Molecular Modeling Connect
ISSN: 3105-3734
2026, Vol. 3, Cite as: doi:10.x/journal.x.x.x

SCIFINITI PUBLISHING

OPEN ACCESS

Research Article

Thermodynamic Quantification of Multi-Sequence DNA Recognition by Prokaryotic Transcription Factor: A Molecular Dynamics Simulation Study

Soumi Das[1]

1 **Independent Researcher, East Burdwan, West Bengal, 713103, India**

Correspondence: Soumi Das

E-Mail: iammandovi@gmail.com

ORCID ID:0009-0005-2869-5849

Abstract

DNA recognition by transcription factors is governed not only by binding affinity but also by DNA sequence-encoded modulation of protein conformational dynamics. Quantitative analysis of residue-specific conformational thermodynamics, based on side-chain torsion angle distributions derived from molecular dynamics simulations and principal component analysis, reveals that the symmetric lactose repressor protein (LacI) undergoes operator-specific conformational redistribution upon binding to the natural lac operator O2 and the synthetic operator O-SymL. Binding to O2 is associated with enhanced conformational entropy in the N- and C-terminal regions, whereas O-SymL binding promotes localized conformational stabilization. Protein dynamics are differentially modulated by polar and hydrophobic residues, with DNA sequence-dependent asymmetry governing this effect. While binding to either operator stabilizes residues at the protein-DNA interface, distal core subdomains retain sequence-dependent conformational landscape remodeling. Long-range modulation of distal domains leads to unequal monomeric dynamics and functional non-equivalence of structurally identical monomers, hallmarks of sequence-dependent allosteric propagation through the LacI scaffold. These findings establish a thermodynamic framework illustrating how distinct DNA sequences reshape the conformational ensemble of a transcription factor, thereby tuning gene regulation beyond static binding affinity. Such sequence-dependent asymmetric recognition provides a mechanistic basis for the stabilization of gene regulatory networks through distinct energetic and entropic coupling. More broadly, this work has implications for synthetic circuit design and for the development of allosteric modulators targeting protein-nucleic acid and protein-protein complexes, potentially improving strategies for addressing traditionally “undruggable” targets.

Keywords: Operator DNA Sequences; Lac Repressor; Protein–DNA Recognition; Asymmetric Protein Dynamics; Conformational Thermodynamics; Molecular Dynamics Simulation

1. Introduction

Protein-DNA interactions (PDIs) play a central role to cellular function, governing gene regulation, DNA replication and repair, chromatin organization, and cell cycle control. PDIs are essential for controlling transcription and preserving genomic integrity. Dysregulation of PDIs is a hallmark of numerous diseases such as cancer, neurodevelopmental and immune related disorders. Although PDIs were historically considered “undruggable” due to the absence of well-defined binding pockets and the predominance of electrostatic interactions, recent advances have shifted this view. PDIs are now recognized as promising therapeutic targets. PDIs rely on both direct base readout and indirect DNA shape recognition. These interactions are stabilized by electrostatics interactions (salt bridges), hydrogen bonding, hydrophobic interactions, and stacking interaction. Multiple therapeutic strategies, which have been developed by targeting PDIs, can be categorized into (i) direct inhibition of DNA binding proteins, (ii) targeting DNA through sequence or structure specific recognition, and (iii) modulation of the protein-DNA interface via allosteric or indirect mechanisms in which small molecules bind distal sites on the protein or DNA to alter conformational dynamics, and reduce binding affinity indirectly. Because protein-DNA interfaces are inherently flexible, indirect strategies can exploit transient, druggable states, providing a promising route for selective and effective therapeutic intervention [1-2].

Transcription factors (TFs) are key regulatory proteins that modulate gene expression by binding to specific DNA sequences, thereby facilitating the activation or repression of target genes. Deciphering the molecular principles by which transcription factors (TFs) discriminate among DNA sequences is fundamental to understanding gene regulation. Although TFs bind multiple genomic loci, their regulatory outputs are highly sequence dependent and cannot be rationalized solely in terms of binding affinity. Instead specificity emerges from the coordinated interplay of several factors such as direct base contacts, DNA shape readout, conformational selection within dynamic protein-DNA ensembles, long range allosteric communication, cooperative interactions with additional regulatory factors, sequence-encoded modulation of protein dynamics, chromatin accessibility. However, the mechanistic basis by which these factors translate into precise sequence-specific transcriptional regulation remain unresolved [3].

In bacteria, the most common DNA binding motif is the helix-turn-helix (HTX). This motif typically interacts with inverted repeat (IR) DNA sequences. These sequences form palindromic structures, allowing TFs such as LacI, TetR, IclR to bind as dimers or tetramers in a head-to-head configuration. In contrast, direct repeat (DR) sequences require a head-to-tail binding mode. This type of interaction is observed in TFs such as Xis, Atox1, BldC. Over the past decade, several databases (e.g.; JASPAR, TRANSFAC, UniPROBE, RegulonDB, CollecTF) have supported the study for prediction of TF binding sites (TFBS). At the same time, many Artificial Intelligence (AI) guided approaches have been developed. DanQ combines convolutional neural networks (CNNs) with recurrent neural networks (RNNs) to predict TFBS. DeepBind, DeepSEA, and DESSO are based on convolutional neural networks (CNNs). In addition, FCNsignal uses fully convolutional neural networks (FCNs) to predict TF binding signals. Deep learning methods have shown strong performance in TFBS prediction. Machine learning (ML) models can accurately predict and classify TFBS. However traditional machine learning (ML) approaches are still less explored [4]. JASPAR 2026 significantly expands its open-access database of transcription factor (TF) binding profiles. It integrates deep learning based models to improve the accuracy and scalability of TF-DNA binding prediction [5].

Recent advances underscore the importance of integrating physical principles with data-driven approaches to better understand transcription factor (TF)-DNA recognition. Physics based machine learning frameworks incorporate explicit energetic descriptors such as electrostatic, van der Waals, and solvation contributions into statistical models, enabling both accurate prediction and mechanistic interpretability of binding. These approaches reveal that binding specificity arises not solely from DNA sequence but from the underlying energetic landscape of protein-DNA interactions. This approach provides physically meaningful feature attribution, identify key amino acid-nucleo base interactions, and overcome the limitations of conventional black-box models [6]. Wang et al. (2026) showed that sequence-based deep learning models, such as CNNs and self-attention transformers, can accurately predict transcription factor-DNA binding for both high and low affinity sites. These models perform better than traditional statistical methods. They can identify low-affinity binding sites and provide insights into how transcription factors recognize DNA, which helps in understanding gene regulation and designing proteins [7]. Combining convolutional neural networks (CNNs) with multi-head self attention, DeepCAC enables more effective modelling of transcription factor (TF)-DNA interactions [8]. DeepCTF shows that integrating DNA sequence information with shape features e.g.; minor groove width, helix twist), in an attention based deep learning framework improves the prediction of transcription factor (TF) binding specificity [9]. Barissi et al. (2022) introduced DNAAffinity, physics based machine learning method that predicts transcription factor binding affinities from DNA structural and mechanical properties derived from atomistic molecular dynamics simulations. This study underscores that binding affinity is influenced not only by direct base contacts but also by indirect readout mechanisms arising from DNA shape and energetic [10]. Sedhom et al. (2022) developed a method to align major-groove hydrogen bond patterns across DNA sequences that bind the same protein, revealing a conserved interaction which was hidden by sequence based analysis alone. By applying this approach to multiple binding sites of the Lac repressor and λ-CI repressor, they showed that key hydrogen bonds remain conserved even when the DNA sequence changes. This study highlights the importance of direct readout information in protein-DNA specificity, with implications for protein design and understanding evolutionary relationships [11]. Recent works demonstrate that the electrostatic properties of intrinsically disordered regions (IDRs) of transcription factor regulate efficient target search through various transient interactions [12-13]. Marklund et al. (2022) showed that DNA binding specificity is mainly determined by how efficiently proteins associate with different sequences. Modulating how fast a protein binds to DNA, rather than how fast it unbinds, reduces the chance of it getting stuck on non-target sequences during search process and suggests that binding rates play the main role in determining specificity in protein-DNA interactions [14].

The lactose repressor (LacI), a sequence-specific transcription factor that regulates the *lac* operon in E. Coli, act as a canonical system for investigating the relationship between sequence-specific DNA recognition and allosteric regulation supported by extensive structural, genetic, and thermodynamic characterization.

While biophysical techniques such as NMR, X-ray crystallography have elucidated the architectures of LacI-DNA complexes, these static structures do not completely capture the dynamic and energetic contributions that distinguish specific recognition from non-specific recognition. Within the cellular environment, LacI samples numerous non-specific DNA sites during its target-search process through several mechanisms such as sliding and diffusion before forming a stable complex at its cognate operator. Understanding the structural and thermodynamic distinctions between these binding mechanisms is therefore essential for explaining recognition specificity.

In the lac operon, the homo-tetrameric LacI repressor binds simultaneously to a primary operator O1 and auxiliary operators O2 and O3 to induce DNA looping, thereby increasing the local concentration of the repressornear the promoter, sterically hindering the access of RNA polymerase, and leads efficient transcriptional silencing. Deletion of auxiliary operators reduces repression dramatically, underscoring the importance of cooperative interactions.

LacI functions as a homo-tetramer (a dimer of dimers) composed of four identical 37.5 kDa monomers, each containing an N-terminal helix-turn-helix (HTH) DNA binding domain (residues 1-62), a core domain (residues 63-329) containing inducer binding site and dimerization interface, and a C-terminal tetramerization domain (residues 330-360). The core domain which is subdivided into N- and C-sub domains is connected to DNA binding domain through hinge helices (residues 50-57). Figure 1 (a) represents architecture of the operator bound tetrameric Lac repressor. Upon binding inducers such as allolatose or IPTG, LacI undergoes an allosteric conformational transition within the core domain that lowers operator/DNA binding affinity, and initiate transcription. N terminal helix-turn-helix motif of lac repressor recognizes the major groove of the operator DNA through base-specific hydrogen bonds. Hinge helices insert into the minor groove, inducing a 45^0 DNA bend through an induced-fit mechanism [15-37].

Fried et al., (2002) revealed that hydration plays significant role in the interaction between the lac repressor and operator DNA. Formation of the specific LacI-O1repressor-operator complex is accompanied by the release of 260 water molecules, producing a substantial entropy gain that drives the high-affinity repressor-operator complex formation while reducing solvent-accessible surface area and stabilizing the interface. Trapped interfacial water molecules mediate hydrogen bonding network that complement direct base-specific hydrogen bonding network. The thermodynamic basis of binding reflects compensation between enthalpic gains (direct and water-mediated interactions) and entropic effects from solvent rearrangement [38]. Barr and van der Vaart (2012)demonstrated through molecular dynamics simulation that the bent DNA conformationin the Lac repressor headpiece-O1 operator complex arises primarily from specific protein-DNA contacts and hydration thermodynamics rather than intrinsic DNA curvature alone [39].

Lüking et al. (2022) shed light on two distinct conformations which lac repressor uses during diffusion; (i) a flexible “search” conformation with disordered hinge regions that enables efficient one-dimensional scanning along straight DNA, and (ii) a helical “recognition” conformation that supports tight binding to bent specific operator DNA. During the search phase, the protein adopts a flexible conformation with disordered hinge regions, allowing lac repressor to remain mobileand facilitating rapid one-dimensional sliding along non-specific largely straight DNA without kinetic trapping. Upon encountering the specific operator sequences, these hinge regions undergo a disorder to order transition, forming α-helices that insert into the minor groove. This helical “recognition” conformation stabilizes lac repressor-operator complex. Thus, switching between disordered (search) and ordered helical (recognition) conformation underpins the efficiency of locating and binding the correct operator site [40]. Using all-atom molecular dynamics simulation, Liao et al., (2019) suggest that LacI preferentially binds to bent A-form DNA, stabilizing protein structure while reducing conformational entropy, with slightly increased hydrogen bonding in specific complexes relative to non-specific DNA [41]. Enhanced sampling indicates the cognate sequence stabilizes specific polar contacts, lowering the free energy of the specific complex while preserving a smooth energy landscape that enables rapid sliding without large conformational changes [42-43]. Xu et al. (2015), examined how the lac repressor discriminates between specific and non-specific DNA by integrating molecular dynamics simulation with correlation network analysis. Specific operator binding by the lac repressor generates a more centralized and densely connected interaction network, enhancing long-range

communication pathways between the DNA-binding domain and distal regions to stabilize allosteric networks. In contrast, non-specific complexes exhibit a more dynamic and weakly coordinated network governed by electrostatic and hydrogen-bonding interactions [44].

Sengupta et al. (2017) indicate that the high-free-energy transition state (TS) of Lac repressor-operator binding preserves the most electrostatic interactions while DNA phosphates remain hydrated and key aromatic and amide groups become buried [45]. Computational and experimental studies show that local protein flexibility, electrostatic interactions, and solvent effects play key roles in modulating LacI dynamics. The hinge helix strengthens non-specific DNA binding by enhancing electrostatic contacts and altering counterion distribution, thereby slowing dissociation [46-47].

Molecular dynamic simulations of full length LacI varying hydrostatic pressures reveal that elevated pressure prevents the transition from the DNA-unbound “open” to the DNA-bound “closed” conformation, similar to the effect of DNA or ligand binding, suggesting that hydration of specific residues, particularly near the hinge helix and domain interfaces, plays a significant role in stabilizing different functional conformers. Water mediated effects act as key contributors to LacI’s functional allostery [48]. High resolution NMR spectroscopy shows that the Lac repressor dimer exists in a dynamic equilibrium between two distinct conformations corresponding to DNA bound and inducer bound states. Binding of either effector shifts this equilibrium consistent with the Monod-Wyman-Changeux (MWC) allosteric model, where redistribution of pre-existing conformational ensembles mediates transcriptional switching [49].

Machine learning approaches significantly improve the reliable predictions how mutations affect LacI mediated transcriptional repression with higher accuracy than traditional approaches [50]. OptoLacI variants were created by coupling Lac repressor scaffold, with a light oxygen voltage 2(LOV2) photosensory domain, enabling direct light dependent control of gene expression without IPTG. Optogenetic engineering of LacI demonstrates precise light mediated control of DNA-binding activity, highlighting applications in synthetic biology without need for chemical inducers [51].

Xu et al. (2022) showed that LacI-mediated DNA loops act as topological barriers whose stability is modulated by transcription generated supercoiling. They demonstrated that torsional stress reduces the lifetime and stability of LacI induced loops, thereby weaking their ability to block RNA polymerase progression. These findings highlight a dynamic interplay between DNA topology and protein mediated looping, suggesting how mechanical strain modulates transcriptional repression in the lac operon [52]. Kinetic analyses support a multistep binding pathway with operator dependent energetic barriers and partial ordering intermediates. NMR, single molecule spectroscopy, and molecular dynamics simulations reveal that recognition mechanism between Lac repressor and operators is not a classical lock and key process but a dynamic multistep mechanisms involving direct and indirect base readout [53-57].

Despite these advances, the impact of operator sequences on distal LacI protein dynamics and intra protein communication remains incompletely understood. It is unclear whether symmetric LacI dimers respond equivalently to different operators or whether sequence-specific binding induces asymmetric dynamics between the monomers. Addressing this question demands residue-specific thermodynamic contribution capable of quantifying entropy redistribution and energetic coupling across the protein architecture.

The histogram based method (HBM) provides a rigorous framework for dissecting the conformational thermodynamic basis of bimolecular function across diverse systems (such as protein-peptide, protein-protein, protein-nucleic acid complex systems).This method can predict functional ligand binding sites in proteins as well as putative binding pockets for non-cognate

ligands targeting nucleic acids, providing structural and thermodynamic insights to guide the design of protein/nucleic acid targeted therapeutics.

