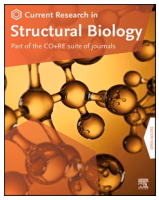
Current Research in Structural Biology 7 (2024) 100150

Contents lists available at ScienceDirect

Current Research in Structural Biology

journal homepage: www.journals.elsevier.com/current-research-in-structural-biology



Research Article

Stability and dynamics of extradenticle modulates its function

Aakanksha Singh , Bidisha Acharya , Beas Mukherjee , Veda Sheersh Boorla [1], Soumendu Boral [2], Snigdha Maiti [3], Soumya De [*]

School of Bioscience, Indian Institute of Technology Kharagpur, Kharagpur, WB, 721302, India

ARTICLE INFO

Handling editor: Dr A Wlodawer

Keywords:
Transcription factor
Protein-DNA interaction
NMR spectroscopy
MD simulation
Hydrogen exchange

ABSTRACT

Extradenticle (EXD) is a partner protein of the HOX transcription factors and plays an important role in the development of Drosophila. It confers increased affinity and specificity of DNA-binding to the HOX proteins. However, the DNA-binding homeodomain of EXD has a significantly weaker affinity to DNA compared to the HOX homeodomains. Here, we show that a glycine residue (G290) in the middle of the EXD DNA-binding helix primarily results in this weaker binding. Glycine destabilizes helices. To probe its role in the stability and function of the protein, G290 was mutated to alanine. The intrinsic stability of the DNA-binding helix increased in the G290A mutant as observed by NMR studies and molecular dynamics (MD) simulation. Also, NMR dynamics and MD simulation show that dynamic motions present in the wild-type protein are quenched in the mutant. This in turn resulted in increased stability of the entire homeodomain ($\Delta\Delta G_{G\rightarrow A}$ of −2.6 kcal/mol). Increased protein stability resulted in three-fold better DNA-binding affinity of the mutant as compared to the wild-type protein. Molecular mechanics with generalized Born and surface area solvation (MMGBSA) analysis of our MD simulation on DNA-bound models of both wild-type and mutant proteins shows that the contribution to binding is enhanced for most of the interface residues in the mutant compared to the wild-type. Interestingly, the flexible N-terminal arm makes more stable contact with the DNA minor groove in the mutant. We found that the two interaction sites i.e. the DNA-binding helix and the unstructured N-terminal arm influence each other via the bound DNA. These results provide an interesting conundrum: alanine at position 290 enhances both the stability and the DNA-binding affinity of the protein, however, evolution prefers glycine at this position. We have provided several plausible explanations for this apparent conundrum. The function of the EXD as a HOX co-factor requires its ability to discriminate similar DNA sequences, which is most likely comprom

1. Introduction

Homeodomain proteins play important roles in the development of multicellular organisms (Gehring et al., 1994; Scott et al., 1989). Most vertebrates have almost 250 homeodomain-containing transcription factors (Holland, 2013), which account for 15–30% of all transcription factors in an organism (De Mendoza et al., 2013). Homeodomains are formed of 60 amino acids, which contain three α-helices and an N-terminal unstructured arm. The second and third helices form a helix-turn-helix motif. The third helix is the DNA-recognition helix and inserts into the DNA major groove (Laughon, 1991). The unstructured N-terminal arm interacts with the DNA minor groove and modulates binding specificity (Joshi et al., 2007). Based on sequence similarity and other features, such as the presence of other domains, homeodomain-containing proteins have been grouped into various classes (Bürglin and Affolter, 2016). Here we have studied the Drosophila transcription factor Extradenticle (EXD), which belongs to the TALE superclass of homeodomains (Mann and Chan, 1996; Bürglin and Affolter, 2016). These homeodomains have a three amino acid loop extension (TALE) between the first and the second helices. The TALE superclass consists of MKX, IRO, TGIF, PREP, MEIS and PBC subclasses (Bürglin and Affolter, 2016). EXD belongs to the PBC subclass of TALE homeodomains.

EXD performs its functions in association with its partner proteins, especially the HOX transcription factors (van Dijk and Murre, 1994; Mann and Chan, 1996). Hox genes play important roles in

* Corresponding author.
E-mail address: somde@iitkgp.ac.in (S. De).
[1] Present address: Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, 16802, USA.
[2] Present address: Structural Biology Initiative, CUNY Advanced Science Research Center, New York, NY 10031, USA.
[3] Present address: Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.

https://doi.org/10.1016/j.crstbi.2024.100150
Received 30 December 2023; Received in revised form 3 April 2024; Accepted 6 May 2024
Available online 9 May 2024

anterior-posterior (AP) axis determination during the development of all bilateral animals from insects to vertebrates (Pearson et al., 2005; Mann et al., 2009). However, HOX proteins have highly conserved DNA-binding homeodomains that bind very similar DNA sequences *in vitro* (Mann et al., 2009). EXD helps the HOX proteins achieve DNA-binding specificity *in vivo* (Pearson et al., 2005; Joshi et al., 2007; Slattery et al., 2011) and also modulates their regulatory properties of transcriptional activation or repression (Joshi et al., 2010). EXD interacts with HOX proteins via a hexapeptide motif, often termed as YPWM motif, appended to the N-terminus of the HOX homeodomain (Foos et al., 2015; Maiti et al., 2019). The EXD homeodomain has a hydrophobic pocket formed by the loop between the first and the second helices that interact with the YPWM motif. This interaction enables the HOX and EXD proteins to cooperatively bind cognate DNA more tightly, which typically have half sites for binding both homeodomains (Joshi et al., 2007; Chang et al., 1995). It is important to note that HOX proteins can bind their cognate DNA with nanomolar affinity, which is enhanced in the presence of EXD (Pellerin et al., 1994; Pinsonneault et al., 1997). Interestingly, most studies report almost no binding for EXD under the same conditions, thus indicating much weaker binding of EXD to cognate DNA (van Dijk and Murre, 1994). The same has been observed for the homologous human protein PBX1 (Van Dijk et al., 1995; Beslu et al., 2004), which has an almost identical homeodomain sequence as EXD (Fig. 1). This is surprising as homeodomains are known to bind cognate DNA with high affinity. Inspection of homeodomain sequences revealed the presence of a glycine residue in the middle of the DNA-binding helix of the PBC subclass of the TALE transcription factors (Fig. 1). Glycine is known to destabilize alpha helices. Here we have investigated the role of this glycine in modulating the stability, dynamics and function of EXD.

The conservation of the glycine residue in the middle of the DNA-binding helix H3 in the PBC subclass of the TALE transcription factors raises important questions regarding its function. Typically, conserved residues are required for one or more of the following reasons: fast and correct folding of the protein (Chikunova et al., 2021); structural integrity of a folded protein (Bellanger et al., 2023); and correct function of the protein (Bellanger et al., 2023; Dulebohn et al., 2006). Mutating such residues results in reduced or complete loss of function. To determine the role of the glycine at position 290, we mutated it to alanine, which is the next smallest residue and hence, should result in minimal structural perturbation. Also, alanine is the most helix-stabilizing amino acid (Pace and Scholtz, 1998). Our NMR studies showed very similar structures of the wild-type (EXD^{WT}) and the mutant (EXD^{G290A}) proteins. Backbone NMR dynamics experiments showed no difference in the fast ps-ns timescale motions between both proteins. However, slower μs-ms timescale motions were observed in the wild-type protein that was quenched in the mutant. Also, NMR-based hydrogen exchange experiments showed increased stability of the mutant than the wild-type protein. DNA-binding experiments showed that the EXD^{G290A} mutant binds three-fold tighter to DNA in comparison to the wild-type protein. Furthermore, 1.5 μs molecular dynamics (MD) simulations were performed on the structural models of free and DNA-bound EXD^{WT} and EXD^{G290A}, which further support our observations and provide important insights into the underlying reasons for enhanced stability and function of EXD^{G290A}.

2. Materials and methods

2.1. Cloning, mutagenesis, protein expression and purification

The minimal homeodomain (residues Ala238−Ile300) of Extradenticle (EXD) (UniProtKB P40427) from *D. melanogaster* was cloned into the pET28a(+) vector. Glycine 290 was mutated to alanine by site-directed mutagenesis using non-overlapping primers (Boral et al., 2023). The methylated template DNA was digested with *Dpn*I. The 5′ end of the PCR product was phosphorylated by T4 polynucleotide kinase enzyme (PNK). The PCR product was ligated by T4 DNA ligase to generate the circular plasmids and transformed into *Escherichia coli* DH5α competent cells. Mutation was confirmed by DNA sequencing. The plasmids were transformed into the *E. coli* BL21(λDE3) strain for protein expression. For the $^{15}N/^{13}C$ -labelled protein sample, 1 L of M9 minimal medium was supplemented with 1 gm of ^{15}N ammonium chloride and 3 gm of ^{13}C glucose as the sole nitrogen and carbon sources, respectively. The culture was grown at 37 °C until cell density (OD_{600}) reached ~0.7. After induction with 1 mM Isopropyl-β-D-thiogalactopyranoside (IPTG) for 5 h at 37 °C, cells were harvested and lysed. The lysate was centrifuged to

Fig. 1. Multiple sequence alignment of homeodomain containing proteins. The conserved amino acids in all the proteins are marked as star and colored in orange. The glycine in middle of helix H3 of PBC subclass is marked in red. Amino acids other than glycine at same position in other families of homeodomain are colored blue.

precipitate the cell debris and the clear supernatant was filtered and passed through Nickel-nitrilotriacetic acid (Ni-NTA) resins for purification. Finally, size exclusion chromatography (HiLoad 16/600 Superdex 200 column) was performed using phosphate buffer (20 mM sodium phosphate, 50 mM NaCl and pH 5.5) to obtain highly pure proteins. The affinity tag (His_6) was removed by thrombin cleavage. The protein concentration was determined by A_{280} using the predicted molar absorptivity (ε280) i.e., 11460 $M^{-1}cm^{-1}$, which was determined by expasy protparam (Gasteiger et al., 2005). The protein purity was determined by SDS-PAGE. The NMR experiments were performed on highly purified proteins and phosphate buffer (20 mM sodium phosphate, 50 mM NaCl and pH 5.5) was used to make the 500 μl NMR samples in 5 mm tube. For EMSA experiments highly purified protein samples were prepared in 1X binding buffer (10 mM Tris (pH 7.8), 100 mM KCl, 6 mM $MgCl_2$, 1 mM EDTA, 50% glycerol, 2 mM BME, 0.02% NP-40 and 0.01 mg/ml BSA).

