



Cite this: *Phys. Chem. Chem. Phys.*,
2017, 19, 14781

Specific DNA sequences allosterically enhance protein–protein interaction in a transcription factor through modulation of protein dynamics: implications for specificity of gene regulation†

Abhishek Mazumder,^a Subrata Batabyal,^b Manas Mondal,^c Tanumoy Mondol,^b Susobhan Choudhury,^b Raka Ghosh,^a Tanaya Chatterjee,^d Dhananjay Bhattacharyya,^{ib,c} Samir Kumar Pal^{ib} and Siddhartha Roy^{ib,*d}

Most genes are regulated by multiple transcription factors, often assembling into multi-protein complexes in the gene regulatory region. Understanding of the molecular origin of specificity of gene regulatory complex formation in the context of the whole genome is currently inadequate. A phage transcription factor λ -CI forms repressive multi-protein complexes by binding to multiple binding sites in the genome to regulate the lifecycle of the phage. The protein–protein interaction between two DNA-bound λ -CI molecules is stronger when they are bound to the correct pair of binding sites, suggesting allosteric transmission of recognition of correct DNA sequences to the protein–protein interaction interface. Exploration of conformation and dynamics by time-resolved fluorescence anisotropy decay and molecular dynamics suggests a change in protein dynamics to be a crucial factor in mediating allostery. A lattice-based model suggests that DNA-sequence induced allosteric effects could be crucial underlying factors in differentially stabilizing the correct site-specific gene regulatory complexes. We conclude that transcription factors have evolved multiple mechanisms to augment the specificity of DNA–protein interactions in order to achieve an extraordinarily high degree of spatial and temporal specificities of gene regulatory complexes, and DNA-sequence induced allostery plays an important role in the formation of sequence-specific gene regulatory complexes.

Received 22nd February 2017,
Accepted 11th May 2017

DOI: 10.1039/c7cp01193h

rsc.li/pccp

Introduction

Transcription factors are core elements of gene regulatory networks and in the formation of gene regulatory complexes.¹ The high specificity of transcription factors towards their target binding site(s) over a vast amount of non-target genomic DNA is essential for their function.² The target sequences of transcription factors are generally very short compared to the length of the whole genome. The ability of a transcription factor to discriminate between target and non-target sequences of similar size through direct protein–DNA interactions alone is often insufficient for

specific binding in the context of the whole genome. Increasingly, it is becoming clear that the protein–protein interaction between two DNA bound transcription factors can be an important contributor to the specificity of gene regulatory complex formation and consequent transcription regulation.³ However, it is not clear how mechanistically additional specificity is acquired through protein–protein interaction.

It has been known for some time that specific DNA sequences can exert allosteric effects on the DNA-binding protein. How the allostery relates to regulation of gene expression or how such effects are transmitted is not known in general. The phenomenon of allostery—the transmission of the effect of ligand binding to distant sites affecting the function—was first discovered many decades back. Early examples include haemoglobin and aspartate transcarbamylase.^{4,5} For many years, the effect was generally believed to be mediated by protein conformational changes.⁶ In the 1980s, an alternate possibility of transmission of allosteric effects through changes in protein dynamics was first proposed.⁷ However, only in the past few years, have such mechanisms been convincingly demonstrated.^{8–10} To date, most of the examples of dynamics mediated allosteric effects are confined to enzymes and

^a Division of Structural Biology and Bioinformatics, CSIR-Indian Institute of Chemical Biology, 4 Raja S.C. Mullick Road, Kolkata 700 032, India

^b Department of Chemical, Biological & Macromolecular Sciences, S. N. Bose National Centre for Basic Sciences, Block JD, Sector III, Salt Lake, Kolkata 700 098, India

^c Computational Science Division, Saha Institute of Nuclear Physics, Kolkata 700064, India

^d Department of Biophysics, Bose Institute, PI/12, C.I.T. Scheme VII M, Kolkata 700054, India. E-mail: sidroykolkata@gmail.com

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7cp01193h

receptors. Only a few examples of such allostery are known in DNA binding proteins, and even those are largely confined to small molecule ligands.^{11,12}

In this article we report that the protein–protein interaction between two transcription factors is enhanced when they are bound to a pair of naturally occurring binding site sequences as opposed to other pairs of DNA sequences. We show that such an effect differentially stabilizes the protein–protein complex at the natural target sequences. We also show that the DNA-sequence induced allosteric effect in this transcription factor is responsible for this differential stabilization and the allostery is mediated by the change in protein dynamics.