Protein-protein interactions, whether in MDM2-p53 or SAP-SLAM complex, induce cooperative intra protein conformational changes that propagate allosteric effects beyond the immediate binding interface. HBM allows quantitative characterization of these conformational rearrangements by analyzing residue-specific thermodynamic changes, providing microscopic insight into structural and thermodynamics basis of ligand binding and allostery. Studies of membrane associated and pathogenic protein-protein interfaces further underscored the role of entropy redistribution in stabilizing transient yet functionally critical interactions.

Thermodynamic quantification using the HBM approach reveals how free energy and entropic contribution collectively govern stability of bio-molecular complex systems and ligand recognition, highlighting the method's ability to connect microscopic conformational dynamics to functional outcomes. Beyond proteins, HBM has been successfully extended to nucleic acid systems, where it has clearly explained the thermodynamic basis of metal ion specificity in DNA cleavage by restriction endonuclease, protein-induced Hoogsteen base pair stabilization, ligand recognition by riboswitch aptamer, and thermosensing behaviour of riboswitch aptamers. Conformational thermodynamics analysis of ligand binding studies of fluoride riboswitch aptamer has identified potential hot spots for RNA targeted antimicrobial drug design. In HBM approach, residue specific torsional probability distribution are extracted from equilibrium molecular dynamics trajectories, enabling direct computation of conformational entropy and free energy following principles of statistical mechanics [58-71]. This method has broad implication in understanding structure-function relationship of biomolecules by exploring conformational stability and order under different perturbation (temperature, pressure, solvent composition, pH, mutations).

The HBM offers distinct advantages over conventional enhanced sampling md simulation such as metadynamics or umbrella sampling, which depend on predefined collective variables (CVs) to accelerate rare conformational transitions. In protein-nucleic acid complex systems, conformational dynamics emerge from wide range microscopic conformational variables such as torsion angle, pseudo torsion angle of nucleic acid; backbone dihedrals and side chain rotamers of amino acid residues; solvent-mediated interactions, making it difficult to define a minimal and unbiased set of collective variables.Incorrect or suboptimal CV selection may introduce sampling bias, poor convergence, and reduced interpretability at the residue level. By removing the dependence on predefined reaction coordinates, HBM enables efficient and unbiased residue-specific characterization of conformational thermodynamics capturing contributions from free energy and entropy. Moreover, HBM complements traditional energy decomposition approach (MM/PBSA, MM/GBSA) by revealing detailed residue-level mechanisms of molecular recognition of the receptor-ligand complex that are often inaccessible through CV dependent enhanced sampling approach [72-76].

Figure 1 (b) represents two target sequences which are considered in this study (i) the naturally occurring pseudo-palindromic operator O2 and (ii) the synthetic operator SymL (perfect palindrome of the left half site of the natural operator O1). Complete twofold symmetry in O-SymL is achieved by deleting the central G:C base pair. O-SymL is widely used in biophysical studies because it binds the Lac repressor with exceptionally high affinity. In operator O2, left (consensus) and right (non-consensus) binding sites exhibit sequence asymmetry. Conserved bases are highlighted in red (left) and blue (right), with asymmetric regions in O2 underlined and italicized (Figure 1 (b)).

Molecular Modeling Connect
ISSN: 3105-3734
2026, Vol. 3, Cite as: doi:10.x/journal.x.x.x

In this study, region-specific thermodynamic changes in the Lac repressor protein (holo) were quantified relative to the unbound (apo) state using HBM. Negative conformational free energy changes (ΔG) and entropy contribution (TΔS) indicate energetic stabilization and increased conformational ordering, whereas positive values of ΔG and TΔS suggest enhanced destabilization and disorder. Further principal component analysis (PCA) is employed to characterize the dominant motions of LacI bound to operator O-SymL and O2 to validate the trends observed in HBM.

Representative snapshots from molecular dynamics simulations for unbound and bound Lac-SymL and Lac-O2 states are shown in Figure 2.

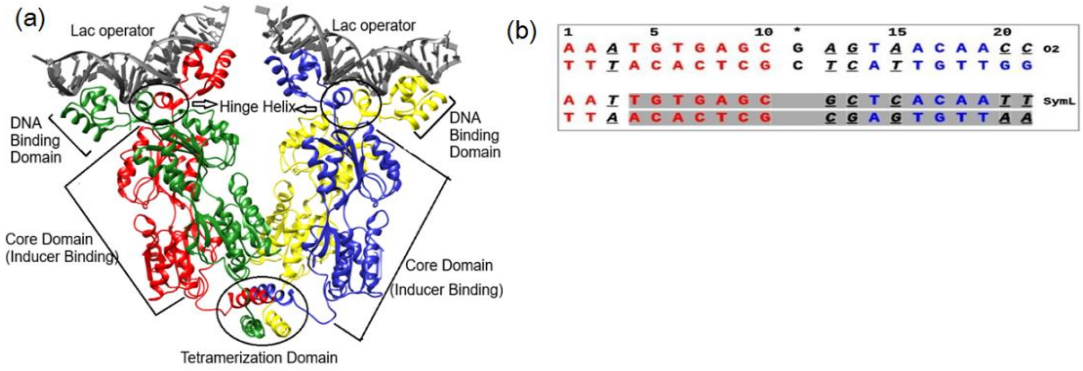

Figure 1(a) Ribbon representation of the tetrameric Lac repressor bound to two DNA operator sites, based on the 4.80 Å resolution crystal structure (PDB ID: 1LBG), visualized using UCSF Chimera. (b) Nucleotide sequences of the naturally occurring pseudo-palindromic O2 operator and the synthetic symmetric (palindromic) SymL operator.

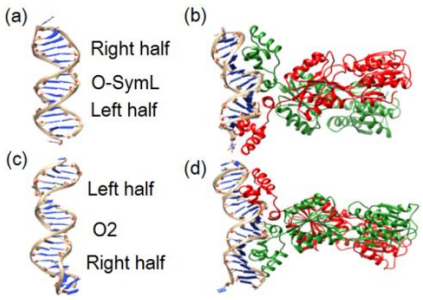

***Figure** 2 Representative average structures derived from molecular dynamics trajectories for the Lac repressor in the (a) lac repressor unbound operator O-SymL, (b) holo state corresponding to lac repressor bound O-SymL, (c) lac repressor unbound operator O2 system, and (d) holo state corresponding to lac repressor bound O2. This figure is reproduced from the author's original work published as a preprint titled "In-silico studies on structural and thermodynamic basis of interaction of lac repressor protein binding to different DNA operators"* Figure modified from Ref. [79].

Overall, this work provides direct thermodynamic evidence bridging side-chain dynamics tosequence-dependent allostery in LacI. Transcriptional repression is therefore governed not only by static binding affinity but also by dynamic redistribution of entropic and energetic contributions across the protein scaffold. By connecting subtle DNA sequence variations to global allosteric reprogramming, this study deepens mechanistic insight into transcription factor regulation and establishes a general framework for interpreting how sequence-encoded dynamics modulate function in diverse protein–DNA systems.

2. Method

2.1 Modeling Strategy

The coordinates of the Lac repressor dimer bound to operator DNA were obtained from the Protein Data Bank (PDB ID: 1EFA) [77], which contains the full-length Lac repressor. Chains A and B were retained for all simulations, while Chain C (residues 46–331) was removed to reduce system size and to maintain a symmetric dimeric architecture of the repressor. The co-crystallized anti-inducer ONPF present in the structure was excluded, as previous studies indicate that ONPF binding does not significantly perturb the architecture of the ligand-binding pocket or alter the overall conformation of the Lac repressor–DNA complex. All crystallographic water molecules were also removed. This processed structure was used as the model for the Lac–O-SymL repressor–operator complex.

Since no high-resolution X-ray or NMR structure is available for the Lac–O2 complex containing the full-length Lac repressor dimer (residues 2–329 encompassing both DNA-binding and core domains), a hybrid modeling approach was employed. The N-terminal DNA-binding domain of the Lac repressor dimer from 1EFA was structurally aligned with the corresponding domain of the NMR ensemble structure 2KEJ. The remaining DNA-binding region present in 2KEJ was subsequently removed to avoid redundancy. Among the NMR ensemble models of 2KEJ, model 8 identified by the WHAT IF server (https://swift.cmbi.umcn.nl/servers/html/index.html) as the closest to the average NMR structure was selected as the representative structure. The resulting Lac–O2 complex model was further subjected to energy minimization to remove steric clashes introduced during modeling. Minimization was performed using 5000 steps of steepest descent followed by 10,000 steps of conjugate gradient optimization in UCSF Chimera [78]. In the modeled Lac–O2 complex, Chain A of the Lac repressor interacts with the consensus half-site of the operator, whereas Chain B binds to the non-consensus half-site of the O2 operator. The free Lac repressor (operator-unbound state) was derived from the same PDB entry (1EFA) by removing the DNA molecule.

Model validation of Lac-O2 repressor-operator complex

The structural integrity and reliability of the modeled of Lac-O2 repressor-operator complex were evaluated using a multi criteria assessment to ensure conformational consistency, stereochemical quality, interface analysis, and structural stability prior to production simulations.

Structural Quality and Stereochemistry

The modelled complex was subjected to stereochemical validation using standard structure assessment tools (Procheck and MolProbity). Ramachandran plot analysis confirmed that the majority of the residues occupied favoured regions, with no significant clustering in disallowed conformational space. Rotamer distributions were within acceptable limits and no severe steric clashes were detected after energy minimization. Bond length and angle deviations were within standard tolerances for high quality protein structures suggesting modeling and minimization procedures did not introduce geometric artifacts.

Structural Conservation Relative to Template

To ensure preservation of the canonical LacI fold, the modeled Lac-02 complex was superimposed on to the reference Lac-O-SymL crystal structure (PDB ID: 1EFA). Backbone RMSD calculation for the N-terminal DNA binding domains yielded values below 0.5 Ådemonstrating retention of the helix-turn-helix architecture and hinge helix positioning. The core domains also exhibited minimal deviation, confirming that global structural features of the Lac repressor dimer were conserved during model construction. DNA bending and minor groove geometry remained consistent with known LacI induced deformation patterns, reinforcing the structural and functional validity of the model [79].

Protein-DNA Interface Validation

The structural validity of the modelled LacI-O2 complexrepressor-operator complex model was further assessed by comparing hydrogen bonds, salt bridges, and van der Waals contacts between 8th conformer of the NMR ensemble of 2KEJ and the modelled LacI-O2 complex at the protein-DNA recognition interface.Major groove base-specific contacts, characteristic of LacI recognition were preserved, particularly involving residues within the helix-turn-helix motif and hinge-helix region. The asymmetric engagement of the two subunits of Lac repressor, where chain A binds the consensus half-site and chain B interacts with the non-consensus half-site was structurally consistent with the sequence asymmetry of the operator O2.

Hybrid modelling strategies are well established [57, 68].

Following the validated protocol ensures that observed sequence-dependent conformational differences reflect intrinsic LacI-DNA dynamics rather than modelling artifacts.

2.2 Molecular Dynamic Simulation Protocol

All-atom molecular dynamics simulations were performed using GROMACS 5.1 for the repressor-operator complexes (Lac-O-SymL and Lac-O2) and operator unbound Lac repressor protein (Table 1) [80]. The proteins and DNAs were parameterized with the Amber ff99SB-ILDN and Amber ParmBSC1 force field, respectively, a validated combination that improves protein side-chain dynamics and corrects DNA backbone conformation.

Table 1: Systems taken for molecular dynamics simulation

System Number	System taken for analysis	Initial Structure	DNA sequence used for MD simulation	Presence of Protein	Duration of MD simulation	Number of solvent	Number of Na^+	Force Field	Average RMSD in Å with mean and SD (within first bracket)
1	Lactose repressor-Operator SymL	As in PDB ID 1EFA	d(TGTGAGCGCTCACAATT).d(AATTGTGAGCGCTCACA)	Yes	500ns	51750	38	Amber ff99SB-ILDN force field for protein and Amber ParmBSC1 force field for DNA	2.0 (0.02)
2	Lactose repressor-operator O2	Model structure	d(AAATGTGAGCGAGTAACAACC).d(GGTTGTTACTCGCTCACATTT)	Yes	500ns	55844	50	Amber ff99SB-ILDN force field for protein and Amber ParmBSC1 force field for DNA	3.4 (0,01)
3	Lactose repressor dimer	As in PDB ID 1EFA without DNA	Not Applicable	Yes	500ns	41017	6	Amber ff99SB-ILDN force field for protein	5.4 (0.10)

The Amber ff99SB-ILDN force field was used to model the proteins due to its improved backbone and side-chain torsion parameters which enhance structural accuracy, reproduce experimental NMR observables, and ensure stability in long time-scale molecular dynamics simulation. DNA was modeled using the Amber ParmBSC1force field, which corrects known artifacts of earlier DNA force fields such as excessive terminal base pair fraying, inaccurate BI/BII backbone balance,and underestimation of DNA helical parameters and enables stable micro-second time-scale simulations of both canonical B-DNA and protein bound DNA [41, 68, 81-85].

In case of His residues, the Nε is considered as secondary amino nitrogen and Nδ as imino nitrogen.The systems were solvated TIP3P water model in cubic box using a distance of 1 nm between the complex and the edge of the box and neutralized with required number of Na^+ and Cl^- ions. Periodic boundary conditions were applied in all directions. Energy minimization was performed using the steepest descent method to relieve steric clashes. Each system was equilibrated for 0.5 ns under NVT equilibration, followed by 1.0 ns under NPT equilibration at 300K temperature and 1 atm pressure, with temperature controlled by the Berendsen thermostat and pressure by the Parrinello-Rahman barostat. Lennard-Jones and

short-range electrostatic interactions were truncated at 1.0 nm, and long range electrostics were treated with the Particle Mesh Ewald (PME) method. All bonds involving hydrogen atoms were constrained using the LINCS algorithm. After equilibration, the production runs were performed for 500 ns with a time step of 2 fs.The atomic coordinates were saved every 10.0 ps, and thus, 50,000 structures were collected for further analysis [86-87].

The Protein-DNA complex systems ((Lac-OSymL) and (Lac-O2) contains 165000 and 179000 atoms respectively.

Three independent simulations per system were performed using different randomized velocity seeds. Each trajectory was simulated for 500 ns under identical equilibration and production protocols. All analyses represent averages over three independent molecular dynamics simulation trajectories.

2.3 Analysis of Structure and Conformational Stability

Periodic boundary condition (PBC) artifacts frequently resulted in apparent spatial separation of the two DNA strands and the two protein subunits during the simulations. This issue persisted across all systems and could not be fully resolved using standard trajectory rewrapping procedures, such as “PBC MOL” option in GROMACS or the wrapping functions in VMD. To address this, each trajectory was converted to PDB format using the GROMACS utility trjconv, followed by post-processing with an in-house FORTRAN program to reconstruct continuous protein–DNA complexes for subsequent analysis.

Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analyses were performed using GROMACS on the corrected trajectories. Protein–DNA hydrogen bonds were identified using a modified version of *pyrHBfind* [88], applying a donor–acceptor distance cutoff of 3.0 Å and a minimum hydrogen bond angle of 150°.

Side-chain χ1 torsion angle distributions were analyzed using VMD. DNA conformational properties were characterized in terms of inter-base pair and base-pair step parameters using the NUPARM software package [89].

Structural stability and convergence of LacI-O-SymL complex were confirmed by backbone RMSD with reference to the corresponding experimental crystal structure and RMSF of protein and DNA backbones. Backbone RMSD of LacI-O-SymL and LacI-O2 are stabilized within 2.0 Å and 4.0 Å respectively. The helix-turn-helix motifs maintain stable insertion within the DNA major groove. No unfolding events were observed. Preservation of key protein-DNA contacts and DNA helical parameters through the simulation confirms the stability of LacI-O2 repressor-operator model.

2.4 Conformational Thermodynamics from Histogram Based Methods (HBM)

The complete protocol of the histogram-based method (HBM) for calculation of the conformational thermodynamics derived from equilibrated molecular dynamics trajectory is described in preceding publications [58-71]. The Histogram of $\chi 1$ for each residue is generated within 0° to 360° and divided into 90 equal bins. The histograms illustrate the probability of obtaining the system in a given conformation. Free energy and conformational entropy of the systems are calculated following the Boltzmann factors and Gibbs formula respectively. The normalized probability distribution of any conformational variable such as χ in complex states (i.e. operator bound) and free states (i.e. operator unbound) are represented by $H_i^{complex}(\chi)$ and $H_i^{free}(\chi)$ respectively.

