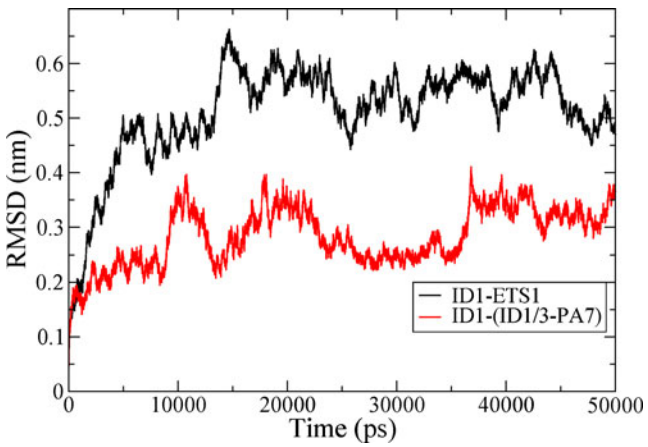


Fig. 10 The backbone RMSD profile of ID1 with ETS-1 and ID1/3-PA7 simulated complexes

L8, E11, L12, H15, W19 and M20) with the aforementioned identified hot spot residues of ID1 also maintains their VdW interactions (Fig. 11b). The average bond length and standard deviation of each contact are given in supplementary Table 1. The stability of the observed interactions indicates their strong contribution to maintain the individual components as a strong complex for propagating their molecular function.

The structural mechanism of ID1 and ETS-1 in $p16^{INK4a}$ regulatory pathway

The docking and simulation analysis of ID1 with ETS-1 and ID1/3-PA7, and docking studies of ETS-1 with $p16^{INK4a}$ provides an insight into the molecular mechanism of $p16^{INK4a}$ regulatory pathway leading to apoptotic process. The structural level schematic model for the involvement of