Experimental procedure

Materials

Acrylamide, isopropyl β -D-thiogalactopyranoside (IPTG), ampicillin, commassie brilliant blue, poly(ethyleneimine) and PMSF were purchased from Sigma Chemical Co. (St. Louis, MO). TCEP and glycerol were purchased from Aldrich Chemical Co. (Milwaukee, WI). Acrylodan was obtained from Molecular Probes Inc. All other reagents were of analytical grade quality. Plasmid pEA305 containing the CI clone was a gift from Prof. Mark Ptashne. O_R1 (5'-CGTA CCTCTGGCGGTGATAG-3') and its complementary oligonucleotide were purchased from Trilink (USA) (see Table S1, ESI†). The concentrations of the oligonucleotides were calculated from absorbance values at 260 nm, using extinction coefficients of $14 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$ for purine and $7 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$ for pyrimidine. The individual oligonucleotide strands were mixed at a 1 : 1 molar ratio and annealed by heating to 80 °C followed by cooling down slowly to room temperature. All experiments were carried out at 25 °C in 0.1 M potassium phosphate buffer containing 2 mM EDTA and 1 mM TCEP.

Interaction studies

All fluorescence anisotropy titrations were done in a PTI Quantimaster-6 T-geometry spectrofluorometer at 25 °C. In experiments involving the study of interactions of DNA protein complexes, 0.2 μM of FITC labelled DNA bound- λ -CI complexes was titrated with increasing concentrations of other DNA bound- λ -CI complexes. All titrations were performed at 25 °C in 0.1 M potassium phosphate buffer, pH 8.0 with 1 mM EDTA and 1 mg ml⁻¹ BSA. An estimation of cooperativity was done by fitting the binding isotherm to the following equation:

$$A_{\text{obs}} = \frac{A_1(1 + 0.5(K_1 + K_2)P) + A_2(0.5(K_1 + K_2)P + \alpha K_1 K_2 P^2)}{1 + K_1 P + K_2 P + \alpha K_1 K_2 P^2}$$

where A_{obs} is the observed anisotropy, A_1 is the initial anisotropy, A_2 is the anisotropy at infinite protein concentrations and α is the cooperativity factor. Full derivation and its use are described by Mazumder *et al.* (2012).¹³

DTNB reaction

The formation of a thionitrobenzoate anion by DTNB reaction with free sulfhydryl groups of lambda repressor was monitored by measuring the absorbance at 412 nm over time.¹⁴ All the

absorbance measurement experiments were carried out in 0.1 M of potassium phosphate buffer, pH 7.0, containing 1 mM EDTA at 25 °C. Each protein sample was mixed with excess DTNB in a quartz cuvette and the absorbance was recorded continuously at 412 nm using a Perkin Elmer LS50 spectrophotometer for 1 hour. Nonlinear least squares fits were done using Origin 8.5 and fitted to the equation:

$$A(412) = A_0 + n \times C \times 13\,600(1 - \exp(-k \times t))$$

where A is the absorbance, A_0 is the absorbance at zero time, n is the number of reacting sulfhydryl groups, C is the concentration of the repressor, 13 600 is the extinction coefficient of thionitrobenzoate (TNB⁻), k is the rate constant, and t is the time.

Chemical modification of G186C

Acrylodan labelling was carried out as described previously.¹⁵ This was followed by ultracentrifugation at 100 000 rpm for 45 minutes to remove any protein aggregates that may have been present. The protein concentration of the modified protein was determined by A_{280} measurement with an appropriate subtraction for the incorporated acrylodan.