The change in free energy for microscopic conformational variable $\chi 1$ due to binding of the repressor protein with operator DNA is defined by the following formula:

$$\Delta G_i^{conf}(\chi) = -k_B T ln\,(H_{max,i}^{complex}(\chi)/H_{max,i}^{free}(\chi))$$

Where 'max' indicates the peak value of the histogram and i represents each of the lac repressor protein residues.

The change in side chain conformational entropy for χ_1 due to binding of the repressor protein with operator DNA is expressed by the following formula:

$$T\Delta S_i^{conf}(\chi) = -k_B T\left[\sum_j H_i^{complex}(\chi) ln H_i^{complex}(\chi) - \sum_j H_i^{free}(\chi) ln H_i^{free}(\chi)\right]$$

Where the sum is taken over all histogram bins j and i represent each of the amino acid residues of the lac repressor protein. $k_B$and T indicate Boltzmann constant and temperature, respectively.

The changes in conformational thermodynamics of chain A and chain B lac repressor (operator bound) are computed with respect to that of chain A and chain B lac repressor (operator unbound), respectively.

Conformational thermodynamic quantities were derived from the χ_1 torsional dergree of freedom, which captures dominant rotameric transitions of interface residues as well as distal residues far from theprotein-DNA interface. The $\chi_1$torsion angle defines the primary rotameric state governing side-chain orientation relative to the backbone and directly modulates intermolecular contacts. While higher order side-chain torsion angle (χ_2, χ_3, χ_4) and backbone (φ/ψ) dihedrals may also contribute precise conformational adjustments; however, on the sampled timescales their contributions are generally smaller or indirectly coupled to χ_1, consistent with the restricted backbone flexibility of the DNA-binding domain.

All the reported thermodynamic quantities (ΔG and TΔS) represent the average of the three independent molecular dynamics trajectories per system, with uncertainties estimated as the standard error of the mean across the replicas.

2.5 Principal Component Analysis (PCA) of Molecular Dynamics Simulation Trajectories

Principal Component Analysis, when integrated with Molecular Dynamics Simulation, transforms high-dimensional trajectory data into a compact set of orthogonal modes representing the most significant collective motions. This is achieved by constructing and diagonalizing the covariance matrix (C_{ij}) of atomic positional fluctuations, thereby projecting the dataset onto a lower-dimensional space that captures the key dynamical features of the protein–DNA complex during the simulation [90–93]. The dominant principal components typically correspond to slow, large-scale conformational transitions, indicating that functionally relevant motions are largely confined to a low-dimensional subspace, where a limited number of eigenvectors account for the most of the conformational variance.

The elements C_{ij} in the covariance matrix are defined by C_{ij}= < (x_i-<x_i>) (x_j-<x_j>) >where <x_i>, <x_j> are the mass-weighted mean Cartesian coordinates of the atoms present in the system and x_i (or x_j) is the coordinate of the i^{th} (or j^{th}) atom of the systems, and "<>" indicates ensemble average. Diagonalization of the covariance matrix yields a set of eigenvectors and corresponding eigenvalues, where the eigenvectors define the directions of correlated atomic displacements and the eigenvalues quantify the magnitude of fluctuations along these directions. The principal components (PCs), ranked by decreasing eigenvalues, therefore capture the dominant motions of the system, with the first few PCs typically representing slow, large-scale collective movements associated with conformational transitions of the complex. The mass-weighted covariance matrix was constructed and diagonalized using the gmx covar module of the GROMACS simulation package, and the resulting eigenvectors were

further analyzed using gmx anaeig. To reduce computational cost while preserving essential global dynamics, only the Cα atoms of the protein and the phosphate atoms of DNA were included in the analysis, given that the dimensionality of the covariance matrix scales quadratically with the number of atoms. The dominant collective fluctuations involving multiple regions of the macromolecule correspond to large-scale "breathing" motions. The extreme projections along the first principal component (PC1) were identified to obtain representative conformations corresponding to minimum and maximum amplitudes (hereafter referred to as PC1-first and PC1-last frames), which define the endpoints of motion along this mode. Porcupine plots were generated using an in-house FORTRAN program, which utilizes these endpoint structures (in PDB format) to compute displacement vectors. These vectors depict both the direction and magnitude of atomic displacements associated with the principal mode of motion.

The Pymol, UCSF Chimera and Visual Molecular Dynamics (VMD) packages were used for visual assessment of the trajectory files and to generate images. Graphs were produced using Gnuplot, Microsoft Excel and Python scripts.

3. Result

3.1 Assessment of MD Simulation Convergence

Structural stability and conformational variability of the Lac repressor were assessed by calculating the root mean square deviations (RMSDs) of Cα atoms for both dimeric protein and DNA in operator-bound and unbound states relative to their initial structures (Figure S1). RMSD analysis shows that the DNA-bound systems (Lac-O-SymL and Lac-O2) are significantly more stable than the free protein, indicating that binding to operator DNA constrains the Lac repressor structure. All systems reach a plateau after ~100-150 ns, suggesting that the simulations are sufficiently long to capture the equilibrated dynamics of LacI in different states. Simulations of the free protein dimer reveal that RMSD values continue to fluctuate over time, reflecting slower relaxation and higher conformational variability in the absence of DNA. This RMSD analysis demonstrates that DNA binding not only stabilizes the LacI conformation but also reveals operator-specific effects, with synthetic symmetric sequences (O-SymL) producing a more rigid protein conformation than the natural operator O2. These differences could underlie variations in binding affinity and specificity observed experimentally. Based on these observations, snapshots from 250 ns to 500 ns were selected for further analyses of DNA-protein interactions, bending, groove dynamics, conformational thermodynamics, and essential dynamics, as all systems reached structural equilibrium by 250 ns.

3.2 Comparative Hydrogen Bonding Interaction Analysis of Lac Repressor Protein with O-Sym and O2 DNA operators

A comparison of hydrogen-bond occupancies over the 250–500 ns molecular dynamics simulation, using a donor–acceptor distance ≤ 3.5 Å and angle ≥ 120° and considering only interactions with ≥ 50% occupancy, reveals a clear asymmetry in Lac repressor–DNA recognition. Residue numbering follows that of Figure 1. Table 2: (a) and (b) represent the hydrogen bonding interactions between the Lac repressor and different DNA operators were analyzed.

Table 2: (a) Hydrogen bonds between Lac repressor and operator O-SymL, identified from the MD trajectory (250ns-500ns) using Amber-ff99SB-ILDN force field for protein and Amber ParmBSC1 force field for DNA. For DNA, chain D corresponds to the first strand and chain E corresponds to the second stand.

Donor	Donor Atoms	Acceptor	Acceptor Atoms	% of Occurrence
Leu-6A	N	Cyt-13E	O2P	100.00

Tyr-7A	OH	Cyt-13E	N4	75.00
Ser-16A	OG	Gua-5D	O2P	100.00
Ser-16A	N	Gua-5D	O2P	100.00
Tyr-17A	OH	Gua-7D	O6	100.00
Tyr-17A	OH	Gua-7D	N7	78.00
Gln-18A	NE2	Gua-5D	O6	95.00
Thr-19A	OG1	Gua-5D	O2P	100.00
Ser-21A	OG	Thy-14E	O2P	100.00
Asn-25A	ND2	Thy-14E	O1P	79.00
Tyr-47A	OH	Cyt-13E	O2P	100.00
Asn-50A	N	Cyt-13E	O5’	98.00
Gln-54A	NE2	Thy-14E	O1P	100.00
Lys-59A	NZ	Cyt-15E	O1P	69.00
Leu-6B	N	Cyt-13D	O2P	100.00
Ser-16B	OG	Gua-5E	O2P	100.00
Ser-16B	N	Gua-5E	O2P	100.00
Ser-16B	OG	Gua-5E	O5’	77.29
Tyr-17B	OH	Gua-7E	N7	74.00
Tyr-17B	OH	Gua-7E	O6	100.00
Thr-19B	OG1	Gua-5E	O2P	100.00
Arg-22B	NH2	Gua-5E	O6	100.00
Arg-22B	NH1	Gua-5E	N7	99.00
Ser-31B	N	Thy-4E	O2P	100.00
Thr-34B	OG1	Thy-4E	O1P	98.00
Asn-50B	ND2	Cyt-13D	O1P	74.00
Gln-54B	NE2	Thy-14D	O1P	97.00

Table 2: (b) Hydrogen bonds between the Lac repressor and operator O2, identified from the MD trajectory (250ns-500ns) using Amber-ff99SB-ILDN force field for protein and Amber ParmBSC1 force field for DNA. For DNA, chain D corresponds to the first strand and chain E corresponds to the second stand.

Donor	Donor Atoms	Acceptor	Acceptor Atoms	% of Occurrence
Leu-6A	N	Cyt-13E	O2P	100.00
Tyr-7A	OH	Cyt-13E	N4	98.50
Ser-16A	N	Gua-5D	O2P	100.00
Ser-16A	OG	Gua-5D	O2P	100.00
Tyr-17A	OH	Gua-7D	O6	100.00
Tyr-17A	OH	Gua-7D	N7	80.30
Thr-19A	OG1	Gua-5D	O2P	100.00
Ser-21A	OG	Thy-14E	O2P	88.00
Arg-22A	NH1	Gua-5D	N7	79.70
Arg-22A	NH2	Gua-5D	O6	86.26
Asn-25A	ND2	Thy-14E	O1P	81.00
Asn-25A	ND2	Thy-14E	O2P	77.00
Ser-31A	N	Thy-4D	O2P	100.00
Thr-34A	OG1	Thy-4D	O1P	99.00
Tyr-47A	OH	Cyt-13E	O2P	77.00
Asn-50A	N	Cyt-13E	O1P	94.00
Asn-50A	N	Cyt-13E	O5’	75.00
Gln-54A	NE2	Thy-14E	O1P	95.00
Lys-59A	NZ	Thy-14E	O1P	98.00
Gln-60A	N	Cyt-15E	O1P	97.00
Arg-118A	NH1	Gua-11D	O1P	65.00
Arg-118A	NH2	Gua-11D	O1P	54.00
Leu-6B	N	Ade-12D	O2P	99.00
Ser-16B	N	Gua-5E	O2P	100.00
Ser-16B	OG	Gua-5E	O2P	94.00
Tyr-17B	OH	Gua-13D	N7	64.00
Gln-18B	NE2	Thy-6E	O4	90.00
Thr-19B	OG1	Gua-5E	O2P	100.00
Arg-22B	NH1	Gua-5E	N7	100.00
Arg-22B	NH2	Gua-5E	O6	100.00
Asn-25B	ND2	Gua-13D	O2P	66.00

Ser-31B	N	Thy-4E	O2P	100.00
Ser-31B	OG	Thy-4E	O1P	74.00
Thr-34B	OG1	Thy-4E	O1P	93.00
Tyr-47B	OH	Ade-12D	O1P	100.00
Asn-50B	N	Ade-12D	O1P	98.00
Gln-54B	NE2	Gua-13D	O1P	89.00

Despite the intrinsic symmetry of the Lac repressor dimer and the palindromic nature of the O-SymL operator, hydrogen-bond occupancy analysis (Table 2a) reveals pronounced asymmetry in protein–DNA interactions. The two monomers do not contribute equivalently to DNA recognition, indicating that sequence symmetry does not translate into a symmetric interaction landscape at the protein–DNA interface. In the O-SymL complex, Chain A is characterized by a combination of base-specific and extensive backbone-mediated interactions, with a strong preference for phosphate contacts. Residues such as Ser16A, Thr19A, and Gln54A exhibit highly persistent occupancies, forming a dense interaction network primarily localized on strand E, while additional contacts from Tyr7A (75%), Asn25A (79%), and Lys59A (69%) further stabilize the consensus half-site. Notably, Tyr17A forms a fully persistent hydrogen bond with Gua7D (O6/N7; ~78–100%), and Gln18A establishes a highly stable base-specific interaction with Gua5D (O6; ~95%), underscoring their roles in sequence-specific recognition. In contrast, Chain B displays a more structured and diverse interaction profile, combining strong base-specific recognition with multiple backbone contacts. This is dominated by Arg22B, which forms highly stable bifurcated hydrogen bonds with Gua5E (O6/N7; ~99–100%), supported by Tyr17B interactions with Gua7E (O6/N7; ~74–100%), indicative of dominant major groove readout. Additionally, Chain B exhibits extensive backbone interactions, including Leu6B - Cyt13D (O2P; 100%), Ser16B - Gua5E (O2P/O5'; ~77–100%), Thr19B - Gua5E (O2P; 100%), and unique contacts with Thy4E via Ser31B and Thr34B (O2P/O1P; ~98–100%). Furthermore, Chain B engages both DNA strands through Asn50B - Cyt13D (O1P; ~74%) and Gln54B - Thy14D (O1P; ~97%), indicating a broader and more spatially distributed interaction interface. However, several interactions expected to be symmetric particularly those involving Gln18 are diminished or absent in Chain B, reflecting a breakdown of ideal symmetry under dynamic conditions.

In contrast, the O2 complex exhibits a more heterogeneous and adaptive interaction network (Table 2b), driven by its intrinsic sequence asymmetry. In Chain A, the interaction pattern is dominated by a highly stabilized and functionally coherent network combining strong phosphate backbone anchoring with selective base readout on strand E. Residues Ser16A, Thr19A, Ser31A, and Thr34A form near-persistent contacts with phosphate groups (O2P/O1P; ~99–100%), establishing a rigid electrostatic scaffold for DNA binding. Superimposed on this backbone framework, Tyr7A (98.5%) contributes a strong base-specific contact with Cyt13E (N4), while Tyr17A maintains fully persistent interactions with Gua7D (O6/N7; ~80–100%), indicating conserved major groove recognition. Additionally, Arg22A provides bifurcated base readout through stable interactions with Gua5D (N7/O6; ~79–86%), and Asn25A engages Thy14E via dual phosphate contacts (O1P/O2P; ~77–81%), further reinforcing half-site specificity. Supporting stabilization is also provided by Gln54A (95%) and Lys59A (98%) toward Thy14E, along with Gln60A (97%) interacting with Cyt15E, collectively defining a dense and highly cooperative interaction network. Notably, Arg118A contributes weaker but persistent long-range phosphate contacts with Gua11D, suggesting auxiliary stabilization beyond the core recognition interface. On the other hand, Chain B exhibits a more distributed yet highly stable interaction architecture, integrating robust base recognition with extensive backbone engagement across both DNA strands. Arg22B serves as the principal specificity determinant, forming fully persistent bidentate hydrogen bonds with Gua5E (N7/O6; 100%), strongly

supported by Thr19B (100%) and Ser16B (94–100%) contacts with the phosphate backbone, thereby reinforcing strand E recognition. Additional sequence readout is mediated by Gln18B, which interacts with Thy6E (O4; 90%), and Tyr17B, which forms a weaker but persistent contact with Gua13D (N7; 64%), indicating asymmetry in major groove engagement relative to Chain A. Backbone stabilization is further extended through Ser31B and Thr34B interactions with Thy4E (O2P/O1P; 74–100%), while Tyr47B and Asn50B establish highly persistent phosphate contacts with Ade12D (100% and 98%, respectively), suggesting a compensatory stabilization strategy on strand D. Moreover, Gln54B contributes additional phosphate recognition with Gua13D (89%), reinforcing cross-strand stabilization.