2.2. NMR experiments and backbone assignment

NMR experiments were performed in a 600 MHz Bruker Avance III spectrometer equipped with a triple-resonance cryogenic probe head at 25 °C. The protein concentrations were 0.6–0.8 mM with 8% D_2O for spin lock. For the long-term stability of the proteins, 0.8 mM PMSF, 2 μl of protease inhibitor cocktail, and 0.04% sodium azide were also added to the final sample. For the sequential backbone assignment, ^{15}N, ^{13}C double-labelled EXD^{WT} was used. Backbone resonances were manually assigned in NMRFAM-SPARKY 1.470 (Lee et al., 2015) using two-dimensional ^{15}N–^{1}H HSQC and three-dimensional NMR experiments such as CBCA(CO)NH, HNCACB, HNCO, and HN(CA)CO (Sattler et al., 1999). The chemical shift assignments are submitted in BMRB with code 51157. The mutant EXD^{G290A} was assigned using ^{15}N-NOESY and ^{15}N-TOCSY experiments (Sattler, Schleucher and Griesinger, 1999). These spectra were processed and analyzed using NMRPipe (Delaglio et al., 1995) and Sparky (Lee et al., 2009), respectively. The secondary structure propensity was determined from the secondary chemical shifts (Mielke and Krishnan, 2009) as follows:

$$\Delta\Delta CS = \left(C_\alpha^{obs} - C_\alpha^{rc}\right) - \left(C_\beta^{obs} - C_\beta^{rc}\right)$$

where C_α^{obs} and C_β^{obs} are the experimentally determined chemical shifts for the α and β carbons of a residue, respectively, and C_α^{rc} and C_β^{rc} are the random coil chemical shifts for the α and β carbons of the same amino acid type, respectively.

The chemical shift perturbation (CSP) for each residue was calculated as follows.

$$CSP = \sqrt{(\Delta H)^2 + (0.154 \times \Delta N)^2}$$

where ΔH and ΔN are the chemical shift differences of the amide protons and nitrogens, respectively, between the EXD^{WT} and EXD^{G290A}.

2.3. Backbone amide ^{15}N relaxation experiments

Amide ^{15}N R_1, R_2 and steady-state heteronuclear {^{1}H}-^{15}N NOE experiments were collected at 25 °C for both the proteins. Spectra for R_1 (50, 100, 150, 200, 400, 600, 900 and 1200 ms) and R_2 (25, 50, 75, 100, 150, 200, 250, and 300 ms) time series were collected in random order to minimize any systematic error. Relaxation rate constants R_1 and R_2 were computed using in-house Matlab (MathWorks) codes by fitting the peak intensities (I) to single exponential decay

$$I = I_0 e(-tR_i)$$

where I_0 is the Initial intensity, 't' is the relaxation delay, and R_i is either R_1 or R_2 (Maiti and De, 2022). Errors of the rate constants were determined by Monte Carlo simulation. The heteronuclear {^{1}H}-^{15}N-NOE for every residue was calculated as a ratio of the peak heights obtained with and without ^{1}H saturation. The recycle delay was set to 5 s. The spectral noise is used to assess the errors in {^{1}H}-^{15}N NOE values by following the error propagation method and by using the given formula

$$\Delta NOE = \{[(1/I_{REF})\delta I_{NOE}]^2 + [(-I_{NOE}/I_{REF}^2)\delta I_{REF}]^2\}^{1/2}$$

Here, ΔNOE is the propagation of error in the heteronuclear {^{1}H}-^{15}N-NOE, $I_{REF,}$ and I_{NOE} are the peak heights of the reference and NOE spectra respectively. δI_{REF} and δI_{NOE} are the spectral noise of the reference and NOE spectra respectively.

2.4. Amide hydrogen exchange experiments

Amide protium-deuterium exchange rates of both the proteins were measured at 25 °C and pH 5.5. The protein samples were lyophilized and dissolved in 100% D_2O, and a series of ^{15}N–^{1}H HSQC spectra were collected over a period of time to monitor the decay of the amide signals as the protium is exchanged by deuterium. The pseudo-first-order rate constants for exchange, k_{ex}, were calculated using in-house MATLAB codes by nonlinear least-squares fitting of the peak intensities, I_t (normalised by the number of transients) to the equation $I_t = (I_0 - I_{00}) \times \exp(-k_{ex}t) + I_{00}$, where t is the midpoint time of each spectrum, I_0 is the initial peak intensity, and I_{00} accounts for the intensity coming from residual water (Boral et al., 2020; De et al., 2016). The time points for each experiment were measured from the time of dissolution of the protein in D_2O. The time between the dissolution of the protein and the start of the first ^{15}N–^{1}H HSQC experiment was 3 min and 2 min 10 s for EXD^{WT} and EXD^{G290A}, respectively. The error in k_{ex} was determined by Monte Carlo simulation. For the calculations of k_{ex}, a corrected pD was used by adding 0.4 to the measured pH of the protein solution (Bai et al., 1993).

Amide protium–protium fast exchange rates for EXD^{WT} were measured at pH 5.5, 7.0, and 8.0 and for EXD^{G290A} at pH 5.5, 6.5, and 7.5 by the CLEANEX-PM method (Hwang et al., 1998) at 25 °C. At each pH, a series of spectra with 10, 20, 30, 40, 50, 60, 80 and 100 ms transfer periods and a reference spectrum using a recycle delay of 12.0 s were collected. The pseudo-first-order rate constants for chemical exchange, k_{ex}, were calculated by nonlinear least-squares fitting of the peak intensities versus transfer time using in-house MATLAB codes. A scaling factor of 0.7 was used to correct for the steady-state water magnetization.

The protection factors (PFs) for each amide proton were calculated as the ratio of the predicted intrinsic exchange rate constant (k_{int}) for an unstructured polypeptide with the same amino acid sequence versus the experimentally calculated exchange rate constant (k_{ex}). The k_{int} values were calculated as described by Englander and co-workers (Zhang, 1995) Here an EX2 mechanism is assumed, where the exchange rate constants (k_{ex}) have a first-order dependence on sample pH and temperature. Also, protein stability is assumed to be independent of pH and temperature.

2.5. Electrophoretic mobility shift assays

The equilibrium dissociation constants (K_D) of EXD^{WT} and EXD^{G90A} for a 25-bp Forkhead 250 double-stranded DNA containing the sequence 5′- CAGCTGG**CGATTAATCTT**GACATTG $^{-3'}$ was measured by EMSA (De et al., 2014**)**. One of the oligonucleotides in the duplex had a fluorescein tag attached to their 5′ end. The lyophilized single strands of DNA were purchased from Integrated DNA Technologies (IDT). The strands were mixed in equimolar concentration, heated to 95 °C for 5 min and slowly cooled down to 4 °C in a thermocycler. For the binding studies, reactions were carried out in 10 mM Tris (pH 7.8), 100 mM KCl, 6 mM $MgCl_2$, 1 mM EDTA, 50% glycerol, 2 mM BME, 0.02% NP-40 and 0.01 mg/ml BSA. The reactions, containing 3 nM DNA and 4 nM to 2 μM protein (EXD^{WT} or EXD^{G90A}), were incubated on ice for 30 min and subjected to

electrophoresis on an 8% native polyacrylamide gel under ice-cold conditions using a BioRad mini-gel system in 0.5X TBE buffer. Gels were scanned with a BioRad ChemiDoc^{TM} MP imaging system. The band intensities of the free DNA were quantified by using the software ImageJ (NIH) (Schneider et al., 2012) and used to calculate the fraction of bound DNA as below.

$$\theta = 1\text{-}([DNA_{free}]/[DNA_{tot}])$$

where $[DNA_{tot}]$ is the band intensity of the unbound DNA in the free DNA lane, and $[DNA_{free}]$ is the band intensity of the unbound DNA in lanes with proteins. The dissociation constant (K_D) is calculated from a nonlinear curve fitting of total protein concentration P_T at each titration point versus the fraction of bound DNA to the following equation assuming.

single-site binding. The nonlinear curve fitting was done using Solver (Microsoft Excel).

$$\theta = P_T/(P_T + K_D)$$

For EXD^{WT} protein, the reported values are average and standard deviation from 4 EMSA experiments. For EXD^{G90A} protein, the reported values are average and standard deviation from 5 EMSA experiments.

2.6. Molecular dynamics simulations and analysis

The model for EXD was extracted from the x-ray structure with PDB code 2R5Z (Joshi et al., 2007). Four residues (A238, R239, R240 and K241) in the N-terminal arm and the loop between H1 and H2 were modeled using Swiss-Model (Waterhouse et al., 2018) with the 2R5Z structure as a template. The G290 was mutated to alanine using the mutagenesis tool in PyMOL (Schrödinger and DeLano, 2020). Both the wild-type and mutant models were validated using the Ramachandran plot and other metrics provided by the MolProbity server (Williams et al., 2018). The structural models of EXD^{WT}-DNA and EXD^{G290A}-DNA complexes were generated by HADDOCK (Van Zundert et al., 2016) using the 2R5Z structure as a template. The structures with the best HADDOCK score and lowest Z-score were selected for molecular dynamics (MD) simulations.