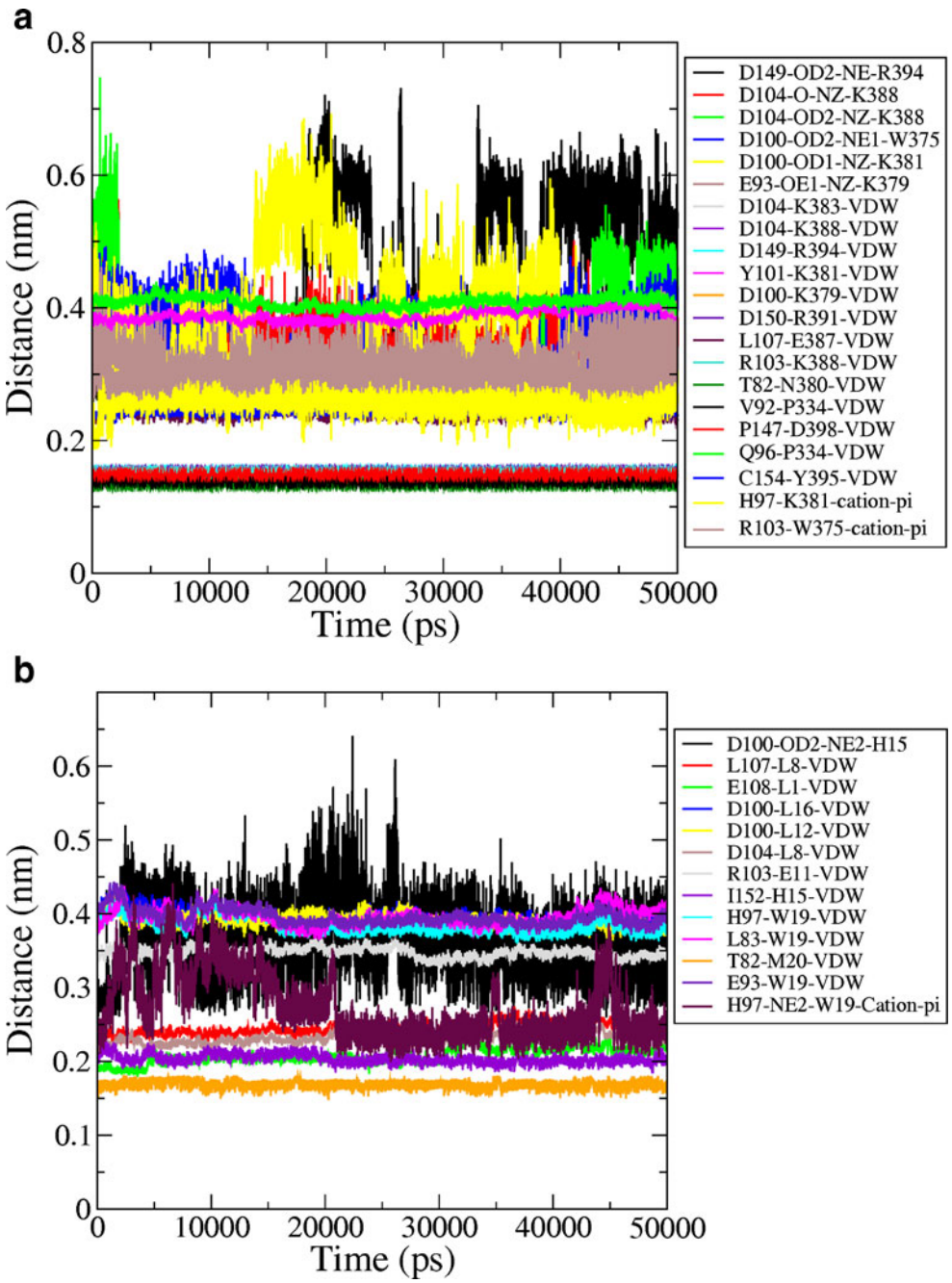


Fig. 11 The HBs and VdW contact profile of simulated complexes. (**a**) The contact profile of ID1 with ETS1 complex. (**b**) Contact profile of ID1 with ID1/3-PA7

ID1 in the regulation of $p16^{INK4a}$ through the inhibition of ETS-1 protein is given in Fig. 12 with the three central bio-molecular interactions. The step 1 bio-molecular interaction is a normal process, in which the ETS DNA binding (H1, H1', H2, H3, B1, B2, B3 and B4) of ETS-1 proteins identifies the 5'-AGGAAG-3' core motif of $p16^{INK4a}$ and regulates the expression of $p16^{INK4a}$ which in turn activates the apoptotic pathway by inhibiting the CDKs. The oncogenic property of ID1 forms type 2 bio-molecular interaction in ovarian cancer cell, in which ID1 via its HLH motif (α3, loop and α4) makes tight interaction with the DNA binding region of ETS-1 and subsequently inhibits the normal expression of $p16^{INK4a}$, which prevents the activation of the apoptotic pathway. As an approach to understand the effect of ID1 inhibition, the 20 amino acid peptide ID1/3-PA7 construct was used, and it binds at the ETS-1 binding region of ID1, which prevents ID1 from binding with ETS-1 and consequently, the $p16^{INK4a}$ will be expressed.

Conclusions

Understanding the interacting mechanism of ID1 with transcriptional regulatory proteins provides structural insights