Calculating the rotational correlation time of λ -CI-O_L1 and the free dimer

The rotational correlation time of the λ -CI-O_L1 complex and that of the free λ -CI dimer was estimated using the program HYDROPRO.¹⁶ The pdb entry 3BDN was used as the guiding crystal structure. The DNA coordinates were removed from the pdb file to generate the free dimer structure. The partial specific volume for the protein was taken to be 0.73 and that for the protein–DNA complex was estimated to be 0.69. The solution viscosity was assumed to be 1 cP and the temperature was fixed to 298° K.

Fluorescence spectroscopy

All steady state fluorescence studies were done in a PTI Quantimaster-6 T-geometry spectrofluorometer at 25 °C. The experiments were carried out either in 1.0 cm or in a 0.5 cm path length quartz cuvette. For anisotropy experiments, the excitation wavelength was at 480 nm and the emission was at 530 nm, while the excitation and emission band passes were 5 nm each.

All fluorescence decays were measured by the picosecond-resolved time-correlated single photon counting technique. A commercially available picosecond diode laser pumped time-resolved fluorescence spectrometer setup from Edinburgh Instrument, UK, was used. It has an instrument response function (IRF) of 50 ps. The pico-second excitation pulse from a Picoquant diode laser was used at 375 nm. A liquid scatterer was used to measure the FWHM of the IRF. Fluorescence from the sample was detected by a micro channel plate photo multiplier tube (Hamamatsu) after dispersion through a grating monochromator.

For anisotropy ($r(t)$) measurements, emission polarization was adjusted to be parallel or perpendicular to that of the excitation, and the decays for I_{vv} and I_{vh} are recorded. Anisotropy at time t is defined as

$$r(t) = (I_{\text{vv}}(t) - G \times I_{\text{vh}}(t)) / (I_{\text{vv}}(t) + 2 \times G \times I_{\text{vh}}(t))$$

G , the grating factor was assumed to be 1. Nonlinear least squares fits of the anisotropy decay profiles that were obtained were done using an equation describing bi-exponential decay using Origin 8.5. The obtained decay curves fit well to the following bi-exponential equation:

$$R(t) = R_0 [A \times \exp(-x/t_1) + (1 - A) \times \exp(-x/t_2)] \quad (1)$$

where $R(t)$ is the anisotropy at time t , R_0 is the limiting anisotropy, t_1 and t_2 are the fast and slow components of the anisotropy decay, and A is the amplitude of the fast component. For all fluorescence decay experiments, the protein concentration was 1 μM in terms of the dimer. The DNA concentrations were stoichiometric.

MD simulations

Simulation details and analysis of the trajectories. We used molecular dynamics (MD) simulations with explicit solvent (water and charge neutralizing counter ions) to study the role of conformational fluctuations of the two sub-units of lambda repressor protein in dimeric binding with DNA. The coordinates of the protein dimer and the dimer bound with DNA were taken from the Protein Data Bank (PDB)¹⁷ with the PDB-ID 3BDN. We modelled residues 215 (Cys) and 216 (Asn) on both the protein subunits, which are missing in the crystal structure, considering all trans conformation as initial guess and energy minimized with CHARMM.¹⁸ Their nearby residues adopt unstructured coil conformation; these are far from the DNA binding region and also from the dimer contacts. Initial set up and minimization of the systems was done using the AMBER 8 suite of programs¹⁹ with parm94²⁰ parameters, which is one of the most widely used parameter sets for biomolecular simulation.^{19,21} In the past few years, few studies have revised this force field by modifying the torsion potentials associated with a few dihedral angles with fitting to additional quantum-level and NMR derived data.^{22–26} On the other hand many reports also exist in the literature indicating the effectiveness of the parm94 force field in biomolecular simulations and for qualitative study using this force field.^{27–31} Considering the controversy about the force field effect, we also considered the recent ff14SB force field³² of AMBER in our study. Each system was solvated in an orthorhombic water box containing TIP3P water molecules and required numbers of sodium ions to maintain the electro neutrality. The systems were solvated in such a manner that there was at least a 15 Å thick layer of water around the solute in all directions. These systems were then energy minimized for 20 000 cycles using a combination of steepest descent and conjugate gradient algorithms and applying periodic boundary conditions. Long range electrostatic energy was calculated using the particle mesh Ewald summation with 1 Å grid spacing and a 10^{-6} convergence criterion. Lennard-Jones and short-range electrostatic interactions were truncated at 10 Å. All the MD simulations were carried out using the NAMD.³³ Initially, for each of the minimized systems, heating to 300° K was carried out slowly during 30 ps with 1 fs time steps. We continued the simulations of the two systems for 180 ns each considering parm94 and for 100 ns each considering the ff14SB