MD simulation was performed as described before (Roy et al., 2022). The simulation was performed using the GROMACS software suite (Van Der Spoel et al., 2005). The topologies for free protein models were built employing Charmm36 all-atom force field (March 2019) and solvated by TIP3P water molecules in a cubic box extending 1.0 nm away from the protein in all directions (Jorgensen et al., 1983). Solvation was followed by the addition of 8 Cl^- ions for neutralization. The topology files for the protein-DNA complexes were generated using the AMBER99SB force field, solvated by the TIP3P water molecules and neutralized with 12 Na^+ ions. Energy minimization was done using the steepest descent algorithm with convergence criteria of <1000.0 kJ/mol/nm. The energy-minimized systems were then equilibrated under NVT and NPT ensemble sequentially for 500 ps each. Initial velocities were assigned from a Maxwell distribution during NVT equilibration using velocity rescale thermostat for simulating constant temperature at 300 K. Isotropic pressure coupling was employed with Parrinello–Rahman barostat used to maintain a constant pressure of 1 bar with time constant 2.0 ps (Bussi et al., 2007 ; Parrinello and Rahman, 1981). Production simulations for the free and DNA-bound wild-type and mutant proteins were then performed for 1500 ns using a leapfrog dynamics integrator with an integration step size of 2 fs. The coordinates and velocities were saved at every 10 ps and periodic boundary conditions (PBC) were considered throughout the simulation in all three dimensions. Particle Mesh Ewald algorithm was used to compute long-range electrostatic interactions with 1.2 nm C cutoff distance. Verlet cutoff scheme was used for short-range neighbour search with a cutoff of 1.2 nm. The PBC corrected trajectory was used for analysis after completion of the production simulation using modules available in GROMACS and in-house Python and Matlab scripts. For analysis, RMSD and RMSF plots were generated using the Python-based software Biotite (Kunzmann and Hamacher, 2018). Secondary structure analysis to study the change in the length of the helices in the free and DNA-bound protein (EXD^{WT} and EXD^{G290A}) complexes over the course of the simulation was performed using the DSSP application of Biotite.

Molecular mechanics Generalized-Born surface-area solvation (MM/GBSA) method (Genheden and Ryde, 2015; Miller, 2012) was employed to estimate the free energy of binding between the proteins (EXD^{WT} or EXD^{G290A}) with DNA as described (Saha et al., 2023). The snapshots taken from each nanosecond of the protein-DNA complex trajectories were analyzed using the tool gmx MMPBSA (Valdés-Tresanco et al., 2021) to investigate the effects of the alanine mutation on the change in free energy of binding ($\Delta G_{binding}$) which is the difference between the free energy of the complex and that of the receptor and ligand, ie., $\Delta G_{binding} = G_{complex} - (G_{protein} + G_{DNA})$ (Weng et al., 2019). The calculations were performed with parameters "igb" and salt concentration as 5 and 150 mM, respectively. $\Delta G_{binding}$ is the total free energy of binding comprising contributions from Van Der Waals and electrostatic interactions, free energy of solvation and solute-dependent entropic terms (Miller, 2012). Residue-wise free energy decomposition was carried out to evaluate the energy contributions of individual residues in the proteins (EXD^{WT} and EXD^{G290A}) and nucleotides in DNA in the binding interaction. Default parameters with idecomp set as 2 were used. DELTA values or energy contributions obtained by decomposition of the total $\Delta G_{binding}$ of the respective amino acids of the EXD^{WT} and EXD^{G290A} protein with DNA were analyzed. The same was carried out for every nucleotide in the 2 chains of DNA to study the free energy of the interaction with the EXD^{WT} and EXD^{G290A} proteins.

3. Results

3.1. G290 is an unusual residue in the DNA-recognition helix of EXD

Glycine is the smallest amino acid. Lack of a sidechain allows glycine to adopt backbone torsion angles that are not accessible to other amino acids. This is reflected in the larger Φ-Ψ space occupied by glycine in the Ramachandran plot of proteins. As a consequence, glycine promotes disorder in proteins (Williams et al., 2001), has the least helical propensity among the 19 non-proline amino acids (Pace and Scholtz, 1998) and often results in the termination of alpha helices (Aurora et al., 1994). In the homeodomain of EXD, the DNA-recognition helix H3 is 19 residues long and extends from T281 to N299 (2R5Z, Joshi et al., 2007). Interestingly, there is a glycine residue (G290) in the middle of helix H3. Inspection of homeodomains from other proteins reveal the absence of glycine residues in the DNA-recognition helix H3 (Fig. 1). Thus, the presence of glycine in H3 appears to be a unique feature of the homeodomain in the PBC subclass of TALE proteins.

The presence of glycine in the middle of helices are typically observed due to packing constraints with other side chains of the protein (Kim et al., 2005) or ligand in the binding interface (Luscombe and Thornton, 2002). Inspection of the DNA-bound structures of EXD reveals that neither is the case for G290 (Fig. 2). It is on the surface and is not packed against any other residue of the protein. In the DNA-bound structures, glycine can be replaced with other small amino acids. Modeling an alanine in place of G290 showed no apparent steric hindrance to DNA-binding. We chose to mutate glycine 290 with an alanine and experimentally test its effect on the stability, structure, dynamics and DNA-binding function of EXD. Glycine to alanine mutation was chosen as alanine is the next smallest amino acid and should have the least structural perturbation. Also, alanine has the highest helical propensity among the natural amino acids (Pace and Scholtz, 1998).

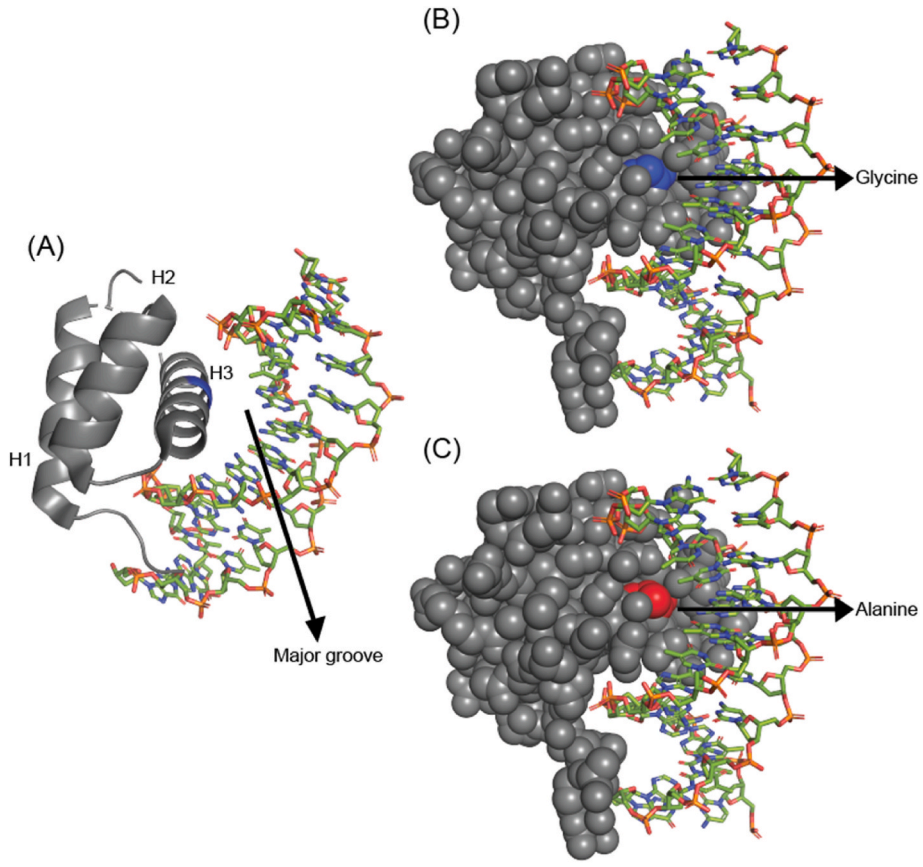


Fig. 2. **(A)** Crystal structure of EXD bound to DNA (2R5Z). EXD consists of three helices H1, H2 and H3. Helix H3 inserts into the major groove of the DNA. **(B)** The crystal structure (2R5Z) is again shown with all heavy atoms of the protein as spheres. Glycine 290 is colored blue. **(C)** In the same structure glycine 290 is mutated to alanine using Pymol (Schrödinger and DeLano, 2020) and shown in red. DNA is shown as sticks. Modeling glycine 290 to alanine results in no apparent steric hindrance in this model.

3.2. Structural characterization of EXD^{WT} and EXD^{G290A} by solution NMR spectroscopy

EXD^{WT} (residues 238 to 300) was prepared as uniformly ^{15}N and ^{13}C labelled protein. Backbone atoms ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{1}H^{\alpha}$, $^{1}H^{\beta}$ and ^{13}CO) were sequentially assigned using triple resonance (^{1}H, ^{15}N, ^{13}C) experiments. EXD^{G290A} was prepared as a ^{15}N-labelled protein. Its backbone amide N–H were assigned by comparing its ^{1}H–^{15}N HSQC spectrum with EXD^{WT}. Residues with significant chemical shift perturbation were assigned using ^{15}N-TOCSY and ^{15}N-NOESY spectra of EXD^{G290A}. Very similar ^{1}H–^{15}N HSQC spectra (Figure SI 1) indicate that EXD^{WT} and EXD^{G290A} have the same fold and that the structure of the protein is not perturbed by G290A mutation (Fig. 3A, B, 3C). Elution of both proteins are at the same volume indicates very similar shape and hydrodynamic radius of EXD^{WT} and EXD^{G290A} (Figure SI 2). The majority of the chemical shift perturbations are observed for the neighboring residues of the mutation site (Fig. 3D).

The secondary structures of EXD^{WT} were determined from the backbone chemical shifts (Fig. 3E) and NOE cross peaks in ^{15}N-NOESY spectra (Fig. 4). It has three alpha helices as expected in a homeodomain. Helix H1 (residues K247 to S258) and helix H2 (residues E269 to C278) are consistent with the DNA-bound structures of EXD. Interestingly, helix H3 (residues V282 to Y296) is shorter in the free EXD^{WT} compared to the DNA-bound protein (residues T281 to N299). Similar shorter helix boundaries were observed for the free form of the homologous protein PBX1 (Jabet et al., 1999). In EXD^{G290A}, helix H3 also extends up to Y296. However, more cross peaks are observed in the NOESY spectra for EXD^{G290A} as compared to EXD^{WT} (Fig. 4). Thus, the G290A mutation slightly increases the stability of helix H3 but not to the same extent as the DNA-bound structures.