Thus, the two subunits differ not only in interaction strength but also in residue utilization and strand preference. Gln60A forms a highly stable hydrogen bond with Cyt-15E (~97%), while Arg118A contributes additional phosphate backbone contacts in non-consensus regions. These interactions collectively expand the recognition interface to include residues beyond the helix–turn–helix (HTH) motif, such as Tyr47 and Asn50, thereby enhancing binding adaptability. The absence of a central GC base pair in O-SymL and the intrinsic asymmetry of the O2 operator result in distinct binding modes and interaction distributions. While O-SymL exhibits partial symmetry breaking at the dynamic level, O2 demonstrates a more flexible and distributed interaction network that accommodates sequence variability through compensatory stabilization. Collectively, these findings demonstrate that Lac repressor–DNA recognition is governed by a combination of stable phosphate backbone anchoring, sequence-specific base readout, and dynamically redistributed auxiliary contacts. The results underscore the critical role of DNA sequence context in shaping protein–DNA interaction landscapes beyond static structural models, emphasizing the importance of dynamic reorganization during molecular recognition.

3.3 Modulation of Protein Side-Chain Dynamics upon Operator DNA Binding

Figure S2 shows that DNA binding changes the Cα fluctuations differently in the two subunits of the Lac repressor dimer, indicating unequal stabilization of the chains upon operator binding. In the Lac-SymL complex, RMSF analysis of DNA indicates minimal fluctuations in bases forming hydrogen bonds with amino acids, while enhanced dynamics are observed in residues His29, Ser31, Ala32, Asn46, Gly58, Lys59, Gln60, Leu61, Arg101, Ser102, Gly103, Asn142, Gly236, Ile237, Gly310, Gln311, Ala312, Val313, and Lys314. Notably, the most pronounced increase in dynamicity upon binding of both SymL and O2 operators occurs in hinge, turn, and coil regions. In the operator-unbound protein, loops exhibit high mobility that is reduced upon DNA binding, indicating stabilization by the protein–DNA interaction. Both monomers insert the hydrophobic side chain of Leu56 into the DNA minor groove, unstacking two contiguous central CpG base pairs and introducing a kink. This van der Waals interaction is stable throughout simulations. RMSF values for Leu56 are markedly reduced in the DNA-bound state (SymL bound: chain A 1.029 Å, chain B 1.0 Å; O2 bound: chain A 0.943 Å, chain B 0.848 Å) compared to the free protein (chain A 1.57 Å, chain B 1.62 Å), confirming the stabilization of the kink region upon binding. Residues in the N- and C-terminal core domains of the Lac-O2 complex, particularly chain A (residues ~75–230), exhibit enhanced fluctuations, whereas the Lac-SymL complex shows no significant alteration in dynamics. Asymmetric dynamics in the operator-unbound protein are also observed at residues 285–310, distal from the DNA-binding interface. These findings indicate that O2 binding induces a distinctly different dynamic behaviour in chain A compared to SymL, suggesting that sequence-dependent differences in overall protein dynamics may play a critical role in coordinating the C-terminal domains during tetramerization.

To assess whether the observed asymmetric conformational dynamics of Lac repressor protein are robust and not artifacts of force field selection, molecular dynamics simulations have been repeated using an alternative protein force field, Amber ff14SB, in addition to Amber ff99SB-ILDN. Residue wise C-α fluctuation of Lac repressor chains A and B are systematically compared for both the operator O2 bound and O-Sym bound complexes using two different force fields (Figure S3). The RMSF values of the O-Sym and O2 bound Lac repressor complexes demonstrate clear operator dependent modulation of protein dynamics and this trend remains consistent across the force fields examined.

In the Lac-O-Sym repressor-operator complex, fluctuations are moderately distributed, with distinct peaks localized to the N-terminal DNA binding domain (residues 1-62) and a pronounced enhancement near the hinge region (residues 50-62), while the core domain displays comparatively restrained yet spatially heterogeneous motions. Both monomers exhibit closely comparable fluctuation amplitude suggesting a symmetric dynamic response upon binding symmetric operator. In contrast, Lac-O2 repressor-operator complex shows enhanced RMSF values within the DNA binding domain, particularly in one monomer, reflecting enhanced local flexibility and dynamic asymmetry. The hinge region of Lac-O2 system remains dynamically prominent but presents less sharply defined broader peaks, consistent with reduced conformational constraint relative to Lac-O-Sym system. Beyond residue 100, the core domain of Lac-O2 repressor-operator reveals generally lower flexibility, with a few localized peaks at specific positions, suggesting sequence-specific redistribution of flexibility rather than uniform or global conformational stabilization.

The strong agreement between two independently parameterized protein force fields demonstrates that the observed conformational dynamics are robust, reproducible, and not artifacts of force-field selection, therefore **reinforcing the validity of** the mechanistic interpretations derived from MD simulations.

3.4 Conformational Thermodynamics of Binding from Side-Chain χ1 Dihedral Angle Histograms

To assess residue-level conformational fluctuations of the Lac repressor, histograms of the side-chain dihedral angle χ_1 were generated for all residues from MD trajectories in both protein-bound and unbound states. Conformational analysis was performed over the 250-500 ns interval of each trajectory, during which χ_1 torsional distributions were stationary and side-chain rotamer transitions produced stable probability distribution.

Representative distributions (Figure 3 (a) and 3 (b)) show multimodal as well as unimodal peaks, reflecting distinct rotameric states and differing side-chain flexibility across residues. Sharper unimodal peaks are observed for chain A His29 (O2 complex), chain B Asp149 (free, O2), chain A Asp219 (free, O-Sym), and chain A Ser224 (free, O-Sym), indicating relative rigidity in these residues. In contrast, bimodal distributions, such as those for chain B Lys84 (free, O-Sym) and chain B Asp149 (O2 complex), reveal increased flexibility.

The comparatively sharp and narrow peaks suggest constrains the conformational dynamics leading to a more rigid and ordered conformational ensemble. In contrast, broadened peaks are indicative of pronounced conformational heterogeneity and dynamic fluctuations. The presence of bimodal distributions further implies inter-conversion between two distinct isomeric conformations suggesting the coexistence of, or dynamic exchange between two discrete isomeric conformations separated by an energy barrier. On the other hand, sharp unimodal peaks are consistent with dominance of a single, energetically favoured conformer, suggesting conformational stabilization and minimized stochastic fluctuations.

Overall, variations in the number, height, and width of peaks demonstrate that Lac repressor exhibits sequence-specific and site-dependent differences in side-chain conformational fluctuations between free and operator-bound states.

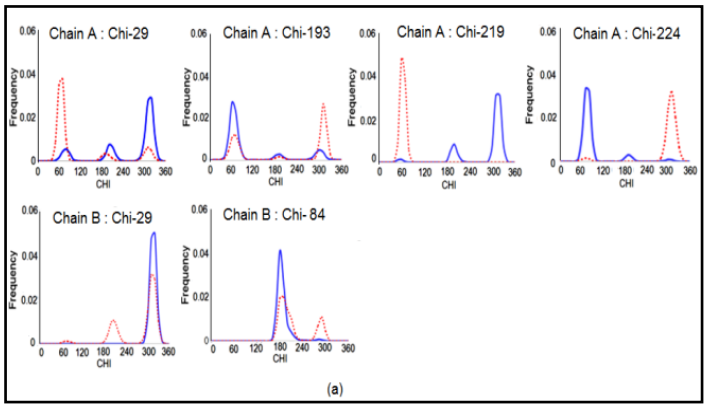

Figure 3(a): *Multimodal histogram distributions of the side-chain χ1 torsion angles for selected residues of the Lac repressor dimer in the absence and presence of operator DNA (O-Sym). Chain A residues include His29, Ser193, Asp219, and Ser224, while chain B includes His29 and Lys84. The apo state is represented by red dotted lines, and the DNA-bound holo state by blue lines.*

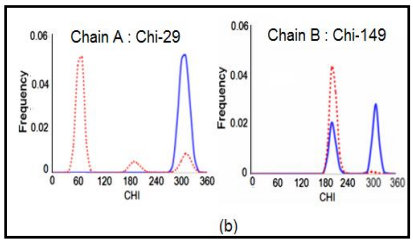

Figure 3: (b) χ1 side-chain torsion angle distributions highlighting distinct rotameric states in the Lac repressor bound to operator DNA (O2), shown for His29 in chain A and Asp149 in chain B. Figure 3 (a) and (b) are reproduced from the author's Ph.D. thesis [Das, S. (2023). "Computational Study Of Structure And Dynamics of Protein-DNAComplexes" (Doctoral thesis). University of Calcutta. *http://hdl.handle.net/10603/508120] and a preprint [Das, S., Roy, S. and Bhattacharyya, D. (2024), "In-silico studies on structural and thermodynamics basis of interaction of lac repressor protein binding to different DNA operators." https://doi.org/10.1101/2024.07.07.602393] available on bioRxiv.*

Based on these observations, residue-wise and domain-wise changes in conformational free energy (ΔG) and entropy (TΔS) were computed using a histogram-based method (HBM) that sums χ1 contributions across residues within specific regions (Figure S4 (a–d)). Negative ΔG and TΔS values correspond to conformational stabilization and increased order upon DNA binding, whereas positive values indicate destabilization and enhanced disorder. The results indicate that binding to both O-Sym and O2 induces conformational stabilization in specific residues, particularly within regions proximal to the DNA-binding interface.

Interestingly, enhanced segmental motions are observed in the core domains (N-terminal and C-terminal sub-domains) of the Lac repressor in the O2-bound state compared to O-Sym, particularly around the protein-protein interaction interface. These motions are more prominently captured by conformational entropy changes (Figure S4 (b)) than by free energy changes (Figure S4 (a)), highlighting the increased flexibility of distal regions induced by operator-specific binding. Table S1 quantitatively illustrates how O-Sym and O2 differently modulate residue-specific conformational thermodynamics across secondary structural elements of the homo-dimeric Lac repressor. Collectively, these analyses demonstrate that side-chain

conformational dynamics and thermodynamic responses of the Lac repressor are highly dependent on the bound DNA sequence, with O2 promoting greater distal flexibility, whereas O-Sym induces more localized stabilization.

Two helices (α2 helix of lac repressor (O-Sym bound) and α4 helix lac repressor (both O-Sym and O2 bound)) as well as two loops (loop3 of chain B lac repressor (O2 bound) ΔG= -7.46 kJ/mol, TΔS= -10.74 kJ/mol and loop5 chain A lac repressor (operator O-Sym bound) ΔG= -7.21 kJ/mol, TΔS= -2.23 kJ/mol and (operator O2 bound) ΔG= -6.14 kJ/mol, TΔS= -10.33 kJ/mol) of DNA binding domain undergo significant energetically stabilized and entropically ordered state upon binding to operators due to formation of an array of hydrogen bonding interaction. α2 helix shows conformational free energy change of -6.40 kJ/mol and -10.83 kJ/mol for chain A and B of lac repressor (O-Sym bound) respectively and corresponding entropy cost include -22.04 kJ/mol and -15.09 kJ/mol respectively. In case of α4 helix, change of conformational free energy comprised of -14.58 kJ/mol, -3.38 kJ/mol for chain A and B of lac repressor (operator O-Sym bound) as well as -8.87 kJ/mol, -10.01 kJ/mol for chain A and B of lac repressor (operator O2 bound) respectively. α4 helix shows conformational entropy change of -11.91 kJ/mol, -23.55 kJ/mol for chain A and B of lac repressor (operator O-Sym bound) as well as -19.17 kJ/mol, -19.71 kJ/mol for chain A and B of lac repressor (operator O2 bound) respectively. Most of the N-terminal residues (DNA binding domain) of both dimers in Lac repressor-O-SymL show a reduction in free energy and entropy due to stable sequence specific H-bond with DNA. But few residues of the DNA binding loop region (Pro3, Gln26, Ser31, Ile48) in contrast show enhancement of conformational entropy in the operator bound state. Some of the residues located in core domain (the N and C-terminal sub domains) of different chains of lac repressor (Asp149, Thr154 and Asn157 of loop12), (Asp219 of loop16), (237 Ile of loop17), (Val111, Leu114, Leu115 of α-Helix6), (163His, 165Asp, 171Val of α-Helix8), (Val192, Arg197, Tyr204 of α-Helix9), (Leu296, Gln298, Thr299, Ser300, Leu304 of α-Helix13), (Thr68 of β-Sheet 1), (Asn125 of β-Sheet 3) have enhanced fluctuation in terms of conformational entropy around the protein–protein interaction site (Figure S4. (b)). A pronounced enhancement of fluctuation is observed in the N-terminal DNA binding residues, namely Tyr7, Ser21, Val24, Met42, Ile48 for the Operator O2-bound complex.Asymmetric motional freedom is pronounced in the Lac-O2 complex. In the operator-bound state, augmented conformational entropy is observed in many residues of the N and C-terminal core domain of both the protein chains at or around the protein–protein interaction site. Inspite of being a homo-dimer with an identical primary sequence, asymmetric motional freedom of different subunits is observed in Lac repressor. Significant disorder is found in α-helix6 (chainB of lac repressor (O-Sym bound)), α-helix8 (chainA and chainB of lac repressor (O-Sym and O2 bound)), α-helix9 (chain A of lac repressor (O-Sym and O2 bound)), α-helix13 (chainA of lac repressor (O-Sym bound)), β-sheet1 (chainB of lac repressor (O-Sym and O2 bound)), β-sheet3 (chainB of lac repressor (O-Sym and O2 bound)), loop12 (chainA of lac repressor (O2 bound), loop16 (chain A of lac repressor (O-Sym and O2 bound)) conformation. A significant decrease in flexibility is observed for DNA-contacting side chains upon complexation. Most of the N-terminal residues of both dimers in Lac repressor-SymL complex show a reduction in entropy due to additional interaction with the DNA. Residue-wise conformational entropy changes suggest ordering of the residues which form quite a stable sequence-specific H-bond with DNA. Further study of their domain motions and essential dynamics of the protein subunits can shed light on this aspect. Few residues of the DNA binding region also show enhancement of conformational entropy in the operator bound state. Thus, an important conclusion may be drawn that some residues at or near the protein–protein and protein-DNA interface attain enhanced mobility upon DNA binding. Figure S4. (c) represents both the N-terminus DNA binding domain of chain A & chain B Lac repressor protein (operator O-Sym and O2 bound) are stabilized and ordered. But the change in conformational free energy and entropy of N-terminus DNA binding

domain of Lac repressor protein (for operator O2 bound) chain A (ΔG =-26.72 kJ/mol, TΔS= -45.63 kJ/mol) and chain B (ΔG= -3.51 kJ/mol, TΔS= -20.40 kJ/mol) differ significantly in comparison to the change in conformational free energy and entropy of chain A (ΔG= -37.0 kJ/mol, TΔS= -40.67 kJ/mol) and chain B (ΔG= -19.01 kJ/mol, TΔS= -57.44 kJ/mol) Lac repressor protein bound to operator O-Sym. Figure S4. (d) shows the thermodynamics of conformational changes of the total N-terminal+ C-terminal sub-domains of Chain A lac repressor is (operator O2 bound) destabilized (ΔG= 5.3kJ/mol) and disordered (TΔS= 7.68 kJ/mol) whereas Chain B lac repressor (operator O2 bound) is stabilized (ΔG= -25.77 kJ/mol) and ordered (TΔS= -18.21 kJ/mol). On the other hand the N-terminal+ C-terminal sub-domains of Chain A lac repressor (operator O-Sym bound) is destabilised (ΔG=12.61 kJ/mol) but ordered (TΔS= -11.66 kJ/mol). The N-terminal+ C-terminal sub-domains of Chain B lac repressor (operator O-Sym bound) is marginally stable (ΔG= -2.07 kJ/mol) but significantly disordered (TΔS= 29.48 kJ/mol).

Amino acids of lac repressor protein are classified into polar (Asn, Cys, Gln, Ser, Thr, and Tyr), hydrophobic (Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, Val), acidic (Asp, and Glu), and basic (Arg, Lys, and His) based on chemical nature. The contributions from such four different kinds of amino acid groups towards the change in conformational thermodynamics of the Lac repressor chain A and chain B (operator bound) with respect to chain A and chain B Lac repressor (free) are displayed in the Figure S4. (e), (f), (g) and (h).