force field at constant temperature (300° K) and pressure (1 bar) using the Langevin-Piston algorithm and periodic boundary conditions. Translational and rotational movements of the centre of mass were removed at intervals of 5 ps. SHAKE constraints were applied to all bonds involving hydrogen atoms. The integration time step was 1 fs. From the extended trajectories the conformations of each of the systems were saved every 1 ps for further analysis. Simulations considering the ff14SB force field are mainly discussed in the main text.

Structural analysis. Root mean square displacement (RMSD) and root mean square fluctuation (RMSF) analyses of the trajectories were performed using CHARMM.¹⁸ Equilibration was ensured from the RMSD variations of the biomolecules. It indicated that after 10 ns the RMSD values converge for both the subunits in the operator bound and unbound states. Therefore, the trajectories after the first 10 ns were considered for further analysis. Protein secondary structure analysis was done using CHARMM following the DSSP³⁴ algorithm for assignment of the secondary structure (α -helical, β -sheet and coil).

Study of conformational entropy. The entropy calculations for the protein subunits were performed using the quasiharmonic analysis method³⁵ implemented in CHARMM. The entropies and their contribution to the free energy (TS) were calculated at 300° K. Along with the main chain, side chain atoms also play an important role in protein dynamics. Therefore, to see the change in protein dynamics throughout the protein chain due to binding with the operator DNA, we calculated the entropy considering all the non-hydrogen atoms of the protein residues.

Along with quasiharmonic analysis, different methods exist to estimate the conformational entropy from the conformational variables of macromolecules. Recently, the dihedral angles of proteins have been widely used as conformational variables. Multidimensional histograms of side chain dihedral angle distributions have been constructed to estimate the conformational entropy.^{36,37} This method, considering correlations among the side chain dihedral angles up to a different order, can give a good estimation of the conformational entropy.^{38,39} However, in biomolecules, long-ranged dihedral correlations have been found to be negligible except for some short-ranged correlations among the side-chain torsions.^{40–43} This also illustrates the significance of completely reduced one-dimensional histograms based on a single dihedral angle. Histograms of a particular side chain dihedral angle indicate the probability of finding the dihedral angle in a given conformation and it can be interpreted as given by the Boltzmann factors of the corresponding effective free energies, while the entropies can be estimated by the Gibbs formula.^{37,44} According to the Gibbs entropy formula, the conformational entropy for a particular side chain dihedral angle, ζ

$$S_{\text{conf}}(\zeta) = -k_B \sum_i H_i(\zeta) \ln H_i(\zeta)$$

where the sum is taken over the histograms of bins i with a nonzero value of probability distribution function, H_i .

We calculated the side chain conformational entropy from the histograms of the side chain dihedral angles χ_1 and χ_2 . We restricted the histogram of χ_1 and χ_2 for each residue within 0°

to 360° and divided it into 90 equal bins. The change in side chain conformational entropy for χ_1 and χ_2 , due to binding of the repressor protein with operator DNA, is

$$\Delta S_{\text{conf}}(\zeta) = -k_{\text{B}} \left[\sum_i H_i^{\text{bound}}(\zeta) \ln H_i^{\text{bound}}(\zeta) - \sum_i H_i^{\text{free}}(\zeta) \ln H_i^{\text{free}}(\zeta) \right]$$

Using the above expression, we studied the change in side chain conformational entropy throughout the protein chain due to binding of the λ -CI dimer with operator DNA.