3.3. G290A mutation enhances the DNA-binding affinity of EXD

The G290A mutation in the DNA-recognition helix results in no significant change in the EXD structure. In order to test the effect of the G290A mutation on the DNA-binding function of EXD, an electrophoretic mobility shift assay (EMSA) was performed. A fluorescently labelled DNA was used containing the sequence $^{5'}$CAGCTGG**CGATTAATCTT**GACATTG$^{3'}$ (Fig. 5A, B and 5C) from the Forkhead250 promoter (Ryoo and Mann, 1999). Forkhead promoter regulates the expression of the Forkhead gene which is involved in the development of embryonic salivary glands (Ryoo and Mann, 1999). The HOX protein sex combs reduced (SCR) and EXD cooperatively bind to this promoter and activate the Forkhead gene (Ryoo and Mann, 1999; Joshi, Sun and Mann, 2010). The protein concentration ranged from 4 nM to 2 μM. The DNA concentration was kept constant at 3 nM for all reaction samples. From several replicate experiments, the dissociation constants (K_D) were determined as 146 ± 13 nM for EXD^{WT} and 38 ± 4 nM for EXD^{G290A}, indicating a 3-fold tighter binding for the mutant (Fig. 5D, Figure SI 3 and 4). Thus, the G290A mutation is not detrimental to the DNA-binding function of EXD; contrary to that, replacement of glycine with an alanine in the DNA-recognition helix enhances its DNA-binding affinity.

In the EMSA gels, two supershifts can be observed for both EXD^{WT} and EXD^{G290A}. These two shifts arise from the binding of one and two molecules of EXD protein to the DNA, respectively (Fig. 5B and C). The Forkhead250 promoter has two binding sites, one for a HOX transcription factor and another for EXD (Fig. 5A). At lower protein concentrations, EXD binds to one of these sites and at higher concentrations, a

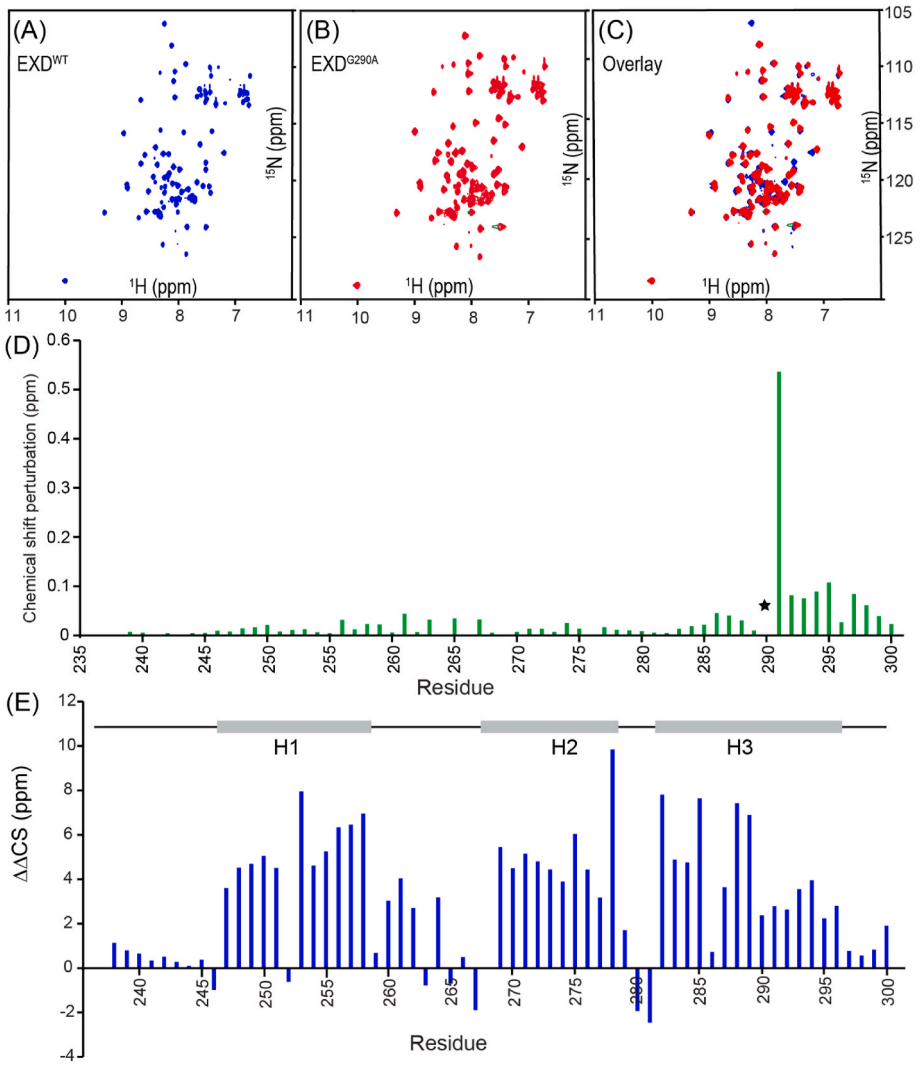


Fig. 3. ^{15}N–^{1}H HSQC of **(A)** EXD^{WT}, **(B)** EXD^{G290A} and **(C)** their overlay. **(D)** Chemical shift perturbation between EXD^{WT} and EXD^{G290A} is shown for each residue. Position of mutation is marked with a star. Major perturbations are observed in the residues following the residue at 290 in middle of helix H3. (**E**) Secondary chemical shifts (ΔΔCS) of EXD^{WT} are plotted for each residue. The resulting secondary structure is shown on the top where helices are represented as gray rectangles.

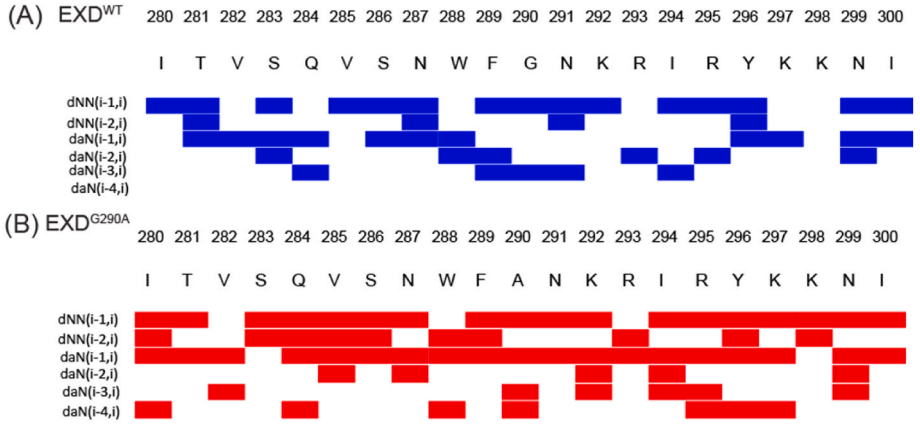


Fig. 4. **(A)** Short and medium range diagnostic NOE patterns for α-helices are shown for EXD^{WT} (blue bars) and for **(B)** EXD^{G290A} (red bars). Unambiguous NOEs pattern between $^{1}H^{N}$ and $^{1}H^{\alpha}$ are mentioned as dNN(i-1, i), dNN(i-2, i), dαN(i-1, i), dαN(i-2, i), dαN(i-3, i) and dαN(i-4, i). More cross peaks are observed in the NOESY spectra for EXD^{G290A} as compared to EXD^{WT}.

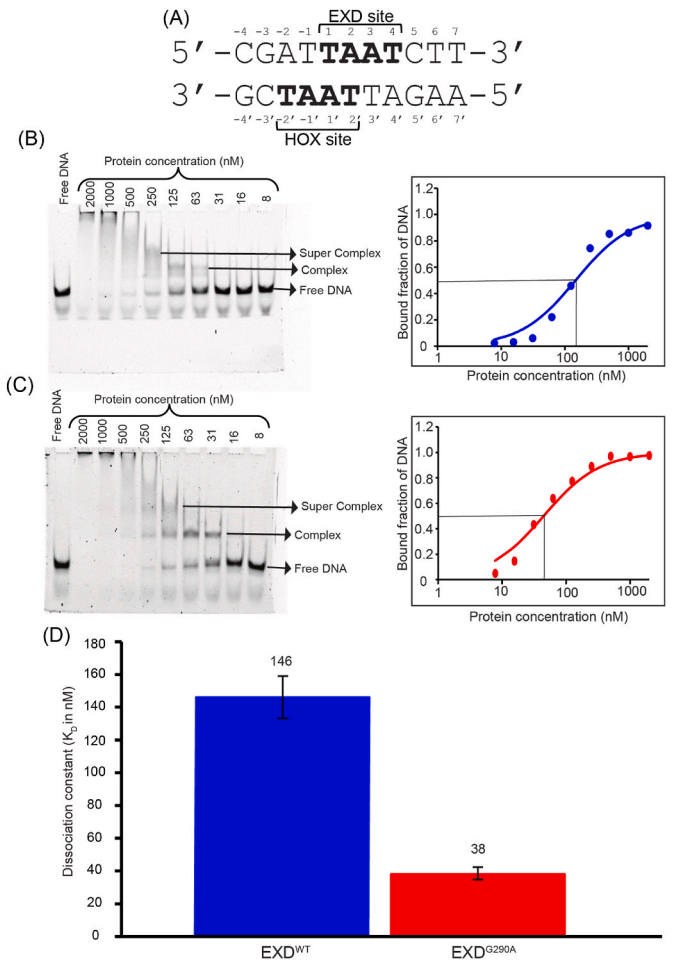


Fig. 5. **(A)** The DNA sequence used for the DNA-binding experiments. Nucleotides labelled in bold are the DNA binding sites on opposite strands for HOX and EXD, respectively. DNA-binding affinities were determined by the EMSA studies are shown for **(B)** EXD^{WT} and **(C)** EXD^{G290A}. The left panel shows representative gel images. The right panel shows the data fitting for the corresponding proteins. In all the lanes 3 nM Forkhead250 DNA was used. Protein concentration varies from 4 nM to 2 μM and mentioned on the top of the gel. Lane labelled as free DNA doesn't contain any protein. **(D)** From EMSA experiments the dissociation constants were measured as 146 ± 13 nM for EXD^{WT} and 38 ± 4 nM for EXD^{G290A}.

second molecule of EXD binds to the remaining site. Thus, for both EXD^{WT} and EXD^{G290A} an apparent dissociation constant is reported here that accounts for EXD binding at two sites on the Forkhead250 DNA.