Acidic amino acids impart marginal conformational stability of chain B of lac repressor (operator O-Sym bound) and chain A of lac repressor (operator O2). On the other hand chain A of lac repressor (operator O-Sym bound) is energetically stabilised whereas chain A and B of lac repressor (operator O-Sym bound) and chain A of lac repressor (operator O2) are entropically ordered significantly by contribution from acidic amino acids. Basic amino acids impart marginal conformational destabilization and disorder of both chain A and B of lac repressor (operator O-Sym bound) and significant conformational destabilization and disorder of chain A lac repressor (operator O2 bound). However significant conformational stabilization and order towards chain B lac repressor (operator O2 bound) are contributed by acidic and basic amino acids. Hydrophobic amino acids impart significant destabilization and disorder for chain B lac repressor (operator O2 bound) while chain A and B lac repressor (operator O-Sym bound), chain A lac repressor (operator O2 bound) get significantly stabilized and ordered. Mostly polar amino acid residues from both Lac repressor chain A and chain B (operator O-Sym and O2 bound) contribute the most conformational stability and order suggesting electrostatic interaction plays the major role in the change in conformational thermodynamics due to repressor-operator complex formation. The change in conformational free energy of Lac repressor chain A and chain B (operator O-Sym bound) are -40.13 kJ/mol and -16.13 kJ/mol respectively whereas for Lac repressor chain A and chain B (operator O2 bound) system, the change in conformational free energy are -23.67 kJ/mol and -31.32 kJ/mol respectively due to contribution from polar residues. On the other hand, -42.14 kJ/mol and -30.81 kJ/mol are the contribution of the conformational entropy cost for Lac repressor chain A and chain B (operator O-Sym bound) respectively as well as change in conformational entropy for Lac repressor chain A and chain B (operator O2 bound) are -49.01 kJ/mol and -31.23 kJ/mol respectively corresponding to polar residues. It is essential to mention that acidic and basic amino acids play the major role in the conformational stability and order of the Lac repressor chain B (operator O2 bound) [ΔG= -3.33 kJ/mol, TΔS= -13.57 kJ/mol corresponding to acidic residues; ΔG= -9.45 kJ/mol, TΔS= -15.71 kJ/mol corresponding to basic residues] instead of Lac repressor chain A (operator O2 bound) [ΔG= 3.72 kJ/mol, TΔS= 12.41 kJ/mol corresponding to acidic residues; ΔG=7.10 kJ/mol, TΔS= 11.31 kJ/mol corresponding to basic residues] and chain A [ΔG=13.51 kJ/mol, TΔS= 7.27 kJ/mol corresponding to acidic residues; ΔG= 2.0 kJ/mol, TΔS=2.24 kJ/mol corresponding to basic residues] and chain B [ΔG=0.81 kJ/mol,

TΔS=10.04 kJ/mol corresponding to acidic residues; ΔG= 2.92 kJ/mol, TΔS= 0.77 kJ/mol corresponding to basic residues] (operator O-Sym bound). On the other hand, the hydrophobic amino acids aid the conformational stability and order of the Lac repressor chain A [ΔG= 0.45 kJ/mol, TΔS= -18.36 kJ/mol] and chain B of (operator O-Sym bound) [ΔG= -8.68 kJ/mol, TΔS= -9.06 kJ/mol] and chain A (operator O2 bound) [ΔG= -8.59 kJ/mol, TΔS= -16.26 kJ/mol] instead of the Lac repressor chain B (operator O2 bound) [ΔG= 14.80 kJ/mol, TΔS= 14.82 kJ/mol]. Therefore in addition to electrostatic interaction, van der Waals interaction play significant role towards thermodynamics of conformational changes of the lac repressor protein in operator O-Sym and O2 bound state with respect to lac repressor (free).

3.5 Domain Motions and Conformational Transitions Revealed by PCA

Correlated motions govern energy transport and allosteric signaling in biomolecules. Because many functional processes are driven by slow, large-scale conformational rearrangements rather than local thermal vibrations, the conformational transitions of the Lac repressor and its DNA-bound complexes were examined by projecting the molecular dynamics trajectories onto a two-dimensional space defined by the first two principal components (PC1 and PC2), where PC1 captures the maximum variance in the dataset.

Porcupine plots were used to visualize sub-domain motions in the Lac-O-SymL and Lac-O2 protein-DNA complexes, as well as in the operator-unbound free protein (Figure 4). For the free Lac repressor, the first principal component confirms RMSF observations, showing the most prominent motions in the helix-turn-helix motif of the DNA-binding domains in both monomers. In the Lac-O-SymL complex, the regulatory domain of monomer B exhibits greater overall motion than monomer A, with the largest displacements occurring in loops connecting α-helices and the DNA-binding domain. Motions in both Lac-O-SymL and Lac-O2 complexes are primarily localized to loop and turn regions, while subunit A binds the left operator site and subunit B binds the right site.

Analysis of domain motions and essential dynamics reveals asymmetric dynamics between the two monomers. Notably, binding to the O2 operator enhances fluctuations in both the N- and C-terminal core domains, and the protein subunits exhibit a tendency to move away from each other. In contrast, in the operator-unbound protein, major dynamics are largely restricted to the DNA-binding domain, with minimal motion observed in other regions.

Essential dynamics extracted from MD trajectories highlight large-scale correlated motions in the protein-DNA complexes, eliminating recurrent modes to uncover dominant structural transitions. The intrinsic asymmetry of the two half-sites in the natural O2 operator, together with the absence of the central GC base pair in the symmetric synthetic SymL operator, suggests significant differences in the principal modes of motion of the dimeric Lac repressor upon binding to these sequences.

In summary, porcupine plots demonstrate that DNA sequence allosterically modulates protein dynamics even in domains distal to the DNA-binding site. How such allosteric effects relate to gene regulation and how they are transmitted throughout the protein remains an open question, underscoring the complex interplay between DNA sequence and protein dynamics.

The analysis of domain motions and essential dynamics indicates that the two monomers of the Lac repressor display asymmetric dynamics, which are strongly influenced by the DNA sequence bound. When bound to the natural O2 operator, fluctuations increase not only in the DNA-binding domain but also propagate to the N- and C-terminal regions of the core domain, and the subunits tend to move slightly apart. This suggests that DNA binding induces long-range dynamic effects, modulating regions far from the interface. In contrast, the operator-free protein exhibits motion largely confined to the DNA-

binding domain, with minimal fluctuations in the rest of the structure, indicating that DNA is necessary to trigger extensive allosteric dynamics. By focusing on the dominant collective modes, it becomes evident that the asymmetry of the natural O2 operator particularly the intrinsic differences between its two half-sites induces distinct principal motions in each monomer. In comparison, binding to a symmetric synthetic operator such as SymL, which lacks the central GC base pair, results in a more uniform dynamic response. This demonstrates that subtle differences in DNA sequence can have profound effects on the pattern and magnitude of protein motions, emphasizing the role of sequence in shaping allosteric communication pathways.

Allostery refers to the phenomenon where binding of a ligand at one site of a protein induces conformational or dynamic changes at a distal site, thereby modulating protein function. In our study, the porcupine plot and essential dynamics analysis show that binding of DNA operators (O-SymL or O2) to the canonical DNA-binding site of the Lac repressor induces asymmetric motions and enhanced fluctuations in regions far from the binding interface, such as the regulatory domain, N-terminal, and C-terminal core domains. These changes are sequence-dependent; different operators elicit distinct patterns of subunit-specific. This provides rigorous evidence of allostery, as sequence-dependent perturbations at the DNA-binding interface are transmitted through the LacI dimer to modulate the dynamics of distal regions, resulting in non-equivalent dynamics between the two symmetric monomers.

This sequence-dependent modulation of global dynamics establishes a mechanistic insight for stabilization of different gene regulatory networks differently, in which operators function as active allosteric effectors that modulate inter-domain coupling and conformational thermodynamics, ultimately affecting transcription repression strength and conformational plasticity of the LacI regulatory complex.

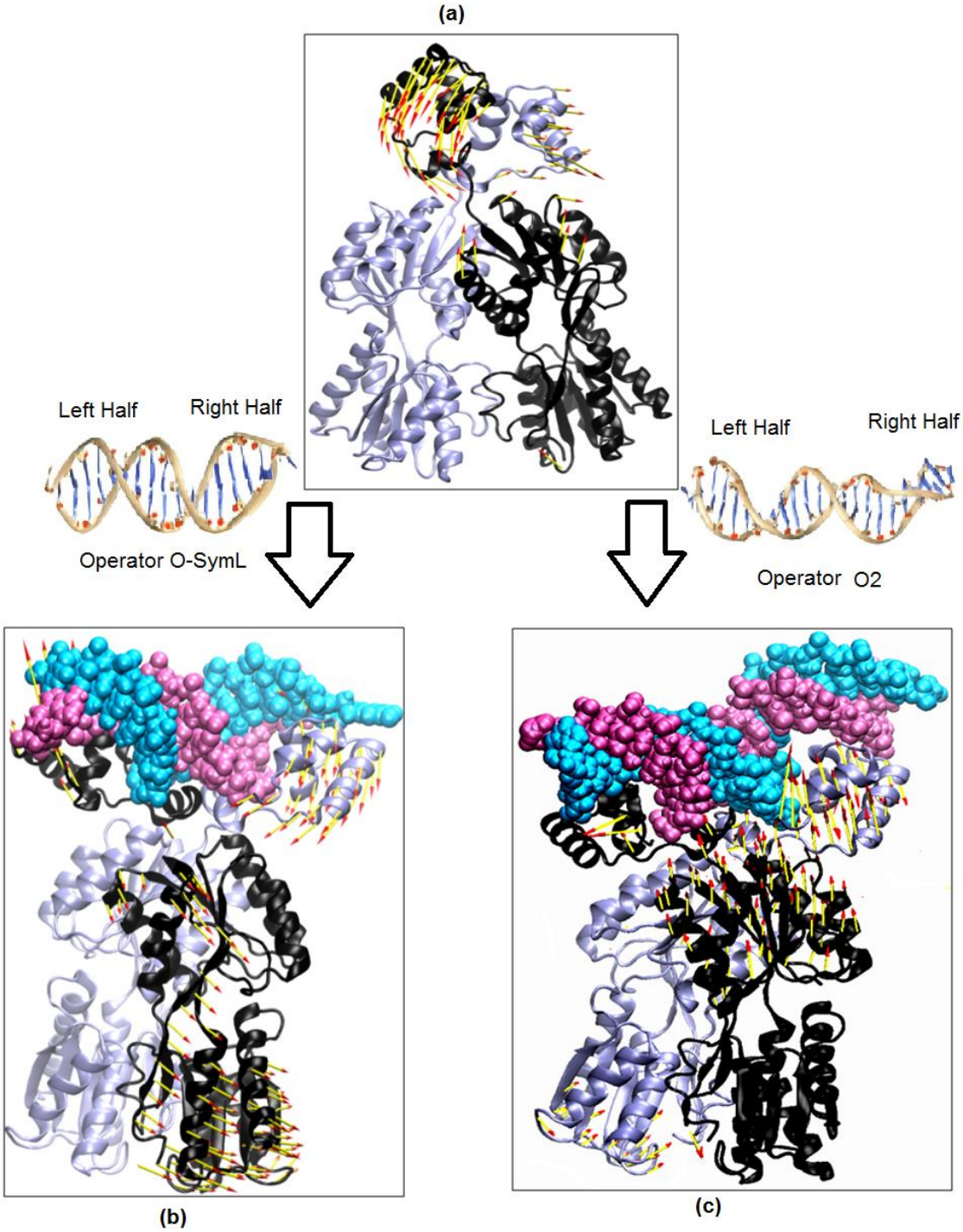

Figure 4: Porcupine plots representing the first principal component (PC1) of motion for (a) free Lac repressor protein, (b) operator SymL bound Lac repressor protein, and (c) operator O2 bound Lac repressor protein. DNA chains are shown as Chain C (cyan) and Chain D (mauve), while protein chains are shown as Chain A (ice blue) and chain B (black). Arrows indicate the direction and magnitude of collective motions. MD simulations were performed using the Amber ff99SB-ILDN force field for the protein and the Amber ParmBSC1 force field for DNA. This figure is reproduced from the author's Ph.D. thesis [Das, S. (2023). "Computational Study Of Structure And Dynamics of Protein-DNA Complexes" (Doctoral thesis). University of Calcutta. http://hdl.handle.net/10603/508120] and a preprint [Das, S., Roy, S. and Bhattacharyya, D. (2024), "In-silico studies on structural and thermodynamics basis of interaction of lac repressor protein binding to different DNA operators." https://doi.org/10.1101/2024.07.07.602393] available on bioRxiv.

The porcupine plot derived from principal component analysis of the MD trajectory of the operator unbound free Lac repressor protein reveals pronounced collective motions primarily localized within the N-terminal DNA binding domains suggesting exploration of multiple conformational sub-states, consistent with a heterogeneous free-energy landscape. The dominant eigenvector (PC1) highlights large-amplitude collective atomic fluctuations of the helix-turn-helix motifs, suggesting substantial conformational flexibility in the absence of operator DNA. In contrast, the core regulatory domains exhibit comparatively smaller eigenvector components, indicating a relatively stable protein scaffold. DNA-binding domains and core domains show oriented displacement patterns. Such anti-correlated motion is hallmarks of hinge-like or twisting dynamics. Such intrinsic flexibility is mechanically consistent with facilitated target search, where conformational pre-equilibration reduces the energetic penalty associated with conformational rearrangement upon operator binding.

4. Discussion

The present study demonstrates that conformational thermodynamic landscape of the symmetric Lac repressor is markedly reshaped asymmetrically in an operator-dependent manner, underscoring the strong coupling between local side-chain fluctuations and long range allosteric communication.

The symmetric O-Sym operator induces pronounced and relatively balanced conformational stabilization and reduced flexibility of the Lac repressor at the recognition interface, consistent with the formation of stable, sequence-specific hydrogen bonding networks. In contrast, the natural operator O2 elicits a more heterogeneous response, characterized by localized stabilization accompanied by increased conformational entropy within the N- and C-terminal core sub-domains as well as in regions proximal to the protein-protein interaction interface. Polar residues primarily contribute to the conformational stabilization through favourable electrostatic and hydrogen-bonding interactions, whereas acidic and basic residues introduce pronounced monomer-specific asymmetry in the Lac-O2 repressor-operator complex. Hydrophobic residues provide additional van der Waals stabilization in the Lac-O-Sym repressor-operator complex and selectively in one monomer of the O2 bound Lac repressor dimer, suggesting operator-dependent modulation of side-chain burial and the underlying enthalpy-entropy balance. .

Principal component analyses corroborate these thermodynamic findings by operator-specific patterns of collective motions. In the Lac-O2 complex, dominant eigen-modes display amplified fluctuations that propagate into the distal N- and C- terminal core domains and accentuate asymmetric motions between the two subunits of Lac repressor protein.

The dominant collective motion of the Lac repressor–O-SymL operator complex is markedly anisotropic and asymmetric. One monomer displays significantly larger displacement vectors, particularly within the core domain and DNA-binding domains, whereas the partner monomer exhibits comparatively restricted mobility. This disparity demonstrates that the two subunits contribute unequally to the principal mode of motion, despite the inherent structural symmetry of the dimer. On the other hand the free Lac repressor shows fluctuations predominantly confined to the DNA-binding domain, especially to the helix-turn-helix motif, with minimal propagation into distal domains.

Each Lac operator has a slightly different base sequence and binding affinity. These sequence differences alter protein-DNA interaction pattern (hydrogen bond, electrostatics, van der Waals interactions), which in turn modulate local side-chain flexibility, inter-residue coupling, collective domain motions, and energetic stabilization pattern.

Previous studies have shown that structural perturbations in DNA propagate long range signal and influence distal regulatory sites [94-97].

Our study extends this understanding by demonstrating that DNA sequences function as active allosteric effectors of the Lac repressor protein through selective reweighting of its conformational entropy, rather than acting as merely passive binding substrates.