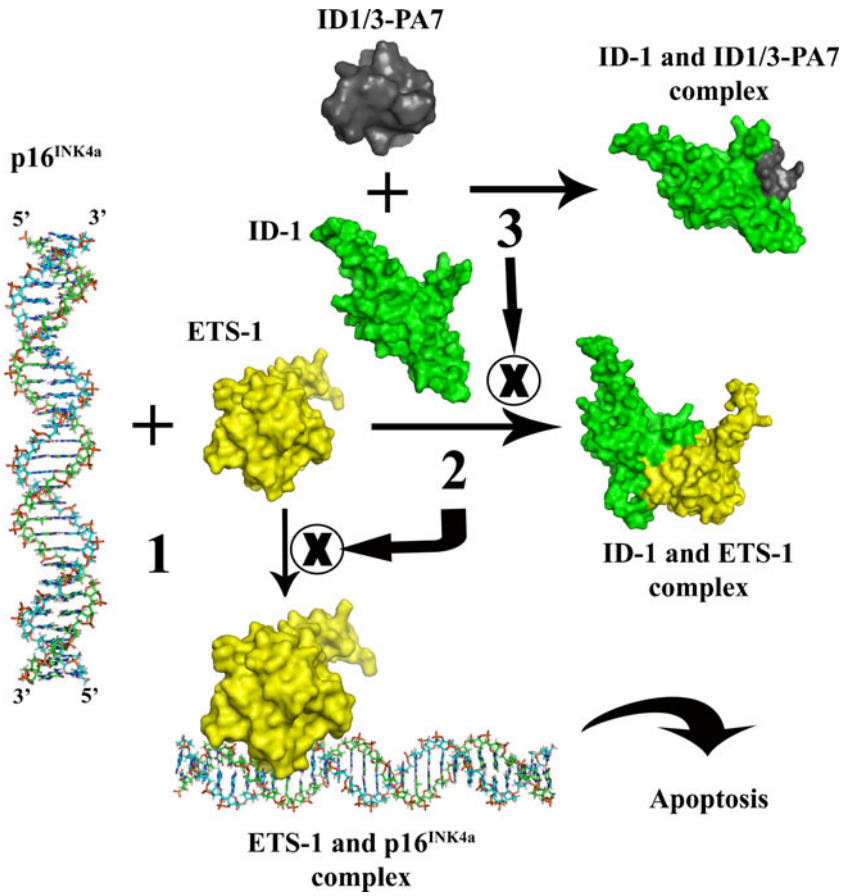


Fig. 12 The structural level mechanism of ID1 in $p16^{INK4a}$ through the mediation of ETS-1. The type 1 interacting mechanism explains the normal process of ETS-1 binding with promoter region of $p16^{INK4a}$ which subsequently induces the apoptotic pathway. Type 2 interacting mechanism restrains the type 1 normal process in which the HLH motif of ID1 holds the ETS DNA binding domain of ETS-1 and this step is controlled by the ID1/3-PA7 peptide aptamer by binding with ID1 in type 3 interaction

for the expression of $p16^{INK4a}$ in CDK pathway. Accordingly, the structure of ID1 and ID1/3-PA7 has been predicted, and analyzed for their stability and reliability of the structure through 50 ns molecular dynamics simulation. The molecular interaction between ID1 with ETS-1 and ID1/3-PA7, and ETS-1 with $p16^{INK4a}$ was analyzed using HADDOCK docking simulation. From this analysis, the hot spot residues of ID1 (HLH motif) responsible for their strong interaction with ETS-1 and ID1/3-PA7 peptide aptamer were identified, and also confirmed through the alanine scanning mutagenesis. The oncogenic role of ID1 against ETS-1 will be restrained by the peptide aptamer ID1/3-PA7 which is proven in the experimental analysis and our results also support these interactions. The stability of the ID1 with ETS-1 and ID1/3-PA7 bio-molecular complexes were analyzed using 50 ns MD simulation. The abovementioned results of docking simulation studies describes the molecular interaction of ID1 with ETS-1 in the p16INK4a regulatory pathway by three types of the reaction mechanisms. These intermolecular interacting mechanisms provide an insight to the structured based drug design in the future to treat ovarian cancer.