3.4. ^{15}N relaxation measurements of EXD^{WT} and EXD^{G290A}

The G290A mutation did not affect the protein structure and significantly improved the DNA-binding affinity. To investigate its effect on protein dynamics, we used NMR spectroscopy. Backbone amide ^{15}N relaxation (R_1, R_2, and steady-state heteronuclear NOE) experiments were performed at 25 °C for both EXD^{WT} and EXD^{G290A} (Fig. 6). The heteronuclear $\{^1H\}$-^{15}N NOE values are in the range of 0.55–0.77 for the residues S246 to Y296 for both proteins, indicating the presence of a well-folded structure with flexible N- and C-terminus (Fig. 6A). The residues in the loop between helices H1 and H2 have low $\{^1H\}$-^{15}N NOE values, indicating local flexibility. The domain boundaries of S246 to Y296 are consistent with the secondary structure segments determined from chemical shift analysis and NOESY experiments, and indicate the presence of smaller helix H3 in the free proteins.

Transverse relaxation rate constants (R_2) have average values of 8.9 s^{-1} and 7.7 s^{-1} for the folded regions of EXD^{WT} and EXD^{G290A}, respectively (Fig. 6B). The R_2 values drop for the flexible residues in both termini. The longitudinal relaxation rate constants (R_1) have average values of 1.9 s^{-1} and 2.0 s^{-1} for the folded regions of EXD^{WT} and EXD^{G290A}, respectively (Fig. 6C). Thus, EXD^{WT} and EXD^{G290A} have very similar $\{^1H\}$-^{15}N NOE and R_1 values, however, the R_2 values for EXD^{WT} are consistently higher compared to EXD^{G290A}. Higher R_2 values may indicate chemical exchange due to slower motions in the μs-ms timescale. The R_1R_2 product can be used to efficiently detect the presence of chemical exchange (Kneller et al., 2002). The R_1R_2 values for both proteins were plotted for each residue and compared to the mean value $<R_1R_2>$, calculated as 10% trimmed mean value after excluding residues with low NOE values (Fig. 6D). Several residues, such as S283 and G290 in helix H3, and L261 in loop L1 of EXD^{WT} show significantly increased R_1R_2 values compared to the mean $<R_1R_2>$, indicating the presence of chemical exchange in these regions. Slower μs-ms timescale motions have been reported for the homologous protein PBX1 (Farber and Mittermaier, 2011). In the mutant, the R_1R_2 values for these residues are lowered, indicating quenching to some extent of the μs-ms timescale motions. To further investigate this, both proteins were subjected to hydrogen exchange studies.

3.5. Amide hydrogen exchange reveals increased stability of EXD^{G290A}

The residue-wise stability of both EXD^{WT} and EXD^{G290A} was determined by NMR-based hydrogen exchange (HX) experiments (Saibo et al., 2023; Maiti, 2024). For both EXD^{WT} and EXD^{G290A} protium-deuterium exchange rates were measured for the backbone amides at 25 °C and pH 5.5. The pseudo-first-order rate constants for exchange, k_{ex}, were obtained for several well-protected residues in both proteins. Amide protium-protium exchange rates were measured at pH 5.5, 7.0 and 8.0 for EXD^{WT} and pH 5.5, 6.5 and 7.5 for EXD^{G290A} by CLEANEX-PM (Hwang, van Zijl and Mori, 1998) at 25 °C. Exchange rates for most of the remaining residues were obtained from this method. The protection factors (PF) for amide protons were calculated as the ratio of the predicted intrinsic exchange rate constant (k_{int}) and the experimentally calculated exchange rate constant (k_{ex}).

HX measurements of the proteins revealed increased PFs for EXD^{G290A} as compared to EXD^{WT} (Fig. 7 and Figure SI 5). The PFs for the residues in the folded domain of EXD^{WT} ranged from 10^2 to 10^4, while they ranged from 10^3 to 10^5 for EXD^{G290A}, indicating a 10-fold increase in PF for most residues throughout the protein due to G290A mutation.

3.6. MD simulation confirms increased stability of DNA-binding helix H3 in EXD^{G290A}

NMR experiments indicate significant changes in the global dynamics and stability of EXD upon G290A mutation. To better understand how the mutation affects the dynamics and stability of the protein 1.5 μs molecular dynamics (MD) simulations were performed for both EXD^{WT} and EXD^{G290A} proteins.

To compare the residue-wise stabilities of EXD^{WT} and EXD^{G290A}, we calculated the root mean square fluctuation (RMSF) of the C-alpha atoms for the entire simulation trajectory of both proteins. Overall, EXD^{G290A} has decreased residue-wise flexibilities compared to those of EXD^{WT} (Figure SI 6A), which is consistent with the NMR dynamics results. On average, the residues of helix H3 had an RMSF of 1.63 Å in EXD^{WT} and an RMSF of 1.10 Å in EXD^{G290A}, indicating a 33% decrease in RMSF upon the mutation. The effect of increase in residue-wise stability is also observed through the rest of the protein with decreases in average RMSF of 22% and 42% for residues in helices H2 and H1 respectively. These findings confirm that the G290A mutation resulted in significant increase in stability of the protein by restricting residue-wise flexibilities.

To get better insight on the effect of G290A mutation on helix H3 stability, we calculated the length of the fully-folded helix H3 by

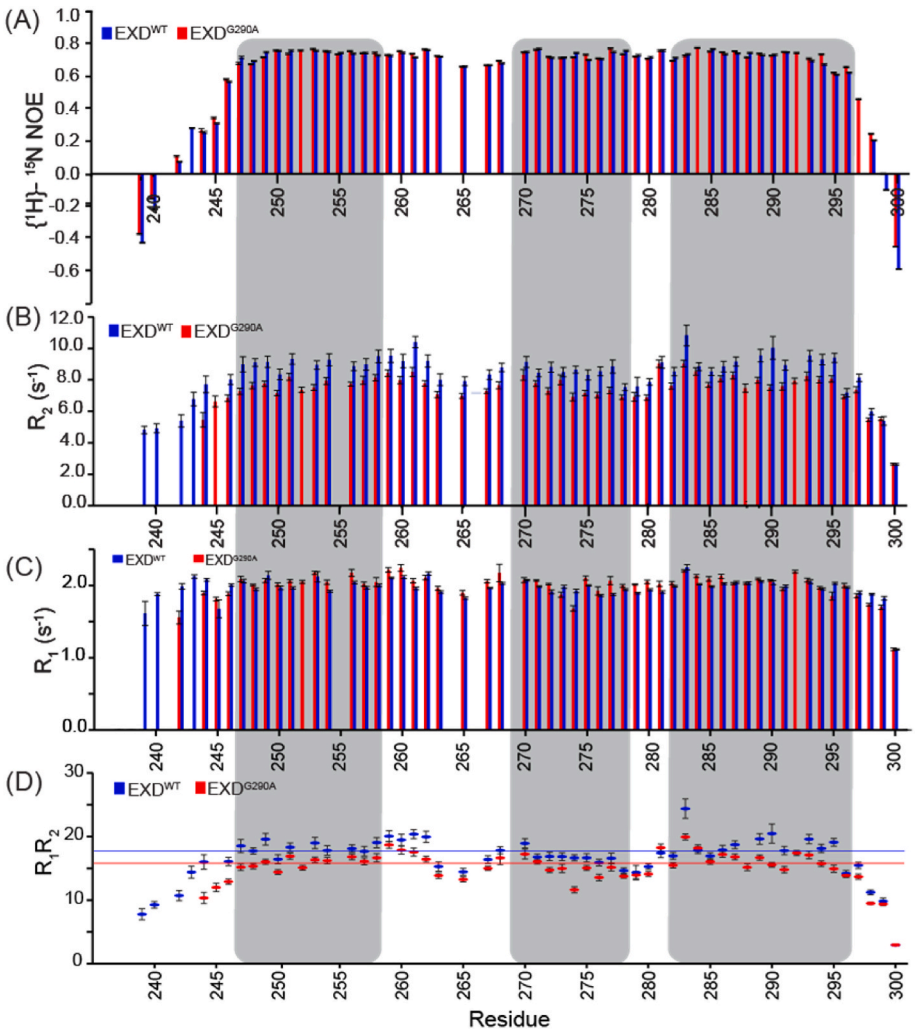


Fig. 6. Fast timescale relaxation (ps-ns) of amides in EXD^{WT} (blue) and EXD^{G290A} (red) are shown for each residue. **(A)** Heteronuclear $\{^1H\}$-^{15}N NOEs, **(B)** transverse relaxation rate constants (R_2), **(C)** longitudinal relaxation rate constants (R_1) and **(D)** R_1R_2 for EXD^{WT} (blue) and EXD^{G290A} (red) are shown. The average of R_1R_2 for EXD^{WT} and EXD^{G290A} are shown as blue and red lines, respectively. Residues in three helices are shaded in gray.