Asymmetry within homo-dimeric LacI has been reported previously in structural studies, mutational analyses, and experimental thermodynamic investigations [33-34].

Rastinejad et al. (1993) showed that the intrinsic sequence asymmetry of the natural Lac operator dictates an asymmetric pattern of contacts with the Lac repressor's DNA-binding domains. The protein interacts symmetrically only when the intrinsic asymmetry at both half sitesis eliminated. These findings emphasize that operator sequence deviations from a perfect palindrome are functionally important for specifying LacI-DNA binding geometry in long range contexts. Frank et al. (1997) demonstrated that operator sequence variation from consensus symmetric Lac operator to variant or non-specific sequences alter binding affinity by several orders of magnitude. This reduction is accompanied by non-additive free energy penalties, altered enthalpy-entropy compensation, diminished heat-capacity changes, suggesting DNA sequences actively modulates LacI conformational dynamics and binding thermodynamics. Flyn et al. (2003) suggested that the LacI allosteric transition from high-affinity operator-bound state to low-affinity inducer-bound states is not a symmetric concerted motion but rather involves asymmetric multi-pathway conformational changes propagating from the inducer site to DNA-binding domains through the core dimer interface [55].

The concept that DNA sequence modulates protein dynamics and cooperative behaviour is well established in other repressor systems [98-104]. McGinnis et al. (2022) indicated that asymmetric DNA sequences can be recognized by mycobacteriophage immunity repressor through cooperative binding of two distinct DNA-binding domains, each providing essential and functionally distinct contacts. This study shows that regulatory specificity does not necessarily depend on symmetric dimerization, but can instead emerge from coordinated multi-domain recognition of non-palindromic operator sequences. Mazumder et al. (2012) showed that the λ-CI repressor binds cooperatively to adjacent operator sites in an orientation-specific manner. In a later study, Mazumder et al. (2017) further demonstrated that DNA sequence can allosterically strengthen distant protein-protein interactions by altering conformational dynamics. Similar sequence-dependent dynamic effects have been observed for both the λ Cro and Gal repressors. Using ultrafast fluorescence spectroscopy, Choudhury et al. (2016) investigated that λ Cro repressor, a transcriptional regulator that controls the lytic-lysogenic switch by binding specific operator DNA sequences. Using ultrafast fluorescence spectroscopy, they showed that the minor-groove binding domains of two operator DNAs exhibit pronounced sequence-dependent flexibility in picoseconds-scale while the overall protein backbone dynamics remain largely unchanged. In the λ Cro repressor system, increased fast DNA motions, especially in the OR3 complexes were associated with differences in the dynamic and entropic contributions to binding specificity. Likewise, in the Gal repressor, distinct operator sequences differentially modulated protein motions, highlighting the importance of dynamic allostery in sequence recognition. Naiya et al. (2016) further proposed conformational selection as a key mechanism for multi-sequence recognition, where specificity arises from preferential stabilization of pre-existing conformational states.

Recent computational and experimental studies further reinforce the emerging paradigm that DNA sequence intrinsically encodes regulatory information through modulation of protein dynamics and thermodynamics. Molecular dynamics simulation of various transcription factors (e.g.; SOC1, ZTA, TBR1, EGR/Sp1, and AP-1 consistently demonstrate that subtle variations in DNA sequence reshape hydrogen bonding networks, electrostatic interactions, binding affinity, and enthalpy-entropy balance

[105-109]. Beyond individual transcription factors, Onyema et al. (2026) revealed that nucleosomes have long range allosteric networks connecting DNA and histone proteins. Specific DNA and histone residues act as allosteic hubs, allowing propagation of local perturbations to influence chromatin accessibility. Quantitative characterization of this allosteric behaviour enables the identification of residues and domains that may serve as potential targets for therapeutic intervention [110].

Taken together these studies strongly support our findings that DNA sequences are not merely passive binding substrates but actively modulate protein conformational ensembles, binding thermodynamics, and long-range allosteric communication. Our results extend the ensemble-based model by providing residue level thermodynamic quantification of sequence-dependent entropy redistribution in LacI, thereby establishing a direct connection between operator DNA sequence and asymmetric modulation of protein's conformational free-energy landscape.

4.1 Significance and Broader Implications

The torsional angle based thermodynamics computation using histogram based method is broadly applicable to protein-DNA, protein-RNA, protein-protein, and protein-ligand systems in which functional regulation arises from redistribution of conformational entropy. Estimation of conformational thermodynamics using the histogram based method provides a precise strategy for identifying allosteric druggable sites. Residues or regions showing pronounced ligand or DNA/RNA induced entropy redistribution define dynamic regulatory node within the allosteic network. Such sites are attractive for drug discovery because modulations of their conformational flexibility can shift the global free-energy landscape. In this perspective, allosteric ligands function as thermodynamic effectors that reshape conformational entropy distributions rather than merely occupying cryptic pockets. Residue level entropy mapping thus integrates dynamic free-energy considerations into selective drug design.

Sequence-dependent redistribution of conformational ensembles emerges as a general mechanistic principle underlying transcription factor specificity. DNA-mediated ensemble reweighting has been reported for p53, Catabolite activator protein, NF-κβ, and the Glucocorticoid receptor, where distinct response elements stabilize alternative conformational substates despite comparable binding affinities. These findings indicate that regulatory discrimination is not determined solely by equilibrium affinity but by DNA-encoded modulation of the protein's conformational free-energy landscape [111-115].

Our results extend this paradigm by demonstrating that regulatory specificity can arise from entropy redistribution and dynamic asymmetry at residue level, rather than from static structural differences alone.

Beyond transcriptional regulation, residue-specific free energy and conformational entropy analysis provides a transferable thermodynamic framework for genetic engineering and synthetic biology. In systems such as CRISPR-Cas9, target discrimination depends not only on binding strength but also on protein flexibility and internal dynamics that influence mismatch tolerance and off-target activity. Quantifying torsional entropy changes between cognate and mismatched complexes can identify residues that govern target discrimination specificity, thereby guiding rational engineering of higher specificity variants. A similar principle can be applied to engineered transcription factors and epigenetic reader proteins, where subtle shifts in conformational entropy modulate recognition of DNA and histone modifications and alter regulatory outcomes without large changes in binding free energy [116].

RNA-guided nucleases further represent entropy-driven allostery. In CRISPR-Cas13a, RNA binding propagates long-range allosteric communication from the REC lobe to the HEPN nuclease domains that stabilizes catalytically competent conformations through reorganizing interdomain interactions and stabilizing catalytically competent conformations [117]. Because RNA-mediated allosteric activation in CRISPR-Cas13a reflects ensemble reorganization rather than discrete

conformational transitions, histogram based method enables residue-wise quantification of conformational entropy, recasting experimentally identified allosteric hotspots as sites of significant torsional entropy change. This approach provides a dynamic and thermodynamic map for engineering RNA-targeting tools [117].

From a translational perspective, DNA-induced reshaping of thermodynamic parameters has important implications for drug discovery. Allosteric pockets may be transiently formed or preferentially stabilized within DNA-bound conformational ensembles, thereby enable selective targeting of transcription factors that are otherwise challenging to modulate through orthosteric inhibition. Instead of blocking the large protein-DNA interface directly, small molecules can work by shifting the balance between different conformations that the protein already samples. By favouring one conformational state over others, such ligands acting as ensemble modulators, effectively reshape the protein's free-energy landscape and alter its functional behaviour. This principle is consistent with topology-guided redesign of allosteric transcription factors within the LacI family underscoring the mechanistic relevance of conformational thermodynamics as a quantitative framework for rational protein design [118]. In drug discovery, engineered protein interfaces are designed using different strategies such as AI guided mutational scanning, interface residue scanning, and machine learning based affinity ranking to develop high affinity inhibitors and biologics that selectively modulate disease relevant targets such as protein-protein and protein-DNA interactions [119].

4.2 Limitations and Future Directions

The present results indicate that DNA sequence alone drives asymmetric redistribution of conformational free energy and entropy across structurally identical subunits, manifested at the conformational ensemble level through coordinated side-chain and collective motions rather than discrete allosteric hubs. This distributed asymmetry supports an entropy driven model of allostery.Although targeted mutational or in silico perturbation analyses could provide complementary validation, the conclusions do not depend on such interventions, as all systems, comprising the identical protein sequence, were simulated using identical force-field parameters and simulation protocols.

MM/PBSA based energy decomposition, dynamic network analysis, computation of electrostatic potential, and detailed hydration-shell or water dynamics analyses, residence time calculations were not performed as they involve additional model dependent assumptions and parameter selections such as dielectric constants, solvent accessible surface definitions, and shell boundaries and may introduce systematic bias. Similarly, correlation network analysis depends on user-defined thresholds that may potentially influence the resulting connectivity patterns. In contrast, principal component analysis (PCA), based on direct diagonalization of the covariance matrix of atomic fluctuation, provides a parameter independent description of dominant collective motions and enables an un-biased evaluation of large-scale dynamic asymmetry.

Conformational thermodynamics quantities were computed exclusively from the side-chain torsion χ_1 using the histogram based method,capturing the principal rotameric transitions underlying side-chain conformational plasticity at both interfacial and distal residues. Therefore conformational thermodynamics results using HBM quantify relative internal conformational changes between ligand-bound and ligand-inbound states, not absolute binding free energies of the full LacI-DNA repressor-operator complexes. Higher order side-chain torsion angle (χ_2, χ_3, χ_4) and backbone (φ/ψ) dihedrals were not included. Thus the reported values represent χ_1 specific entropy contributions rather than full multidimensional conformational entropy estimate. This reduction of microscopic conformational variables may underestimate absolute entropy and neglect certain torsional cross-correlations. Nevertheless, applying the χ_1-specific approaches consistently across all systems, and its agreement with essential dynamics, makes it reliable for comparing sequence-dependent allosteric effects.

Future microsecond-scale molecular dynamics simulations of LacI-O1 and LacI-O3 complexes incorporating all microscopic conformational variables would enable complete thermodynamic characterization of the protein-DNA recognition landscape and elucidate sequence-dependent differences in global dynamics. In addition, targeted mutational analyses could further decipher the role of protein dynamics in allostery-driven protein-protein communication. Systematic base substitution of DNA could quantify how base-specific energetic contributions modulate redistributions of free energy and entropy. Mutations affecting protein-DNA interactions (PDIs) are major drivers of disease. Mutations in DNA binding proteins or their target sequences can disrupt interaction energetic, leading to aberrant gene expression, replication stress, and DNA damage. These changes often cause transcriptional misregulation, aberrant signalling, and genomic instability. Disease associated mutations frequently alter the thermodynamics or kinetics of protein-DNA interactions, producing phenotypic consequences without complete loss of molecular function [2].

5. Conclusion

In summary, this work demonstrates that multi-sequence DNA recognition by LacI is governed by asymmetric redistribution of conformational entropy and free energy across structurally identical subunits. These dynamic disparities originate from variations in hydrogen-bonding patterns, deformation of DNA base-pair step, and minor groove geometry of DNA, underscoring the role of DNA shape and intrinsic sequence asymmetry in modulating protein motion. The absence of a central GC base pair in O-SymL and the inherent half-site asymmetry of operator O2 promote distinct recognition modes. Sequence-dependent differences in overall protein dynamics upon DNA binding play a crucial role in modulating recognition and communication between the C-terminal domains during tetramerization of Lac repressor, thereby contributing to its functional regulation. Operator DNA acts as an active allosteric ligand that reshapes the Lac repressor protein's conformational landscape, thereby differentially stabilizing gene regulatory networks in a sequence-dependent manner. Distinct domain motions of symmetric subunits of Lac repressor characterized by essential dynamics and residue-specific conformational thermodynamics advances a quantitative framework for understanding entropy-driven allostery in transcription factors and establishes a generalized paradigm for sequence-encoded dynamic regulation.By defining the dynamic and thermodynamic basis of sequence-dependent allostery, this study provides a rational basis for engineering Lac repressor variants with broad implications for gene regulation and synthetic biology and guiding drug discovery through selective modulation of protein-DNA interactions and allosteric pathways.

List of Abbreviation

CVs	Collective Variables
GROMACS	GROningen Machine for Chemical Simulations
HBM	Histogram Based Method
HTH	Helix Turn Helix
IPTG	Isopropyl β-D-1thiogalactopyranoside
LacI	Lactose Repressor Protein
LINCS	Linear Constraint Solver
MD	Molecular Dynamics
MM/PBSA	Molecular Mechanics Poisson-Boltzmann surface area
MM/GBSA	Molecular Mechanics Generalized Born surface area
NMR	Nuclear Magnetic Resonance
OptoLacI	Optogenetic Engineering of LacI
PBC	Periodic Boundary Condition

PCA	Principal Component Analysis
PDB	Protein Data Bank
PME	Particle Mesh Ewald
PYMOL	Python Molecular Graphics System
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
TFs	Transcription factors
TS	Transition State
VMD	Visual Molecular Dynamics

Author Contribution:
The author (Soumi Das) solely conceived the Conceptualization, methodology, software; Validation, formal analysis, funding acquisition; Investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, project administration. The author has reviewed and approved the published version of the article.

Data Availability: Data supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The author declares no competing interests.

Funding: This work was supported by the Council of Scientific and Industrial Research (CSIR), Government of India, under the CSIR–NET fellowship scheme in Chemical Sciences (awarded to Soumi Das). No formal grant number was issued.

AI Declaration: Artificial Intelligence tool (ChatGPT) was used solely for language refinement. All scientific concepts, data analyses, data interpretation, and intellectual contributions are independently conceived, critically evaluated, and validated by the author. The author conducted all substantive revisions, approved the final version of the manuscript, and assumes full responsibility for its content and scientific integrity.

Acknowledgements: The author sincerely acknowledges Prof. Siddhartha Roy (Retd.) and Prof. Dhananjay Bhattacharyya (Retd.) for insightful discussions. Computational facilities were provided by the Bose Institute, the Saha Institute of Nuclear Physics, and the Technical Research Centre at the S. N. Bose National Centre for Basic Sciences, Kolkata, India. This work is author's original contribution and is partly derived from the author's Ph.D. thesis [Das, S. (2023). "*Computational Study Of Structure And Dynamics of Protein-DNA Complexes"*, University of Calcutta, http://hdl.handle.net/10603/508120] and a preprint [Das, S., Roy, S. and Bhattacharyya, D. (2024), "In-silico studies on structural and thermodynamics basis of interaction of lac repressor protein binding to different DNA operators." https://doi.org/10.1101/2024.07.07.602393] available on bioRxiv.

References

1. Mishra R, Ahmad S. Protein-DNA interactions in disease and drug discovery. Chem Commun (Camb). 2026 Feb 12;62(12):3702-3729. doi: 10.1039/d5cc06239j. PMID: 41631674.
2. Genz LR, Nair S, Sweeney A, Topf M. Drug targeting of protein-nucleic acid interactions. Curr Opin Struct Biol. 2025 Dec;95:103165. doi: 10.1016/j.sbi.2025.103165. Epub 2025 Oct 7. PMID: 41061560.
3. Das, S. (2025). Decoding the Mechanistic Landscape of Protein-DeoxyriboNucleic Acid Recognition: A Comprehensive Review. ES Chemistry and Sustainability. https://doi.org/10.30919/cs1610
4. Borges Farias A, Sganzerla Martinez G, Galán-Vásquez E, Nicolás MF, Pérez-Rueda E. Predicting bacterial transcription factor binding sites through machine learning and structural characterization based on DNA duplex stability. Brief Bioinform. 2024 Sep 23;25(6):bbae581. doi: 10.1093/bib/bbae581. PMID: 39541188; PMCID: PMC11562833.
5. Ovek Baydar D, Rauluseviciute I, Aronsen DR, Blanc-Mathieu R, Bonthuis I, de Beukelaer H, Ferenc K, Jegou A, Kumar V, Lemma RB, Lucas J, Pochon M, Yun CM, Ramalingam V, Deshpande SS, Patel A, Marinov GK, Wang AT, Aguirre A, Castro-Mondragon JA, Baranasic D, Chèneby J, Gundersen S, Johansen M, Khan A, Kuijjer ML, Hovig E, Lenhard B, Sandelin A, Vandepoele K, Wasserman WW, Parcy F, Kundaje A, Mathelier A. JASPAR 2026: expansion of transcription factor binding profiles and integration of deep learning models. Nucleic Acids Res. 2026 Jan 6;54(D1):D184-D193. doi: 10.1093/nar/gkaf1209. PMID: 41325984; PMCID: PMC12807658.
6. Masri CA, Yu J. Combining Physics-Based Protein-DNA Energetics with Machine Learning to Predict Interpretable Transcription Factor-DNA Binding. J Chem Inf Model. 2025 Nov 10;65(21):11804-11817. doi: 10.1021/acs.jcim.5c01143. Epub 2025 Oct 23. PMID: 41131660.