Acknowledgments R. Krishna and P. Manivel thanks University Grant Commission (UGC), Government of India for providing financial assistance (F.No.37-313/2009 (SR)) to carry out the Research work. Kannan M. (No.F. 14-2(SC)/2009 (SA-III)) and Nishith Saurav Topno (No.F. 14-2(ST)/2010 (SA-III)) thank UGC for Rajiv Gandhi National Fellowship to pursue their PhD degree. J. Muthukumaran (No.09/559/(0076)/2011/EMR-I) thanks CSIR (Council of Scientific and Industrial Research) for Senior Research fellowship. R. Krishna also thanks Department of Biotechnology and Department of Information technology, Government of India, New Delhi for their financial support to Centre for Bioinformatics, Pondicherry University. We thank E. Malanco, P. Elavarasan and S. Manikandan, Senior technical assistants, Centre for Bioinformatics, Pondicherry University for their technical supports. Authors also extend their greetings and regards to Jayakanthan M. and Bhuvaenshwari S. for their timely help to carry out the research work.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P (2001) Estimating the world cancer burden: Globocan 2000. Int J Cancer 94(2):153–156. doi:10.1002/ijc.1440
2. Ohtani N, Zebedee Z, Huot TJ, Stinson JA, Sugimoto M, Ohashi Y, Sharrocks AD, Peters G, Hara E (2001) Opposing effects of Ets and Id proteins on p16INK4a expression during cellular senescence. Nature 409(6823):1067–1070. doi:10.1038/35059131

3. Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell 61(1):49–59
4. Biggs J, Murphy EV, Israel MA (1992) A human Id-like helix-loop-helix protein expressed during early development. Proc Natl Acad Sci U S A 89(4):1512–1516
5. Christy BA, Sanders LK, Lau LF, Copeland NG, Jenkins NA, Nathans D (1991) An Id-related helix-loop-helix protein encoded by a growth factor-inducible gene. Proc Natl Acad Sci U S A 88 (5):1815–1819
6. Riechmann V, van Cruchten I, Sablitzky F (1994) The expression pattern of Id4, a novel dominant negative helix-loop-helix protein, is distinct from Id1, Id2 and Id3. Nucleic Acids Res 22(5):749–755. doi:10.1093/nar/22.5.749
7. Sun XH, Copeland NG, Jenkins NA, Baltimore D (1991) Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. Mol Cell Biol 11(11):5603–5611
8. Davis RL, Cheng PF, Lassar AB, Weintraub H (1990) The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell 60(5):733–746
9. Ellenberger T, Fass D, Arnaud M, Harrison SC (1994) Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev 8(8):970–980. doi:10.1101/gad.8.8.970
10. Ma PC, Rould MA, Weintraub H, Pabo CO (1994) Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell 77 (3):451–459
11. Murre C, McCaw PS, Baltimore D (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell 56(5):777–783
12. Henthorn P, Kiledjian M, Kadesch T (1990) Two distinct transcription factors that bind the immunoglobulin enhancer microE5/kappa 2 motif. Science 247(4941):467–470. doi:10.1126/science.2105528
13. Hu JS, Olson EN, Kingston RE (1992) HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. Mol Cell Biol 12 (3):1031–1042
14. Massari ME, Murre C (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol 20(2):429–440
15. Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB et al. (1989) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell 58(3):537–544
16. Skerjanc IS, Truong J, Filion P, McBurney MW (1996) A splice variant of the ITF-2 transcript encodes a transcription factor that inhibits MyoD activity. J Biol Chem 271(7):3555–3561. doi:10.1074/jbc.271.7.3555
17. Zhang Y, Babin J, Feldhaus AL, Singh H, Sharp PA, Bina M (1991) HTF4: a new human helix-loop-helix protein. Nucleic Acids Res 19(16):4555. doi:10.1093/nar/19.16.4555
18. Lassar AB, Davis RL, Wright WE, Kadesch T, Murre C, Voronova A, Baltimore D, Weintraub H (1991) Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell 66(2):305–315. doi:10.1016/0092-8674(91)90620-E
19. srael MA, Hernandez MC, Florio M, Andres-Barquin PJ, Mantani A, Carter JH, Julin CM (1999) Id gene expression as a key mediator of tumor cell biology. Cancer Res 59(7 Suppl):1726s–1730s
20. Norton JD, Atherton GT (1998) Coupling of cell growth control and apoptosis functions of Id proteins. Mol Cell Biol 18(4):2371–2381
21. Norton JD (2000) ID helix-loop-helix proteins in cell growth, differentiation and tumorigenesis. J Cell Sci 113(Pt 22):3897–3905
22. Wibley J, Deed R, Jasiok M, Douglas K, Norton J (1996) A homology model of the Id-3 helix-loop-helix domain as a basis for structure-function predictions. Biochim Biophys Acta 1294 (2):138–146
23. Kiewitz SD, Cabrele C (2005) Synthesis and conformational properties of protein fragments based on the Id family of DNA-binding and cell-differentiation inhibitors. Biopolymers 80(6):762–774. doi:10.1002/bip.20287
24. Garrell J, Modolell J (1990) The Drosophila extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. Cell 61(1):39–48
25. Li R, Pei H, Watson DK (2000) Regulation of Ets function by protein - protein interactions. Oncogene 19(55):6514–6523. doi:10.1038/sj.onc.1204035
26. Perk J, Iavarone A, Benezra R (2005) Id family of helix-loop-helix proteins in cancer. Nat Rev Cancer 5(8):603–614. doi:10.1038/nrc1673
27. Mern DS, Hasskarl J, Burwinkel B (2010) Inhibition of Id proteins by a peptide aptamer induces cell-cycle arrest and apoptosis in ovarian cancer cells. Br J Cancer 103(8):1237–1244. doi:10.1038/sj.bjc.6605897
28. Mern DS, Hoppe-Seyler K, Hoppe-Seyler F, Hasskarl J, Burwinkel B (2010) Targeting Id1 and Id3 by a specific peptide aptamer induces E-box promoter activity, cell cycle arrest, and apoptosis in breast cancer cells. Breast Cancer Res Treat 124(3):623–633. doi:10.1007/s10549-010-0810-6
29. Lin CQ, Singh J, Murata K, Itahana Y, Parrinello S, Liang SH, Gillett CE, Campisi J, Desprez PY (2000) A role for Id-1 in the aggressive phenotype and steroid hormone response of human breast cancer cells. Cancer Res 60(5):1332–1340
30. Schindl M, Oberhuber G, Obermair A, Schoppmann SF, Karner B, Birner P (2001) Overexpression of Id-1 protein is a marker for unfavorable prognosis in early-stage cervical cancer. Cancer Res 61(15):5703–5706
31. Langlands K, Down GA, Kealey T (2000) Id proteins are dynamically expressed in normal epidermis and dysregulated in squamous cell carcinoma. Cancer Res 60(21):5929–5933
32. Maruyama H, Kleeff J, Wildi S, Friess H, Buchler MW, Israel MA, Korc M (1999) Id-1 and Id-2 are overexpressed in pancreatic cancer and in dysplastic lesions in chronic pancreatitis. Am J Pathol 155(3):815–822. doi:10.1016/S0002-9440(10)65180-2
33. Ouyang XS, Wang X, Lee DT, Tsao SW, Wong YC (2002) Over expression of ID-1 in prostate cancer. J Urol 167(6):2598–2602. doi:10.1016/S0022-5347(05)65044-6
34. Zhang X, Ling MT, Feng H, Wong YC, Tsao SW, Wang X (2004) Id-I stimulates cell proliferation through activation of EGFR in ovarian cancer cells. Br J Cancer 91(12):2042–2047. doi:10.1038/sj.bjc.6602254
35. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5(4):725–738. doi:10.1038/nprot.2010.5
36. Zhang Y (2007) Template-based modeling and free modeling by I-TASSER in CASP7. Proteins 69(Suppl 8):108–117. doi:10.1002/prot.21702
37. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16(4):404–405. doi:10.1093/bioinformatics/16.4.404
38. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theor Comput 4(3):435–447. doi:10.1021/ct700301q

39. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718. doi:10.1002/jcc.20291
40. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78(8):1950–1958. doi:10.1002/prot.22711
41. Kawata M, Nagashima U (2001) Particle mesh Ewald method for three-dimensional systems with two-dimensional periodicity. Chem Phys Lett 340(1–2):165–172. doi:10.1016/S0009-2614(01)00393-1
42. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18(12):1463–1472. doi:10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h
43. Miyamoto S, Kollman PA (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13(8):952–962. doi:10.1002/jcc.540130805
44. Martonak R, Laio A, Parrinello M (2003) Predicting crystal structures: the Parrinello-Rahman method revisited. Phys Rev Lett 90(7):075503. doi:10.1103/PhysRevLett.90.075503
45. Goyal A, Muthu K, Panneerselvam M, Pole AK, Ramadas K (2011) Molecular dynamics simulation of the Staphylococcus aureus YsxC protein: molecular insights into ribosome assembly and allosteric inhibition of the protein. J Mol Model 17(12):3129–3149. doi:10.1007/s00894-011-0998-3
46. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26(2):283–291. doi:10.1107/S0021889892009944
47. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2(9):1511–1519. doi:10.1002/pro.5560020916
48. Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125(7):1731–1737. doi:10.1021/ja026939x
49. de Vries SJ, van Dijk AD, Krzeminski M, van Dijk M, Thureau A, Hsu V, Wassenaar T, Bonvin AM (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. Proteins 69(4):726–733. doi:10.1002/prot.21723
50. Lamber EP, Vanhille L, Textor LC, Kachalova GS, Sieweke MH, Wilmanns M (2008) Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. EMBO J 27(14):2006–2017. doi:10.1038/emboj.2008.117
51. Garvie CW, Pufall MA, Graves BJ, Wolberger C (2002) Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. J Biol Chem 277(47):45529–45536. doi:10.1074/jbc.M206327200
52. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8(2):127–134. doi:10.1093/protein/8.2.127
53. Manivel P, Muthukumaran J, Kannan M, Krishna R (2011) Insight into residues involved in the structure and function of the breast cancer associated protein human gamma synuclein. J Mol Model 17(2):251–263. doi:10.1007/s00894-010-0718-4
54. Kruger DM, Gohlke H DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. Nucleic Acids Res 38(Web Server issue):W480–W486. doi: 10.1093/nar/gkq471
55. Pesce S, Benezra R (1993) The loop region of the helix-loop-helix protein Id1 is critical for its dominant negative activity. Mol Cell Biol 13(12):7874–7880
56. Yokota Y, Mori S (2002) Role of Id family proteins in growth control. J Cell Physiol 190(1):21–28. doi:10.1002/jcp.10042
57. Bounpheng MA, Dimas JJ, Dodds SG, Christy BA (1999) Degradation of Id proteins by the ubiquitin-proteasome pathway. FASEB J 13(15):2257–2264
58. Deed RW, Armitage S, Norton JD (1996) Nuclear localization and regulation of Id protein through an E protein-mediated chaperone mechanism. J Biol Chem 271(39):23603–23606. doi:10.1074/jbc.271.39.23603
59. Weiss MS, Brandl M, Suhnel J, Pal D, Hilgenfeld R (2001) More hydrogen bonds for the (structural) biologist. Trends Biochem Sci 26(9):521–523
60. Lesk A (2001) Introduction to protein architecture. Oxford University Press, Oxford
61. Guo Z, Mohanty U, Noehre J, Sawyer TK, Sherman W, Krilov G (2010) Probing the alpha-helical structural stability of stapled p53 peptides: molecular dynamics simulations and analysis. Chem Biol Drug Des 75(4):348–359. doi:10.1111/j.1747-0285.2010.00951.x
62. Gallivan JP, Dougherty DA (1999) Cation-pi interactions in structural biology. Proc Natl Acad Sci U S A 96(17):9459–9464. doi:10.1073/pnas.96.17.9459