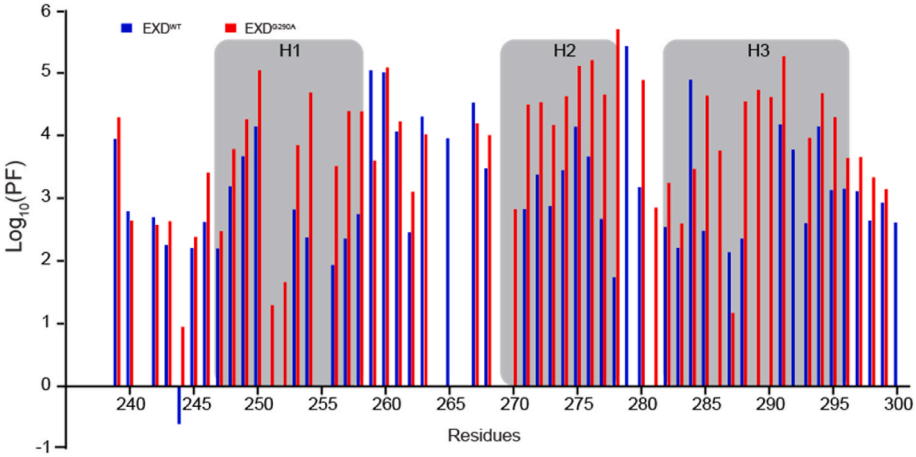


Fig. 7. The residue-wise stability of both EXD^{WT} (blue) and EXD^{G290A} (red) was determined by NMR-based hydrogen exchange (HX) experiments. For both EXD^{WT} and EXD^{G290A} protium-deuterium and protium-protium exchange rates were measured for the backbone amides. Residues in three helices are shaded in gray.

performing DSSP helix assignment along the EXD^{WT} and EXD^{G290A} trajectories. The length of helix H3 varied significantly more for EXD^{WT} than EXD^{G290A} (Figures SI 6B, 6C, 6D). This again confirms that the intrinsic stability of the helix H3 is enhanced as a helix destabilizing residue (glycine) is replaced with a helix stabilizing residue (alanine). Next, we investigated the relative motion of helix H3 within the protein.

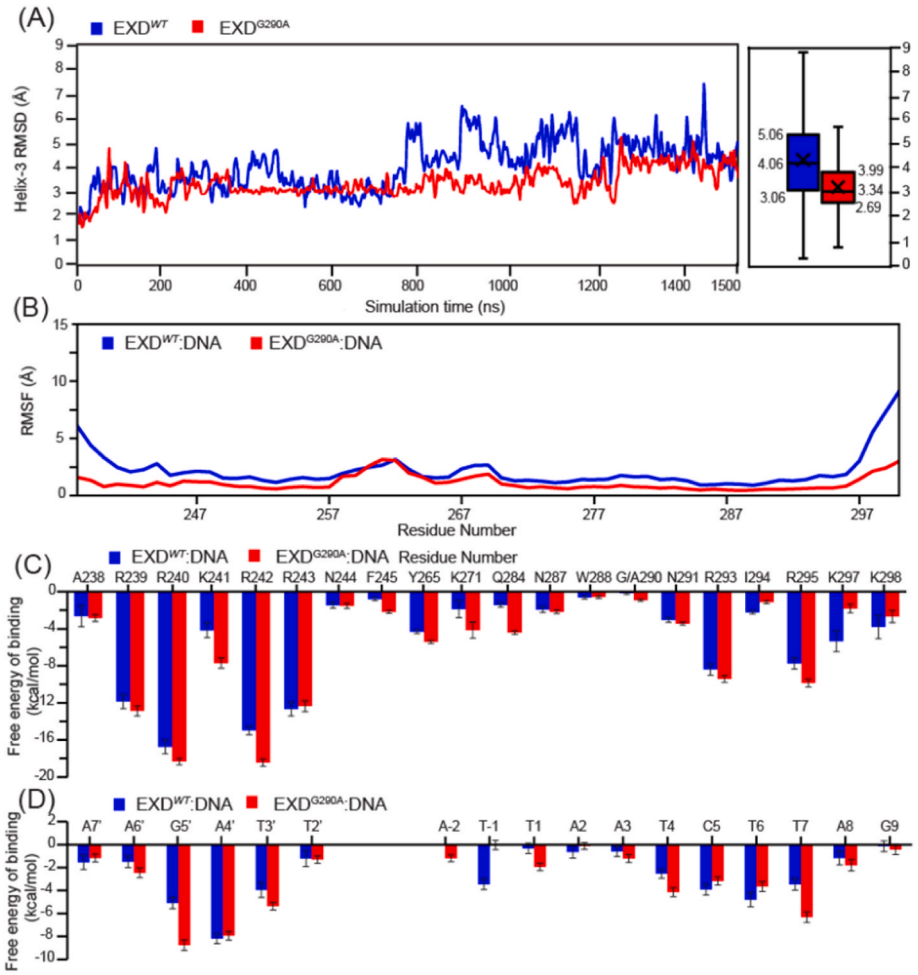


Fig. 8. **(A)** The root mean square deviation (RMSD) in Å of helix H3 from 1.5 μs simulation was obtained after aligning helices H1 and H2 in each frame. The panel on the right shows a boxplot of the RMSD values throughout the simulation grouped together. Average and SD values of RMSD for EXD^{WT} (blue) and EXD^{G290A} (red) are (4.06 ± 1.06) Å and (3.34 ± 0.65) Å, respectively. (B) The root mean square fluctuation (RMSF) is plotted for each residue in Å for DNA bound EXD^{WT} (blue) and EXD^{G290A} (red). The residues at N-terminal and C-terminal of EXD^{WT} showed more fluctuations compared to residues at N- and C-terminal of EXD^{G290A}. (C) Residue wise free energy decomposition is plotted for the amino acids in the interface of EXD^{WT}-DNA complex (blue) and EXD^{G290A}-DNA complex (red). (D) Residue wise free energy decomposition is plotted for the nucleotides in the interface of EXD^{WT}-DNA complex (blue) and EXD^{G290A}-DNA complex (red).

We computed the RMSD of the helix H3 obtained after superimposing helices H1 and H2 throughout the trajectories of both proteins (Fig. 8A). H3 deviates by an average of 4.06 Å (std. dev. 1.06) in EXD^{WT} whereas it deviates by an average of 3.34 Å (std. dev. 0.65) in EXD^{G290A}. This implies that the mutation also leads to a noticeable restriction of the movement of helix H3 within the protein. This is consistent with observed reduction in R_2 relaxation of EXD^{G290A} compared to EXD^{WT}.

3.7. MD simulation of DNA-bound EXD^{WT} and EXD^{G290A}

Mutation of glycine 290 to alanine resulted in tighter binding of DNA by the protein. In order to get a better insight into the role of this mutation in DNA-binding, 1.5 μs molecular dynamics (MD) simulations were performed for the DNA-bound models of EXD^{WT} and EXD^{G290A} proteins. Decreased flexibility was observed for EXD^{G290A} as compared to EXD^{WT} (Fig. 8B). Again, the length of helix H3 varied significantly more for EXD^{WT} than EXD^{G290A} in the DNA bound state (Figures SI 6E, 6F, 6G). The N-terminal arm (residues A238-S246) is unstructured in both proteins, interacts with the DNA minor groove and becomes highly ordered. Again, this unstructured region is more ordered in EXD^{G290A} compared to EXD^{WT} (Figure SI 7).

Furthermore, we estimated the free energy of DNA binding to EXD^{WT} and EXD^{G290A} proteins using molecular mechanics with generalized Born and surface area solvation (MM/GBSA). The same amino acids were found to interact with DNA for both the proteins. Residue-wise free energy contributions were determined and found to increase for most of the interacting residues in EXD^{G290A} as compared to EXD^{WT}, which ultimately results in tighter binding (Fig. 8C and D). Interestingly, the residue at the position 290 (glycine or alanine) has negligible contribution in the overall DNA binding by the respective proteins. Thus, the methyl sidechain of alanine 290 plays no role in enhancing the DNA-binding affinity of EXD^{G290A}, instead the enhanced stability and lower flexibility of helix H3 enables the protein to bind DNA more tightly. Also, a major contribution to DNA-binding comes from the unstructured N-terminal arm, which increases for EXD^{G290A} as compared to EXD^{WT}. This is consistent with the more ordered N-terminal arm observed for EXD^{G290A} (Figure SI 7).

The radius of gyration (R_g) was calculated for the folded domain (residues 247–300) for both EXD^{WT} and EXD^{G290A} proteins (Figure SI 6A). Very similar and constant R_g was observed for both proteins, which is consistent with our SEC measurements (Figure SI 2). Similarly, R_g was measured for the DNA-bound complexes of both proteins considering all residues of both protein and DNA (Figure SI 8B). The R_g of the DNA-bound complex of EXD^{G290A} is consistent throughout the trajectory whereas the R_g of the DNA-bound complex of EXD^{WT} varies along the trajectory. This indicates the flexible nature of the N-terminal arm in the

DNA-bound complex of EXD^{WT} compared to EXD^{G290A}.

4. Discussion

Glycine in the middle of the DNA-binding helix H3 in the PBC family of transcription factors is unusual from a structural biology perspective in two major ways. First, glycine is known to destabilize helices (Williams et al., 2001) and any other amino acid (except proline) is more helix stabilizing (Pace and Scholtz, 1998). Second, glycine in H3 of EXD is surface exposed and does not contribute to ligand binding due to lack of a side chain. Consequently, it is not surprising to find other residues, such as alanine, serine, cysteine, histidine, glutamine, isoleucine and lysine, in this position in other homeodomains (Fig. 1). Hence, this study investigated the effect of glycine290 on the structure, stability, dynamics and DNA-binding function of EXD.