Modelling of anisotropy decay. The fast part of the anisotropy decay data was modelled using a wobbling-in-cone model.⁴⁵ According to this model, the fluorescence anisotropy decay is a product of three independent motions:

- (a) wobbling of the probe $r_{\text{w}}(t)$ with a time constant t_{w} ,
- (b) translation of the probe $r_{\text{D}}(t)$, along the surface of the protein, with a time constant t_{D} ,
- (c) overall rotation $r_{\text{P}}(t)$ of the protein with a time constant t_{P} .

In this case the acrylodan is covalently attached to a cysteine in the protein so we may neglect the motion associated with the translation of the probe along the surface of the protein. Thus, one may decompose $r(t)$ as a product of two independent motions

$$r(t) = r_{\text{w}}(t) r_{\text{P}}(t) \quad (2)$$

$r(t)$ may be written in terms of the order parameter S as

$$r(t) = r_0[S^2 + (1 - S^2)\exp(-t/t_{\text{w}})]\exp(-t/t_{\text{P}}) \quad (3)$$

where S is related to the semi-cone angle θ in the wobbling in cone model^{46,47} as

$$S = 0.5 \cos \theta (1 + \cos \theta) \quad (4)$$

Comparing eqn (1) and (3) one obtains

$$S^2 = (1 - A) \quad (5)$$

$$1/t_1 = 1/t_{\text{w}} + 1/t_{\text{P}} \quad (6)$$

$$1/t_2 = 1/t_{\text{P}} \quad (7)$$

Considering t_{P} to be equal to the overall rotational correlation time for the protein it can be readily seen that $1/t_{\text{P}}$ is negligibly small compared to $1/t_1$. So the wobbling time constant, t_{w} , can be safely assumed to be equal to t_1 obtained from eqn (1).

Cloning and purification of the mutant repressors. The Y60C and G186C mutants were created by site-directed mutagenesis on the pEA305 plasmid containing the gene for λ -CI using a Quikchange Site Directed Mutagenesis Kit from Stratagene. The clones were verified by DNA sequencing of the whole gene. Plasmids containing the Y60C and G186C clones were then transformed into XL1B cells, grown at 37 °C and induced in a manner similar to that used for the wild type λ -CI. The respective proteins were then purified according to Saha *et al.*⁴⁸

Results and discussion

The lytic-lysogenic switch of bacteriophage λ has long been a model system for the study of gene regulatory networks and has some characteristics of gene regulatory networks of higher eukaryotes.⁴⁹ Maintenance of the lysogenic state is achieved by two inter-convertible multi-protein-DNA complexes, each consisting of several interacting molecules of the transcription factor λ -CI (Fig. 1(A)). In λ -CI, as in many other prokaryotic and eukaryotic transcription factors, DNA binding and protein-protein interaction domains are distinct.⁵⁰ In this protein, specific recognition of DNA sequences occurs through the N-terminal domain, whereas interactions with other sequence-specifically bound repressor molecules to form the gene regulatory multi-protein complex occur through the C-terminal domain^{50,51} (Fig. 1(B)). Among the six binding sites of λ -CI in the bacteriophage genome, we primarily focused on $O_{\text{R}1}$ and $O_{\text{R}2}$ (the others being $O_{\text{R}3}$, $O_{\text{L}1}$, $O_{\text{L}2}$ and $O_{\text{L}3}$). To address the question of what influences DNA sequences have on protein-protein interaction, we previously used sedimentation equilibrium studies where the self-association of $O_{\text{R}1}$ - λ -CI was found to have a dissociation constant of $41.6 \pm 0.8 \mu\text{M}$. In contrast, $O_{\text{R}1}$ - λ -CI association with $O_{\text{R}2}$ - λ -CI was found to have a dissociation constant of $4.4 \pm 0.14 \mu\text{M}$.⁵² Thus, the former interaction is significantly weaker than the latter, which has the DNA ligand sites which are contiguous and functional in the natural context. The difference between self-association and hetero-association is thus clear.