7. Wang Y, Li J, Chiu TP, Xin B, Rohs R. Sequence-based modeling of low-affinity transcription factor-DNA binding through deep learning. NAR Genom Bioinform. 2026 Mar 5;8(1):lqag027. doi: 10.1093/nargab/lqag027. PMID: 41799016; PMCID: PMC12961433.
8. Zhang, J., Liu, B., Wu, J. *et al*. DeepCAC: a deep learning approach on DNA transcription factors classification based on multi-head self-attention and concatenate convolutional neural network. *BMC Bioinformatics* 24, 345 (2023). https://doi.org/10.1186/s12859-023-05469-9
9. Tariq, S., Amin, A. DeepCTF: transcription factor binding specificity prediction using DNA sequence plus shape in an attention-based deep learning model. *SIViP* 18, 5239–5251 (2024). https://doi.org/10.1007/s11760-024-03229-7
10. Sandro Barissi, Alba Sala, Miłosz Wieczór, Federica Battistini, Modesto Orozco, DNAffinity: a machine-learning approach to predict DNA binding affinities of transcription factors, Nucleic Acids Research, Volume 50, Issue 16, 9 September 2022, Pages 9105–9114, https://doi.org/10.1093/nar/gkac708
11. Jacklin Sedhom, Jason Kinser, Lee A Solomon, Alignment of major-groove hydrogen bond arrays uncovers shared information between different DNA sequences that bind the same protein, *NAR Genomics and Bioinformatics*, Volume 4, Issue 4, December 2022, lqac101, https://doi.org/10.1093/nargab/lqac101
12. Sakong S, Fierz B, Suter DM. Electrostatic properties of disordered regions control transcription factor search and pioneer activity. Nat Commun. 2026 Feb 8;17(1):2512. doi: 10.1038/s41467-026-69284-5. PMID: 41656376; PMCID: PMC12996321.
13. Jonas F, Navon Y, Barkai N. Intrinsically disordered regions as facilitators of the transcription factor target search. Nat Rev Genet. 2025 Jun;26(6):424-435. doi: 10.1038/s41576-025-00816-3. Epub 2025 Feb 21. PMID: 39984675.
14. Marklund E, Mao G, Yuan J, Zikrin S, Abdurakhmanov E, Deindl S, Elf J. Sequence specificity in DNA binding is mainly governed by association. Science. 2022 Jan 28;375(6579):442-445. doi: 10.1126/science.abg7427. Epub 2022 Jan 27. PMID: 35084952.
15. Wilson CJ, Zhan H, Swint-Kruse L, Matthews KS. The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. Cell Mol Life Sci. 2007, 64(1):3-16. doi: 10.1007/s00018-006-6296-z.
16. Swint-Kruse L, Matthews KS. Allostery in the LacI/GalR family: variations on a theme. CurrOpinMicrobiol. 2009, 12(2):129-37. doi: 10.1016/j.mib.2009.01.009.
17. Kalodimos CG, Bonvin AM, Salinas RK, Wechselberger R, Boelens R, Kaptein R. Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator 01 through alternative conformations of its DNA-binding domain. EMBO J. 2002, 21(12):2866-76. doi: 10.1093/emboj/cdf318.
18. Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, Kaptein R. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science. 2004, 305(5682):386-9. doi: 10.1126/science.1097064. PMID: 15256668.
19. Kalodimos CG, Boelens R, Kaptein R. Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the Lac repressor system. Chem Rev. 2004, 104(8):3567-86. doi: 10.1021/cr0304065. PMID: 15303828.
20. Romanuka J, Folkers GE, Biris N, Tishchenko E, Wienk H, Bonvin AM, Kaptein R, Boelens R. Specificity and affinity of Lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein-DNA complexes. J Mol Biol. 2009, 390(3):478-89. doi: 10.1016/j.jmb.2009.05.022. Epub 2009 May 18. PMID: 19450607.
21. Bell CE, Lewis M. A closer view of the conformation of the Lac repressor bound to operator. Nat Struct Biol. 2000, 7(3):209-14. doi: 10.1038/73317. PMID: 10700279.
22. Daber R, Stayrook S, Rosenberg A, Lewis M. Structural analysis of lac repressor bound to allosteric effectors. J Mol Biol. 2007, 370(4):609-19. doi: 10.1016/j.jmb.2007.04.028. Epub 2007 Apr 19. PMID: 17543986; PMCID: PMC2715899.
23. Lewis M. The lac repressor. C R Biol. 2005 Jun;328(6):521-48. doi: 10.1016/j.crvi.2005.04.004. PMID: 15950160.
24. Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science. 1996, 271(5253):1247-54. doi: 10.1126/science.271.5253.1247. PMID: 8638105.
25. Bell CE, Lewis M. Crystallographic analysis of Lac repressor bound to natural operator O1. J Mol Biol. 2001, 312(5):921-6. doi: 10.1006/jmbi.2001.5024.
26. Sadler JR, Sasmor H, Betz JL. A perfectly symmetric lac operator binds the lac repressor very tightly. ProcNatlAcadSci U S A. 1983, 80(22):6785-9. doi: 10.1073/pnas.80.22.6785.
27. Spronk CA, Bonvin AM, Radha PK, Melacini G, Boelens R, Kaptein R. The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. Structure. 1999, 7(12):1483-92. doi: 10.1016/s0969-2126(00)88339-2.
28. Kalodimos CG, Folkers GE, Boelens R, Kaptein R. Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. ProcNatlAcadSci U S A. 2001, 98(11):6039-44. doi: 10.1073/pnas.101129898.
29. Markiewicz P, Kleina LG, Cruz C, Ehret S, Miller JH. Genetic studies of the lac repressor. XIV. Analysis of 4000 altered Escherichia coli lac repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence. J Mol Biol. 1994, 240(5):421-33. doi: 10.1006/jmbi.1994.1458. PMID: 8046748.
30. Nelson BD, Manoil C, Traxler B. Insertion mutagenesis of the lac repressor and its implications for structure-function analysis. J Bacteriol. 1997, 179(11):3721-8. doi: 10.1128/jb.179.11.3721-3728.1997. PMID: 9171422; PMCID: PMC179170.
31. Milk, L.; Daber, R.; Lewis, M. Functional rules for lac repressor–operator associations and implications for protein–DNA interactions. Protein Science 2010, 19 (6), 1162–1172. https://doi.org/10.1002/pro.389.
32. Betz JL, Sasmor HM, Buck F, Insley MY, Caruthers MH. Base substitution mutants of the lac operator: in vivo and in vitro affinities for lac repressor. Gene. 1986; 50(1-3):123-32. doi: 10.1016/0378-1119(86)90317-3.
33. Rastinejad F, Artz P, Lu P. Origin of the asymmetrical contact between lac repressor and lac operator DNA. J Mol Biol. 1993, 233(3):389-99. doi: 10.1006/jmbi.1993.1519. PMID: 8411152.
34. Frank, D. E., Saecker, R. M., Bond, J. P., Capp, M. W., Tsodikov, O. V., Melcher, S. E., Levandoski, M. M., & Record, M. (1997). Thermodynamics of the interactions of Lac repressor with variants of the symmetric Lac operator: effects of converting a consensus site to a non-specific site. Journal of Molecular Biology, 267(5), 1186–1206. https://doi.org/10.1006/jmbi.1997.0920
35. Mossing MC, Record MT Jr. Thermodynamic origins of specificity in the lac repressor-operator interaction. Adaptability in the recognition of mutant operator sites. J Mol Biol. 1985, 186(2):295-305. doi: 10.1016/0022-2836(85)90106-8.
36. Whitson, P. A., Olson, J. S., & Matthews, K. S. (1986). Thermodynamic analysis of the lactose repressor-operator DNA interaction. Biochemistry, 25(13), 3852–3858. https://doi.org/10.1021/bi00361a017
37. Daber R, Sochor MA, Lewis M. Thermodynamic analysis of mutant lac repressors. J Mol Biol. 2011, 409(1):76-87. doi: 10.1016/j.jmb.2011.03.057. Epub 2011 Apr 1. PMID: 21459098; PMCID: PMC4225698.
38. Fried MG, Stickle DF, Smirnakis KV, Adams C, MacDonald D, Lu P. Role of hydration in the binding of lac repressor to DNA. J Biol Chem. 2002, 277(52):50676-82. doi: 10.1074/jbc.M208540200. Epub 2002 Oct 11. PMID: 12379649.
39. Barr D & van der Vaart A. The natural DNA bending angle in the lac repressor headpiece–O1 operator complex is determined by protein–DNA contacts and water release. PhysChemChem Phys. 2012;14(6):2070–2077. doi:10.1039/c2cp23780f. PMID:22234444.
40. Lüking M, Elf J, Levy Y. Conformational Change of Transcription Factors from Search to Specific Binding: A *lac* Repressor Case Study. J PhysChem B. 2022, 126(48):9971-9984. doi: 10.1021/acs.jpcb.2c05006.
41. Liao Q, Lüking M, Krüger DM, Deindl S, Elf J, Kasson PM, Lynn Kamerlin SC. Long Time-Scale Atomistic Simulations of the Structure and Dynamics of Transcription Factor-DNA Recognition. J PhysChem B. 2019, 123(17):3576-3590. doi: 10.1021/acs.jpcb.8b12363. Epub 2019 Apr 18. PMID: 30952192.

42. Furini S, Domene C. DNA recognition process of the lactose repressor protein studied via metadynamics and umbrella sampling simulations. J PhysChem B. 2014, 118(46):13059-65. doi: 10.1021/jp505885j. Epub 2014 Nov 6. PMID: 25341013.

43. Furini S, Barbini P, Domene C. DNA-recognition process described by MD simulations of the lactose repressor protein on a specific and a non-specific DNA sequence. Nucleic Acids Res. 2013, 41(7):3963-72. doi: 10.1093/nar/gkt099.

44. Xu L, Ye W, Jiang C, Yang J, Zhang J, Feng Y, Luo R, Chen HF. Recognition mechanism between Lac repressor and DNA with correlation network analysis. J PhysChem B. 2015, 119(7):2844-56. doi: 10.1021/jp510940w. Epub 2015 Feb 9. PMID: 25633018.

45. Sengupta R, Capp MW, Shkel IA, Record MT Jr. The mechanism and high-free-energy transition state of lac repressor-lac operator interaction. Nucleic Acids Res. 2017, 45(22):12671-12680. doi: 10.1093/nar/gkx862. PMID: 29036376; PMCID: PMC5727403.

46. Sun L, Tabaka M, Hou S, Li L, Burdzy K, Aksimentiev A, Maffeo C, Zhang X, Holyst R. The Hinge Region Strengthens the Nonspecific Interaction between Lac-Repressor and DNA: A Computer Simulation Study. PLoS One. 2016;11(3):e0152002. doi:10.1371/journal.pone.0152002. PMID:27008630; PMCID:PMC4805274.

47. Seckfort D, Montgomery Pettitt B. Price of disorder in the lac repressor hinge helix. Biopolymers. 2019,110(1):e23239. doi: 10.1002/bip.23239. Epub 2018 Nov 28. PMID: 30485404; PMCID: PMC6335174.

48. Kariyawasam NL, Ploetz EA, Swint-Kruse L, Smith PE. Simulated pressure changes in LacI suggest a link between hydration and functional conformational changes. Biophysical Chemistry. 2024, 304:107126. doi:10.1016/j.bpc.2023.107126. PMID:37924711; PMCID:PMC10842697.

49. Romanuka J, Folkers GE, Gnida M, Kovačič L, Wienk H, Kaptein R, Boelens R. Genetic switching by the Lac repressor is based on two-state Monod-Wyman-Changeuxallostery. ProcNatlAcadSci U S A. 2023, 120(49):e2311240120. doi: 10.1073/pnas.2311240120.

50. Garruss AS, Collins KM, Church GM. Deep representation learning improves prediction of LacI-mediated transcriptional repression. ProcNatlAcadSci U S A. 2021, 118(27):e2022838118. doi: 10.1073/pnas.2022838118. PMID: 34187888; PMCID: PMC8271634.

51. Liu M, Li Z, Huang J, Yan J, Zhao G, Zhang Y. OptoLacI: optogenetically engineered lactose operon repressor LacI responsive to light instead of IPTG. Nucleic Acids Res. 2024, 52(13):8003-8016. doi: 10.1093/nar/gkae479. PMID: 38860425; PMCID: PMC11260447.

52. Xu W, Yan Y, Artsimovitch I, Dunlap D, Finzi L. Positive supercoiling favors transcription elongation through lac repressor-mediated DNA loops. Nucleic Acids Res. 2022, 50(5):2826-2835. doi: 10.1093/nar/gkac093. PMID: 35188572; PMCID: PMC8934669.

53. Goodson KA, Wang Z, Haeusler AR, Kahn JD, English DS. LacI-DNA-IPTG loops: equilibria among conformations by single-molecule FRET. J PhysChem B. 2013, 117(16):4713-22. doi: 10.1021/jp308930c. Epub 2013 Feb 13. PMID: 23406418.

54. Glasgow A, Hobbs HT, Perry ZR, Wells ML, Marqusee S, Kortemme T. Ligand-specific changes in conformational flexibility mediate long-range allostery in the lac repressor. Nat Commun. 2023, 14(1):1179. doi: 10.1038/s41467-023-36798-1. Erratum in: Nat Commun. 2023, 14(1):4208. doi: 10.1038/s41467-023-39918-z.

55. Flynn TC, Swint-Kruse L, Kong Y, Booth C, Matthews KS, Ma J. Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. Protein Sci. 2003, 12(11):2523-41. doi: 10.1110/ps.03188303.

56. Yonetani Y, Kono H. Dissociation free-energy profiles of specific and nonspecific DNA-protein complexes. J PhysChem B. 2013, 117(25):7535-45. doi: 10.1021/jp402664w. Epub 2013 Jun 11. PMID: 23713479.

57. Villa E, Balaeff A, Schulten K. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. ProcNatlAcadSci U S A. 2005, 102(19):6783-8. doi: 10.1073/pnas.0409387102.

58. Das A, Chakrabarti J, Ghosh M. Conformational contribution to thermodynamics of binding in protein-peptide complexes through microscopic simulation. Biophys J. 2013, 104(6):1274-84. doi: 10.1016/j.bpj.2012.12.058.

59. Das, A., Chakrabarti, J., & Ghosh, M. (2013). Conformational thermodynamics of metal-ion binding to a protein. Chemical Physics Letters, 581, pp. 91-95. doi:10.1016/j.cplett.2013.07.022

60. Sikdar S, Chakrabarti J, Ghosh M. A microscopic insight from conformational thermodynamics to functional ligand binding in proteins. MolBiosyst. 2014, 10(12):3280-9. doi: 10.1039/c4mb00434e.

61. Samanta S, Mukherjee S. Microscopic insight into thermodynamics of conformational changes of SAP-SLAM complex in signal transduction cascade. J Chem Phys. 2017, 146(16):165103. doi: 10.1063/1.4981259.

62. Samanta S, Mukherjee S. Co-operative intra-protein structural response due to protein-protein complexation revealed through thermodynamic quantification: study of MDM2-p53 binding. J Comput Aided Mol Des. 2017, 31(10):891-903. doi: 10.1007/s10822-017-0057-y.