Glycine290 was mutated to alanine as it is the next smallest amino acid and is also the most helix stabilizing amino acid (Pace and Scholtz, 1998). Structural and dynamic characterization by NMR spectroscopy showed both EXD^{WT} and EXD^{G290A} have an unstructured N-terminal arm followed by three alpha helices, which is characteristic of a homeodomain. Both NMR and MD studies show an increase in intrinsic stability of helix H3 due to G290A mutation. This is reflected in more NOESY cross-peaks observed for EXD^{G290A} (Fig. 4). No $d_{aN(i-4,\ i)}$ cross-peaks were observed for EXD^{WT}, whereas several such cross-peaks were observed for EXD^{G290A}. Similarly, in MD simulations the average length of the helix H3 was shorter and also varied significantly more for EXD^{WT} than EXD^{G290A} (Figure SI 6B, 6C and 6D), indicating more intrinsic stability of the mutant helix H3.

Interestingly, the increase in stability due to G290A mutation is not limited to helix H3 but increases the overall stability of the EXD homeodomain as evident from our hydrogen-exchange studies. The cooperative effect of the mutation most likely stems from better packing interactions of the core residues in the stabilized H3, such as W288 and F289 that precede G290, of the protein, which in turn enhanced the overall stability of the protein. For the core residues, the protection factor increased by an average of ~90 fold, which corresponds to $\Delta\Delta G_{G\rightarrow A}$ of −2.6 kcal/mol. The increase in stability is also evident from the lower RMSF for the entire EXD^{G290A} compared to EXD^{WT} (Figure SI 6A) and persistently longer length of helix H1 in addition to helix H3 in EXD^{G290A} (Figure SI 6B and 6C) from our MD simulations.

The increase in overall stability resulted in better DNA-binding by EXD^{G290A} (Fig. 5D). In homeodomains, DNA binding is accomplished by two distinct elements of the domain (Laughon, 1991). The first element is the helix H3 that inserts into the major groove of the DNA and forms electrostatic, H-bond and hydrophobic interactions. In EXD, the conserved N291 in H3 forms two H-bonds with the base of A4' (Joshi et al., 2007) of the TAAT motif of the DNA (Fig. 5A). N287 forms a H-bond with the phosphate group on A4'. I294 makes hydrophobic contacts with the base of A2 and the sugar moiety of A1. Typically, other residues at position equivalent to 290 make contacts to the DNA bases and add to both affinity and specificity. In EXD, the glycine at this position makes no base contact due to the lack of a side chain. The second element is the unstructured N-terminal arm, which is rich in basic residues and inserts into the DNA minor groove. For HOX homeodomain, it has been shown that the N-terminal unstructured region is able to sense the variations in the electrostatic potential of the minor groove, which depends on the minor groove width (Joshi et al., 2007). AT tracts in DNA form narrower minor grooves and result in tighter interaction with the positively charged side chains of basic amino acids (Rohs et al., 2009). Indeed, both interactions play important roles in DNA recognition and binding by EXD. A cooperative effect of G290A mutation is observed in enhanced interaction of both elements with DNA. MMGBSA analysis of our MD simulation on DNA-bound models of both EXD^{WT} and EXD^{G290A} show the relative contribution of each residue in the helix H3 and the N-terminal arm. Contribution to binding is enhanced for most of these residues in EXD^{G290A} compared to EXD^{WT} (Fig. 8C). More stable helix H3 of the mutant protein results in better and stable interactions of its residues with the major groove of the DNA (Fig. 8D). This in turn stabilizes the DNA conformation and quenches its flexibility (Figure SI 9A and 9B). The nucleotide bases preceding the TAAT motif form a narrow minor groove (Figure SI 9C). The basic side chains of the unstructured N-terminal arm are inserted into this minor groove (Figure SI 7). More stable DNA conformation in the EXD^{G290A} bound DNA model results in a stable minor groove structure and hence stable interactions with the unstructured N-terminal arm.Thus a stable helix H3 results in better interaction of the N-terminal arm with the DNA, and this communication between the two interaction sites is transmitted through the bound DNA.

Overall, this work demonstrates that glycine to alanine substitution in helix H3 makes EXD homeodomain a more stable protein and a better DNA-binding domain. It is important to note that glycine to alanine substitution is neutral and is not selected against in natural proteins. In fact, alanine, serine and asparagine are the only three residues with a substitution score of zero in the BLOSUM62 matrix, and all other amino acids have negative substitution scores (Henikoff and Henikoff, 1992). Thus, the presence of glycine290 in the PBC family is surprising, as evolutionary tinkering optimizes the sequence of a protein for better function and stability. In view of this, it is important to consider the function of EXD as well as the PBC family of transcription factors in the context of their role as co-transcription factors. These proteins primarily work in partnership with other transcription factors, especially the HOX family of transcription factors (Ryoo and Mann, 1999). Relatively weak binding of the EXD ensures its dependence on co-transcription factors for the regulation of downstream gene expression. Moreover, it has been shown that the partnership of EXD with HOX transcription factors modulates their DNA-binding specificity (Slattery et al., 2011). The DNA-binding specificity by HOX and EXD proteins is driven by their unstructured N-terminal arm (Joshi et al., 2007; Rohs et al., 2010). Hence, it can be postulated that tighter binding of the N-terminal arm in EXD^{G290A} may compromise its ability to discriminate between different DNA sequences i.e. compromise its DNA-binding specificity. Thus, a more stable EXD^{G290A} may compromise the EXD function by affecting both its partnership with HOX proteins and the specificity of DNA-binding by the HOX-EXD complexes.

5. Conclusions

Overall, we show that a single mutation of glycine to alanine in the DNA-binding helix H3 of EXD results in significant enhancement of its stability and DNA-binding affinity but lowers its flexibility. However, the presence of glycine in the natural sequence of EXD and other members of the PBC transcription factors indicates a finer optimization of the protein sequence by evolution beyond the fitness parameters of stability and affinity. The function of the PBC proteins as HOX co-transcription factors requires their ability to discriminate similar DNA sequences, which is most likely compromised by the quenching of the flexibility and enhancement of stability and affinity.

CRediT authorship contribution statement

Aakanksha Singh: Conceptualization, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Bidisha Acharya:** Conceptualization, Writing – review & editing. **Beas Mukherjee:** Investigation, Formal analysis, Writing – review & editing. **Veda Sheersh Boorla:** Formal analysis, Writing – review & editing. **Soumendu Boral:** Formal analysis, Writing – review & editing. **Snigdha Maiti:** Conceptualization, Writing – review & editing. **Soumya De:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work is funded by the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India (CRG/2020/004984). The authors acknowledge Prof. Praphulla C Shukla for the use of the ChemiDoc gel imaging system and Prof. Dibyendu Samanta for helping with SEC studies. The authors acknowledge the National Supercomputing Mission (NSM) for providing computing resources of "PARAM Shakti" at IIT Kharagpur, which is implemented by C-DAC and supported by the Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology (DST), Government of India. The NMR experiments were done at the Central Research Facility (CRF) at IIT Kharagpur.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2024.100150.

References

Aurora, R., Srinivasan, R., Rose, G.D., 1994. Rules for α-helix termination by glycine. Science 264 (5162), 1126–1130. https://www.science.org/doi/abs/10.1126/science.8178170.

Bai, Y., et al., 1993. Primary structure effects on peptide group hydrogen exchange. Proteins: Struct., Funct., Bioinf. 17 (1), 75–86. https://doi.org/10.1002/prot.340170110.

Bellanger, T., et al., 2023. Significant influence of four highly conserved amino-acids in lipochaperon-active sHsps on the structure and functions of the Lo18 protein. Sci. Rep. 13 (1), 1–13.

Beslu, N., et al., 2004. Molecular interactions involved in HOXB4-induced activation of HSC self-renewal. Blood 104 (8), 2307–2314. https://doi.org/10.1182/blood-2004-04-1653.

Boral, S., et al., 2020. Structural, dynamic, and functional characterization of a DnaX mini-intein Derived from Spirulina platensis provides important insights into intein-Mediated Catalysis of protein Splicing. Biochemistry 59 (50), 4711–4724.

Boral, S., et al., 2023. Extein residues regulate the catalytic function of Spl DnaX intein enzyme by restricting the near-attack conformations of the active-site residues. Protein Sci. 32 (7), 1–16. https://doi.org/10.1002/pro.4699.

Bürglin, T.R., Affolter, M., 2016. Homeodomain proteins: an update. Chromosoma 125 (3), 497–521. https://doi.org/10.1007/s00412-015-0543-8.

Bussi, G., Donadio, D., Parrinello, M., 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126 (1), 014101. https://doi.org/10.1063/1.2408420.

Chang, C.P., et al., 1995. Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. Gene Dev. 9 (6), 663–674. https://doi.org/10.1101/gad.9.6.663.

Chikunova, A., et al., 2021. Conserved residues Glu37 and Trp229 play an essential role in protein folding of β-lactamase. FEBS J. 288 (19), 5708–5722. https://doi.org/10.1111/febs.15854.

De, S., et al., 2014. Steric mechanism of auto-inhibitory regulation of specific and non-specific dna binding by the ets transcriptional repressor ETV6. J. Mol. Biol. 426 (7), 1390–1406. https://doi.org/10.1016/j.jmb.2013.11.031.

De, S., et al., 2016. Autoinhibition of ETV6 DNA binding is established by the stability of its inhibitory helix. J. Mol. Biol. 428 (8), 1515–1530. https://doi.org/10.1016/j.jmb.2016.02.020.

Delaglio, F., et al., 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6 (3), 277–293. https://doi.org/10.1007/BF00197809.

Van Der Spoel, D., et al., 2005. GROMACS: fast, flexible, and free. J. Comput. Chem. 26 (16), 1701–1718. https://doi.org/10.1002/jcc.20291.

van Dijk, M.A., Murre, C., 1994. Extradenticle Raises the DNA binding specificity of homeotic selector gene products. Cell 78 (4), 617–624. https://doi.org/10.1016/0092-8674(94)90526-6.

Van Dijk, M.M.A., Peltenburg, L.T.C., Murre, C., 1995. Hox gene products modulate the DNA binding activity of Pbx1 and Pbx2. Mech. Dev. 52 (1), 99–108. https://doi.org/10.1016/0925-4773(95)00394-G.