We also attempted to see if the DNA sequence modulates the protein-protein interaction energy between the two λ -CI dimers when bound in *cis*; that is, on the same piece of DNA (like in the genome). Fig. 2 shows the binding of λ -CI to an oligonucleotide containing two $O_{\text{R}1}$ sites separated by the same base pairs as between $O_{\text{R}1}$ and $O_{\text{R}2}$ in the natural context. The binding is distinctly weaker when compared to an otherwise identical oligonucleotide containing $O_{\text{R}1}$ and $O_{\text{R}2}$.¹³ When fitted to a binding equation that includes a cooperativity parameter, the $O_{\text{R}1}$ - $O_{\text{R}1}$ oligonucleotide gives an average mid-point of transition about $38.1 \pm 4.4 \text{ nM}$, whereas the $O_{\text{R}1}$ - $O_{\text{R}2}$ construct previously showed an average mid-point of transition about 10 nM .¹³ An estimate of the cooperativity factor for the former is 0.93, indicating no cooperativity (a value of 1 indicates no cooperativity and higher values indicate its presence). In the previous paper the cooperativity factor for $O_{\text{R}1}$ - $O_{\text{R}2}$ binding was reported to be around 10. Clearly, the cooperativity between the two dimers is significantly reduced, if not lost completely, suggesting that the pairing of correct binding sites is important in the *cis* configuration as well. Thus, we conclude that the correct DNA sequences can allosterically modulate protein-protein interactions through a domain distant from the DNA-protein interaction site.

How the information about recognition of the target DNA sequence is allosterically transmitted to the protein-protein interaction interface leading to the modulation of interaction energy is not known. The established notion in the field is that allosteric effects in proteins are mediated by a change of conformation.⁵³ However, in recent times we have seen the

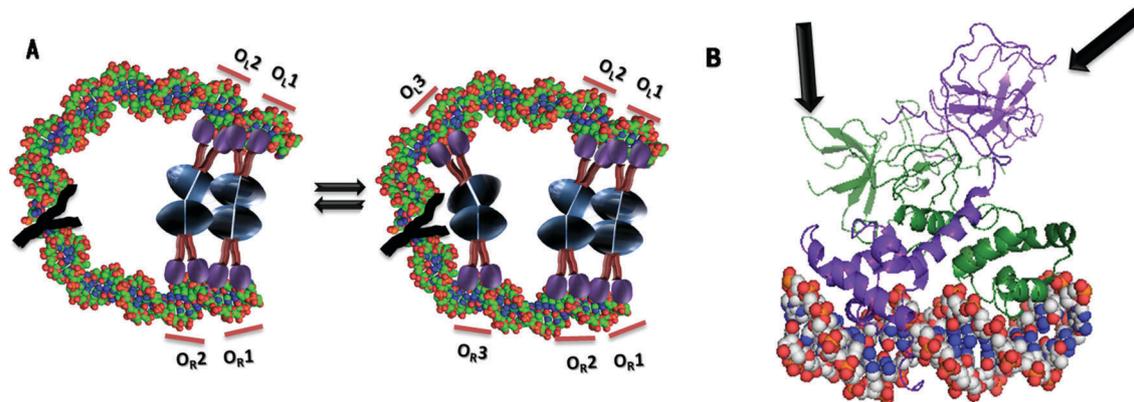


Fig. 1 (A) A cartoon figure depicting the architectures of the two multi-protein complexes of bacteriophage λ that maintain the lysogenic state. The black wiggly lines signify longer DNA sequences. The lavender/red coloured objects are λ -CI subunits. The octamers are known to bind to O_{R1} – O_{R2} and O_{L1} – O_{L2} regions, whereas the additional tetramer is thought to bind to O_{R3} and O_{L3} regions, further stabilizing the complex. The octameric complex is believed to repress all genes except that of the λ -CI, whereas the dodecameric complex represses the λ -CI gene as well. (B) Crystal structure of the λ -CI– O_{L1} complex (pdb 3bdn). The two subunits are in different colours. The lavender coloured subunit binds to the non-consensus half (subunit A), whereas the green coloured subunit binds to the consensus half (subunit B). The arrows indicate approximate locations of the protein–protein interaction surfaces.