63. Maganti L, Ghosh M, Chakrabarti J. Molecular dynamics studies on conformational thermodynamics of Orai1-calmodulin complex. J Biomol Struct Dyn. 2018 Oct;36(13):3411-3419. doi: 10.1080/07391102.2017.1388289. Epub 2017 Nov 10. PMID: 28978262.

64. Maganti, L., Dutta, S., Ghosh, M., & Chakrabarti, J. (2019). Allostery in Orai1 binding to calmodulin revealed from conformational thermodynamics. *Journal of Biomolecular Structure and Dynamics*, *37*(2), 493–502. https://doi.org/10.1080/07391102.2018.1430617

65. Moulick AG, Chakrabarti J. Fluctuation-Dominated Ligand Binding in Molten Globule Protein. J Chem Inf Model. 2023 Sep 11;63(17):5583-5591. doi: 10.1021/acs.jcim.3c00642. Epub 2023 Aug 30. PMID: 37646788.

66. Mondal M, Chakrabarti J, Ghosh M. Molecular dynamics simulations on interaction between bacterial proteins: Implication on pathogenic activities. Proteins. 2018, 86(3):370-378. doi: 10.1002/prot.25446.

67. Dutta S, Ghosh M, Chakrabarti J. In-silico studies on conformational stability of flagellin-receptor complexes. J BiomolStructDyn. 2020, 38(8):2240-2252. doi: 10.1080/07391102.2019.1630317. Epub 2019 Jun 23. PMID: 31232224.

68. Mandal SC, Maganti L, Mondal M, Chakrabarti J. Microscopic insight to specificity of metal ion cofactor in DNA cleavage by restriction endonuclease EcoRV. Biopolymers. 2020, 111(10):e23396. doi: 10.1002/bip.23396.

69. Kole K, Gupta AM, Chakrabarti J. Conformational stability and order of Hoogsteen base pair induced by protein binding. Biophys Chem. 2023, 301:107079. doi: 10.1016/j.bpc.2023.107079.

70. Das S. Decoding the effect of temperatures on conformational stability and order of ligand unbound thermosensing adenine riboswitch using molecular dynamics simulation. J BiomolStructDyn. 2025, 43(18):11042-11055. doi: 10.1080/07391102.2025.2484662. Epub 2025 Apr 3. PMID: 40177813.

71. Das S., Unraveling Conformational Thermodynamics of Ligand Binding to Fluoride RiboswitchAptamer: Implications for Therapeutic Design. *Molecular Modeling Connect* 2 (2025): 0012. https://doi.org/10.69709/MolModC.2025.170203.

72. Shen, W., Zhou, T. & Shi, X. Enhanced sampling in molecular dynamics simulations and their latest applications—A review. *Nano Res.* 16, 13474–13497 (2023). https://doi.org/10.1007/s12274-023-6311-9

73. Mehdi S, Smith Z, Herron L, Zou Z, Tiwary P. Enhanced Sampling with Machine Learning. Annu Rev Phys Chem. 2024, 75(1):347-370. doi: 10.1146/annurev-physchem-083122-125941.

74. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, Hou T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem Rev. 2019 Aug 28;119(16):9478-9508. doi: 10.1021/acs.chemrev.9b00055. Epub 2019 Jun 24. PMID: 31244000.

75. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015 May;10(5):449-61. doi: 10.1517/17460441.2015.1032936. Epub 2015 Apr 2. PMID: 25835573; PMCID: PMC4487606.

76. Jiang D, Du H, Zhao H, Deng Y, Wu Z, Wang J, Zeng Y, Zhang H, Wang X, Wang E, Hou T, Hsieh CY. Assessing the performance of MM/PBSA and MM/GBSA methods. 10. Prediction reliability of binding affinities and binding poses for RNA-ligand complexes. Phys Chem Chem Phys. 2024 Mar 27;26(13):10323-10335. doi: 10.1039/d3cp04366e. PMID: 38501198.

77. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000, 28(1):235-42. doi: 10.1093/nar/28.1.235.

78. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004, 25(13):1605-12. doi: 10.1002/jcc.20084.

79. Das, S., Roy, S. and Bhattacharyya, D. (2024), In-silico studies on structural and thermodynamics basis of interaction of lac repressor protein binding to different DNA operators. https://doi.org/10.1101/2024.07.07.602393

80. Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., &Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1–2, 19–25. doi.org/10.1016/j.softx.2015.06.001
81. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010, 78(8):1950-8. doi: 10.1002/prot.22711.
82. Ivani I, Dans PD, Noy A, Pérez A, Faustino I, Hospital A, Walther J, Andrio P, Goñi R, Balaceanu A, Portella G, Battistini F, Gelpí JL, González C, Vendruscolo M, Laughton CA, Harris SA, Case DA, Orozco M. Parmbsc1: a refined force field for DNA simulations. Nat Methods. 2016, 13(1):55-8. doi: 10.1038/nmeth.3658.
83. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput. 2015, 11(8):3696-713. doi: 10.1021/acs.jctc.5b00255. Epub 2015 Jul 23. PMID: 26574453; PMCID: PMC4821407.
84. Wang X, Xiong D, Zhang Y, Zhai J, Gu YC, He X. The evolution of the Amber additive protein force field: History, current status, and future. J Chem Phys. 2025, 162(3):030901. doi: 10.1063/5.0227517. PMID: 39817575.
85. Liebl K, Zacharias M. The development of nucleic acids force fields: From an unchallenged past to a competitive future. Biophys J. 2023 Jul 25;122(14):2841-2851. doi: 10.1016/j.bpj.2022.12.022. Epub 2022 Dec 20. PMID: 36540025; PMCID: PMC10398263.
86. Darden T, Perera L, Li L, Pedersen L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. Structure. 1999, 7(3):R55-60. doi: 10.1016/s0969-2126(99)80033-1.
87. Parrinello, M. and Rahman, A. (1981) Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. Journal of Applied Physics, 52, 7182-7190. doi.org/10.1063/1.328693
88. Mukherjee S, Bansal M, Bhattacharyya D. Conformational specificity of non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. J Comput Aided Mol Des. 2006, 20(10-11):629-45. doi: 10.1007/s10822-006-9083-x.
89. Bansal M, Bhattacharyya D, Ravi B. NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. ComputApplBiosci. 1995, 11(3):281-7. doi: 10.1093/bioinformatics/11.3.281.
90. Abdi, H., & Williams, L. J. Principal component analysis. WIREs Computational Statistics, (2010), 2(4), 433-459. doi.org/10.1002/wics.101
91. David CC, Jacobs DJ. Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol Biol. 2014;1084:193-226. doi: 10.1007/978-1-62703-658-0_11. PMID: 24061923; PMCID: PMC4676806.
92. Moradi S, Nowroozi A, AryaeiNezhad M, Jalali P, Khosravi R, Shahlaei M. A review on description dynamics and conformational changes of proteins using combination of principal component analysis and molecular dynamics simulation. ComputBiol Med. 2024 Dec;183:109245. doi: 10.1016/j.compbiomed.2024.109245. Epub 2024 Oct 9. PMID: 39388840.
93. Fei J, Ha T. Watching DNA breath one molecule at a time. ProcNatlAcadSci U S A. 2013, 110(43):17173-4. doi: 10.1073/pnas.1316493110.
94. Segers, M., Voorspoels, A., Sakaue, T., &Carlon, E. (2023). Mechanisms of DNA-Mediated Allostery. *Physical Review Letters*, *131*(23), 238402. https://doi.org/10.1103/physrevlett.131.238402
95. Kim S, Broströmer E, Xing D, Jin J, Chong S, Ge H, Wang S, Gu C, Yang L, Gao YQ, Su XD, Sun Y, Xie XS. Probing allostery through DNA. Science. 2013 Feb 15;339(6121):816-9. doi: 10.1126/science.1229223. PMID: 23413354; PMCID: PMC3586787.
96. Traverso, J., Manoranjan, V., Bishop, A. et al. Allostery through protein-induced DNA bubbles. Sci Rep 5, 9037 (2015). https://doi.org/10.1038/srep09037
97. Rosenblum G, Elad N, Rozenberg H, Wiggers F, Jungwirth J, Hofmann H. Allostery through DNA drives phenotype switching. Nat Commun. 2021, 12(1):2967. doi: 10.1038/s41467-021-23148-2. PMID: 34016970; PMCID: PMC8170675.
98. Mondol T, Batabyal S, Mazumder A, Roy S, Pal SK. Recognition of different DNA sequences by a DNA-binding protein alters protein dynamics differentially. FEBS Lett. 2012, 586(3):258-62. doi: 10.1016/j.febslet.2011.12.032. Epub 2012 Jan 8. PMID: 22240202.
99. Mazumder A, Bandyopadhyay S, Dhar A, Lewis DE, Deb S, Dey S, Chakrabarti P, Roy S. A genetic network that balances two outcomes utilizes asymmetric recognition of operator sites. Biophys J. 2012, 102(7):1580-9. doi: 10.1016/j.bpj.2012.01.052. Epub 2012 Apr 3. PMID: 22500758; PMCID: PMC3318126.
100. Mazumder A, Batabyal S, Mondal M, Mondol T, Choudhury S, Ghosh R, Chatterjee T, Bhattacharyya D, Pal SK, Roy S. Specific DNA sequences allosterically enhance protein-protein interaction in a transcription factor through modulation of protein dynamics: implications for specificity of gene regulation. PhysChemChem Phys. 2017, 19(22):14781-14792. doi: 10.1039/c7cp01193h. PMID: 28548177.
101. Choudhury S, Naiya G, Singh P, Lemmens P, Roy S, Pal SK. Modulation of Ultrafast Conformational Dynamics in Allosteric Interaction of Gal Repressor Protein with Different Operator DNA Sequences. Chembiochem. 2016, 17(7):605-13. doi: 10.1002/cbic.201500657. Epub 2016 Mar 3. PMID: 26914958.
102. Choudhury S, Ghosh B, Singh P, Ghosh R, Roy S, Pal SK. Ultrafast differential flexibility of Cro-protein binding domains of two operator DNAs with different sequences. PhysChemChem Phys. 2016, 18(27):17983-90. doi: 10.1039/c6cp02522f. Epub 2016 Jun 21. PMID: 27326896.
103. Naiya G, Raha P, Mondal MK, Pal U, Saha R, Chaudhuri S, Batabyal S, Kumar Pal S, Bhattacharyya D, Maiti NC, Roy S. Conformational selection underpins recognition of multiple DNA sequences by proteins and consequent functional actions. PhysChemChem Phys. 2016, 18(31):21618-28. doi: 10.1039/c6cp03278h. Epub 2016 Jul 18. PMID: 27426617.
104. McGinnis, R.J., Brambley, C.A., Stamey, B. *et al.* A monomeric mycobacteriophage immunity repressor utilizes two domains to recognize an asymmetric DNA sequence. *Nat Commun*13, 4105 (2022). https://doi.org/10.1038/s41467-022-31678-6
105. Kaur S, Sisodia R, Pareek M, Sri T, Chattopadhyay S, Yaduvanshi P, Asthana S, Madhurantakam C, Singh A. Molecular dynamics and evolutionary conservation of FUL-SOC1 regulatory interactions in polyploid Brassica juncea. Funct Integr Genomics. 2026 Mar 16;26(1):67. doi: 10.1007/s10142-026-01831-1. PMID: 41838285.
106. Duraisamy B, Pramanik D. Influence of DNA Sequences on the Thermodynamic and Structural Stability of the ZTA Transcription Factor–DNA Complex: An All-Atom Molecular Dynamics Study. J Phys Chem B. 2025 May 8;129(18):4282-4297. doi: 10.1021/acs.jpcb.4c07713. Epub 2025 Apr 23. PMID: 40266646.
107. Hartman R, Blane A, Kubrom F, Mayet R, Penedo JC, Fanucchi S. Molecular Insights Into the Binding Dynamics of Transcription Factor TBR1 to T-box DNA Sequences. J Mol Biol. 2026 Jan 15;438(2):169543. doi: 10.1016/j.jmb.2025.169543. Epub 2025 Nov 12. PMID: 41237949.
108. Patra P, Gao YQ. Structural and Thermodynamic Insights into Competitive Binding of the EGR1 and Sp1 Zinc Finger Class of Transcription Factors with a GC-Rich DNA Motif. J Chem Inf Model. 2025 Nov 10;65(21):11988-12005. doi: 10.1021/acs.jcim.5c01991. Epub 2025 Oct 29. PMID: 41159612.
109. Patra P, Gao YQ. Structural and dynamical aspect of DNA motif sequence specific binding of AP-1 transcription factor. J Chem Phys. 2024 Mar 21;160(11):115103. doi: 10.1063/5.0196508. PMID: 38506297.
110. Onyema AC, Dikeocha C, Patel R, Moussa J, Loverde SM. Mapping Allosteric Communication in the Nucleosome with Conditional Activity. J Chem Inf Model. 2026 Mar 9;66(5):2889-2899. doi: 10.1021/acs.jcim.6c00317. Epub 2026 Feb 24. PMID: 41733570; PMCID: PMC12977051.
111. Ning S, Zeng C, Wang H, Zhang J, Xue Y, Zhao Y. The Allosteric Regulation of the DNA-Binding Domain of p53 by the Intrinsically Disordered C-Terminal Domain. Pharmaceuticals (Basel). 2026, 19(1):124. doi: 10.3390/ph19010124. PMID: 41599722; PMCID: PMC12845269.
112. Lawson CL, Swigon D, Murakami KS, Darst SA, Berman HM, Ebright RH. Catabolite activator protein: DNA binding and transcription activation. CurrOpinStruct Biol. 2004, 14(1):10-20. doi: 10.1016/j.sbi.2004.01.012. PMID: 15102444; PMCID: PMC2765107.

113. Prabhakant A, Panigrahi A, Krishnan M. Allosteric Response of DNA Recognition Helices of Catabolite Activator Protein to cAMP and DNA Binding. J ChemInf Model. 2020, 60(12):6366-6376. doi: 10.1021/acs.jcim.0c00617. Epub 2020 Oct 27. PMID: 33108170.
114. Ashkenazi S, Plotnikov A, Bahat A, Ben-Zeev E, Warszawski S, Dikstein R. A Novel Allosteric Mechanism of NF-κB Dimerization and DNA Binding Targeted by an Anti-Inflammatory Drug. Mol Cell Biol. 2016, 36(8):1237-47. doi: 10.1128/MCB.00895-15. PMID: 26830231; PMCID: PMC4836272.
115. Watson LC, Kuchenbecker KM, Schiller BJ, Gross JD, Pufall MA, Yamamoto KR. The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. Nat StructMol Biol. 2013, 20(7):876-83. doi: 10.1038/nsmb.2595. Epub 2013 Jun 2. PMID: 23728292; PMCID: PMC3702670.
116. Yao W, Hu X, Wang X. Crossing epigenetic frontiers: the intersection of novel histone modifications and diseases. Signal Transduct Target Ther. 2024 Sep 16;9(1):232. doi: 10.1038/s41392-024-01918-w. PMID: 39278916; PMCID: PMC11403012.
117. Sinha, S., Vargas, A. M. M., Arantes, P. R., Patel, A., O'Connell, M. R., & Palermo, G. (2023). Unveiling the RNA-mediated allosteric activation discloses functional hotspots in CRISPR-Cas13a. Nucleic Acids Research, 52(2), 906–920. https://doi.org/10.1093/nar/gkad1127
118. Hersey AN, Kay VE, Lee S, Realff MJ, Wilson CJ. Engineering allosteric transcription factors guided by the LacI topology. Cell Syst. 2023, 14(8):645-655. doi: 10.1016/j.cels.2023.04.008. PMID: 37591203.
119. Ridha F, Harini K, Shanmugam NRS, Nikam R, Gromiha MM. Computational design of protein complexes: influence of binding affinity. Chem Commun (Camb). 2026 Jan 13;62(3):752-762. doi: 10.1039/d5cc04821d. PMID: 41311232.