Dulebohn, D.P., Hye, J.C., Karzai, A.W., 2006. Role of conserved surface amino acids in binding of SmpB protein to SsrA RNA. J. Biol. Chem. 281 (39), 28536–28545. https://doi.org/10.1074/jbc.M605137200.

Farber, P.J., Mittermaier, A., 2011. Concerted dynamics link allosteric sites in the PBX homeodomain. J. Mol. Biol. 405 (3), 819–830.

Foos, N., et al., 2015. A Flexible extension of the drosophila ultrabithorax homeodomain defines a Novel Hox/PBC interaction mode. Structure 23 (2), 270–279. https://doi.org/10.1016/j.str.2014.12.011.

Gasteiger, E., et al., 2005. The Proteomics Protocols Handbook. The Proteomics Protocols Handbook 571–608. https://doi.org/10.1385/1592598900.

Gehring, W.J., Affolter, M., Bürglin, T., 1994. Homeodomain proteins. Annu. Rev. Biochem. 63 (4), 487–526. https://doi.org/10.1146/annurev.bi.63.070194.002415.

Genheden, S., Ryde, U., 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expet Opin. Drug Discov. 10 (5), 449–461. https://doi.org/10.1517/17460441.2015.1032936.

Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U.S.A. 89 (22), 10915–10919. https://doi.org/10.1073/pnas.89.22.10915.

Holland, P.W.H., 2013. Evolution of homeobox genes. Wiley Interdisciplinary Reviews: Dev. Biol. 2 (1), 31–45.

Hwang, T.-L., van Zijl, P.C., Mori, S., 1998. Accurate quantitation of water-amide proton exchange rates using the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. J. Biomol. NMR 11, 221–226. https://doi.org/10.1023/A:1008276004875. https://link.springer.com/article/.

Jabet, C., et al., 1999. NMR studies of the Pbx1 TALE homeodomain protein free in solution and bound to DNA: Proposal for a mechanism of HoxB1-Pbx1-DNA complex assembly. J. Mol. Biol. 291 (3), 521–530.

Jorgensen, W.L., et al., 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79 (2), 926–935.

Joshi, R., et al., 2007. Functional specificity of a hox protein Mediated by the recognition of minor groove structure. Cell 131 (3), 530–543. https://doi.org/10.1016/j.cell.2007.09.024.

Joshi, R., Sun, L., Mann, R., 2010. Dissecting the functional specificities of two Hox proteins. Gene Dev. 24 (14), 1533–1545. https://genesdev.cshlp.org/content/24/14/1533.short.

Kim, S., et al., 2005. Transmembrane glycine zippers: Physiological and pathological roles in membrane proteins. Proc. Natl. Acad. Sci. U.S.A. 102 (40), 14278–14283. https://doi.org/10.1073/pnas.0501234102.

Kneller, J.M., Lu, M., Bracken, C., 2002. An effective method for the discrimination of motional anisotropy and chemical exchange. J. Am. Chem. Soc. 124 (9), 1852–1853. https://doi.org/10.1021/ja017461k.

Kunzmann, P., Hamacher, K., 2018. Biotite: a unifying open source computational biology framework in Python. BMC Bioinformatics 19 (346), 1–8. https://link.springer.com/article/10.1186/s12859-018-2367-z.

Laughon, A., 1991. DNA binding specificity of homeodomain. Biochemistry 30 (48), 11357–11367.

Lee, W., et al., 2009. PINE-SPARKY: Graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy. Bioinformatics 25 (16), 2085–2087.

Lee, W., Tonelli, M., Markley, J.L., 2015. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 31 (8), 1325–1327. https://doi.org/10.1093/bioinformatics/btu830.

Luscombe, N.M., Thornton, J.M., 2002. Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. J. Mol. Biol. 320 (5), 991–1009. https://doi.org/10.1016/S0022-2836(02)00571-5.

Maiti, S., et al., 2024. Experimental methods to study the structure and dynamics of intrinsically disordered regions in proteins. Current Research in Structural Biology 100138. https://doi.org/10.1016/j.crstbi.2024.100138.

Maiti, S., De, S., 2022. Identification of potential short linear motifs (SLiMs) in intrinsically disordered sequences of proteins by fast time-scale backbone dynamics. Journal of Magnetic Resonance Open 10–11, 100029. https://doi.org/10.1016/j.jmro.2021.100029.

Maiti, S., et al., 2019. Dynamic studies on intrinsically disordered regions of two Paralogous transcription factors reveal Rigid segments with important Biological functions. J. Mol. Biol. 431 (7), 1353–1369. https://doi.org/10.1016/j.jmb.2019.02.021.

Mann, R.S., Chan, S.K., 1996. Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends Genet. 12 (7), 258–262. https://doi.org/10.1016/0168-9525(96)10026-3.

Mann, R.S., Lelli, K.M., Joshi, R., 2009. Hox specificity: unique roles of cofactors and Collaborators. Curr. Top. Dev. Biol. 88 (9), 63–101. https://doi.org/10.1016/S0070-2153(09)88003-4.

De Mendoza, A., et al., 2013. Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. U.S.A. 110 (50) https://doi.org/10.1073/pnas.1311818110.

Mielke, S.P., Krishnan, V.V., 2009. Characterization of protein secondary structure from NMR chemical shifts. Prog. Nucl. Magn. Reson. Spectrosc. 54 (3–4), 141–165. https://doi.org/10.1016/j.pnmrs.2008.06.002.

Miller, B.R., et al., 2012. *MMPBSA.py*: An efficient program for end-state free energy calculations. J. Chem. Theor. Comput. 8 (9), 3314–3321. https://doi.org/10.1021/ct300418h.

Pace, C.N., Scholtz, J.M., 1998. A helix propensity scale based on experimental studies of peptides and proteins. Biophys. J. 75 (1), 422–427. https://doi.org/10.1016/s0006-3495(98)77529-0.

Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. J. Appl. Phys. 52 (12), 7182–7190.
Pearson, J.C., Lemons, D., McGinnis, W., 2005. Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. 6 (12), 893–904. https://www.nature.com/articles/nrg1726.
Pellerin, I., et al., 1994. Hox proteins have different affinities for a Consensus DNA site that Correlate with the positions of their genes on the hox Cluster. Mol. Cell Biol. 14 (7), 4532–4545. https://doi.org/10.1128/mcb.14.7.4532-4545.1994.
Pinsonneault, J., et al., 1997. A model for extradenticle function as a switch that changes HOX proteins from repressors to activators. EMBO J. 16 (8), 2032–2042. https://doi.org/10.1093/emboj/16.8.2032.
Rohs, R., West, S.M., Sosinsky, A., et al., 2009. The role of DNA shape in protein-DNA recognition. Nature 461 (7268), 1248–1253. https://doi.org/10.1038/nature08473.
Rohs, R., et al., 2010. Origins of specificity in protein-DNA recognition. Annu. Rev. Biochem. 79, 233–269. https://doi.org/10.1146/annurev-biochem-060408-091030.
Roy, S., et al., 2022. Structural and dynamic studies of the human RNA binding protein RBM3 reveals the molecular basis of its oligomerization and RNA recognition. FEBS J. 289 (10), 2847–2864. https://doi.org/10.1111/febs.16301.
Ryoo, H.D., Mann, R.S., 1999. The control of trunk Hox specificity and activity by extradenticle. Gene Dev. 13 (13), 1704–1716. https://doi.org/10.1101/gad.13.13.1704.
Saha, R., et al., 2023. Deciphering the conformational stability of MazE7 antitoxin in Mycobacterium tuberculosis from molecular dynamics simulation study. J. Biomol. Struct. Dyn. 0 (0), 1–17. https://doi.org/10.1080/07391102.2023.2280675.
Saibo, N.V., et al., 2023. Nuclear magnetic resonance spectroscopy to Analyse protein folding and dynamics. Protein Folding Dynamics and Stability: Experimental and Computational Methods 61–81. https://doi.org/10.1007/978-981-99-2079-2_4.
Sattler, M., Schleucher, J., Griesinger, C., 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc. 34 (2), 93–158. https://doi.org/10.1016/S0079-6565(98)00025-9.
Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9 (7), 671–675. https://www.nature.com/articles/nmeth.2089.
Schrödinger, L., DeLano, W., 2020. PyMOL. http://www.pymol.org/pymol.
Scott, M.P., Tamkun, J.W., Hartzell, G.W., 1989. The structure and function of the homeodomain. BBA - Rev. Cancer 989 (1), 25–48. https://doi.org/10.1016/0304-419X(89)90033-4.
Slattery, M., et al., 2011. Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. Cell 147 (6), 1270–1282. https://doi.org/10.1016/j.cell.2011.10.053.
Valdés-Tresanco, M.S., et al., 2021. Gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS. J. Chem. Theor. Comput. 17 (10), 6281–6291. https://doi.org/10.1021/acs.jctc.1c00645.
Waterhouse, A., et al., 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. 46 (W1), W296–W303.
Weng, G., et al., 2019. Assessing the performance of MM/PBSA and MM/GBSA methods. 9. Prediction reliability of binding affinities and binding poses for protein-peptide complexes. Phys. Chem. Chem. Phys. 21 (19), 10135–10145.
Williams, R.M., et al., 2001. The protein non-folding problem: amino acid determinants of intrinsic order and disorder. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing 100, 89–100. https://doi.org/10.1142/9789814447362_0010.
Williams, C.J., et al., 2018. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27 (1), 293–315.
Zhang, Y.Z., 1995. Protein and Peptide Structure and Interactions Studied by Hydrogen Exchanger and NMR. University of Pennsylvania, Philadelphia. Ph.D. Thesis.
Van Zundert, G.C.P., et al., 2016. The HADDOCK2.2 Web server: User-Friendly integrative modeling of biomolecular complexes. J. Mol. Biol. 428 (4), 720–725.