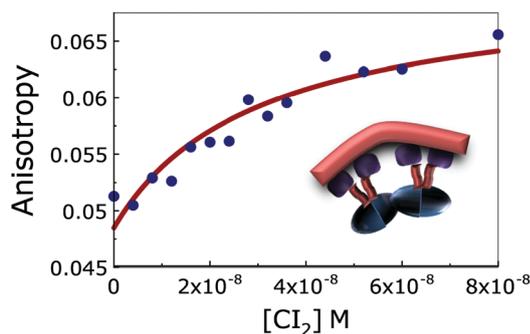


Fig. 2 Binding isotherm of unlabeled λ -CI with an oligonucleotide (sequence: 5'TATCACC GCCAGAGGTA AACCATACCGTTACCTCTGGCGGTGATA3' and its complementary strand) containing two O_{R1} sites with the same orientation and separation as that of the wild-type O_{R1} – O_{R2} site (*cis*). The red line represents the best-fit to the equation for estimation of cooperativity as stated above. 2 nM of FITC labelled oligonucleotide was titrated with increasing concentrations of λ -CI. All titrations were performed at 25 °C in 0.1 M potassium phosphate buffer, pH 7.0, containing 0.1 mM EDTA and 1 mg ml^{−1} BSA. The data shown are the average of three independent experiments. The inset shows a cartoon figure emphasizing the nature of the complex that is forming under these conditions.

emergence of protein dynamics as an important mechanism for transmission of allosteric effects.⁵⁴

Conformation around the dimer–dimer interface is not significantly changed upon DNA binding

In order to investigate how binding to the specific DNA sequence allosterically modulates protein–protein interaction, we studied the conformation and dynamics of λ -CI at or near the site of the protein–protein interaction in the C-terminal domain. For this purpose, a cysteine residue was introduced through site-directed mutagenesis into the C-terminal domain (G186C) of the protein, in close proximity to the dimer–dimer interaction surface, which is responsible for the cooperative assembly of λ -CI (Fig. 3A).^{55,56} Three naturally occurring cysteine

residues in the wild-type λ -CI are all un-reactive⁵⁷ and hence the reaction of λ -CI–G186C with a sulfhydryl reactive fluorescence probe will label C186 exclusively. DTNB reaction of λ -CI–G186C indeed showed only approximately one reactive sulfhydryl per subunit (data not shown). Acrylodan is a fluorophore whose fluorescence is highly sensitive to its environment and it shows large (tens of nanometers) shifts of its emission maximum upon a change in environmental polarity.⁵⁸ The acrylodan conjugated λ -CI–G186C shows a peak at around 502 nm in the emission spectra, but does not exhibit any significant shift of maximum upon the addition of O_{R2} and only exhibits a very minor shift upon the addition of O_{R1} (Fig. 3B). This suggests that upon complex formation with the target DNA, the polarity of the environment and, by proxy, the conformation around residue 186 remains largely unperturbed. The binding of proteins to DNA is sometimes coupled to folding–unfolding events.⁵⁹ In order to find out whether there was any unfolding event coupled to the DNA binding in this case, we compared the backbone structure of the C-terminal domain in the unbound state (pdb2hnf) and in the DNA bound state in the full length protein (3bdn). The two structures align well, suggesting no significant alteration of backbone conformation, and hence no coupled folding–unfolding event, upon DNA binding (Fig. 3C).

Enhanced dynamics in the C-terminal domain upon O_{R1} binding

Since conformational change around the protein–protein interaction site was not detected, we explored whether the dynamics of protein residues around the site was affected. Nuclear Magnetic Resonance is probably the most useful experimental technique for studying protein dynamics at the residue level. However, λ -CI is unsuitable for NMR studies as it aggregates at low micro-molar concentrations.⁶⁰ Thus, for a direct examination of the dynamics near the protein–protein interaction interface of λ -CI, the time-resolved anisotropy decay of fluorescent probes

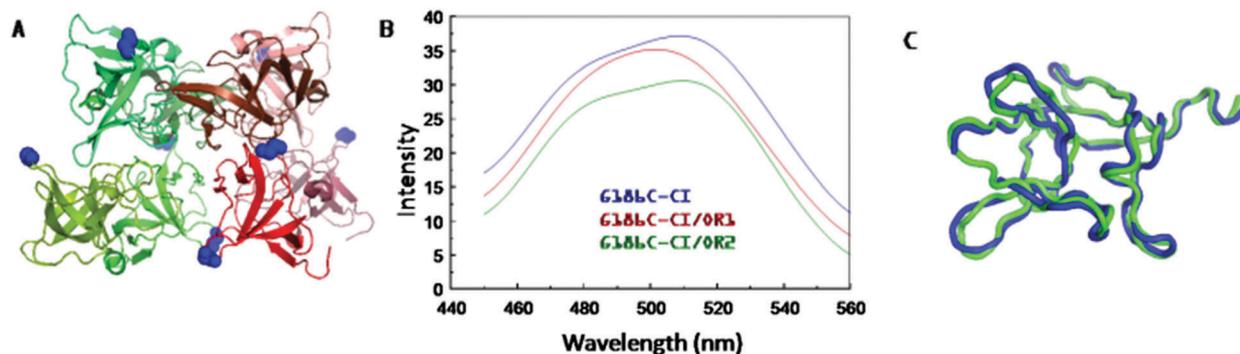


Fig. 3 (A) Ribbon representation of the octamer of the C-terminal domain (1KCA), showing the proximity of G186 (space fill blue coloured residues) to the dimer-dimer interface. (B) Change in the emission spectra of the λ -CI-C186-acrylodan conjugate upon binding of O_{R1} and O_{R2} . (C) Alignment of the C-terminal domains of the A subunit of 3bdn and the A subunit of 2hnf (isolated C-terminal domain of λ -CI). The two chains are coloured blue (3bdnA) and green (2hnf). The alignment was done using the web server Superpose.

directly attached to G186C on the λ -CI was determined at protein concentrations where it was predominantly dimeric. In principle, it is possible to extract dynamical information from the time-resolved decay of anisotropy, $r(t)$, and to estimate diffusion coefficients arising from different kinds of motions.^{61,62} Depolarization of fluorescence anisotropy in the time span of <1 ns generally results from very fast motions largely reflecting the dynamics of the attached probe and the amino acid to which it is conjugated. Depolarization in the longer time-scale reflects motions involving larger segments bearing the conjugated amino acid and the overall tumbling motion of the protein.

In contrast to the invariance of the emission wavelength of the environment sensitive probe acrylodan attached to C186, the steady state anisotropy value at 500 nm decreases from 0.200 to 0.175 upon binding to O_{R1} . The measured average lifetime for the labelled acrylodan in the G186C dimer and the G186C- O_{R1} complex is similar (2.27 and 2.20 ns, respectively; Table S2, ESI[†]). Thus, this decrease of steady-state anisotropy may be due to a greater motional freedom of the acrylodan probe. In order to obtain a more detailed picture of the dynamics of the protein, time-resolved anisotropy decay studies of the G186C-acrylodan conjugated λ -CI were performed at an emission wavelength of 500 nm (Fig. 4). The decay profiles were fitted to a bi-exponential equation and were analyzed using the wobbling-in-cone model.^{45,47} The time constants for the wobbling of the probe (t_w) were calculated to be 169 ps and 126 ps for the G186C- λ -CI dimer and the O_{R1} -G186C- λ -CI complex, respectively, suggesting somewhat faster rotational diffusion in the O_{R1} complex. The semi-cone angle available for the wobbling of acrylodan in the free repressor dimer is about 18.3 degrees while that in the O_{R1} - λ -CI complex is somewhat higher (21.9 degrees), indicating that upon binding to O_{R1} , a modest increase in acrylodan mobility occurs. The slower part of the anisotropy decay (t_2) largely arises from the overall rotational diffusion or internal motions of segments of the protein, which is significantly larger than the probe. The acrylodan labelled λ -CI-G186C- O_{R1} complex exhibits a lower t_2 value (41 ± 1.5 ns) relative to the free λ -CI-G186C-acrylodan conjugate (52 ± 2.7 ns) (Table S3, ESI[†]). For a rigid molecule, the reduction in time

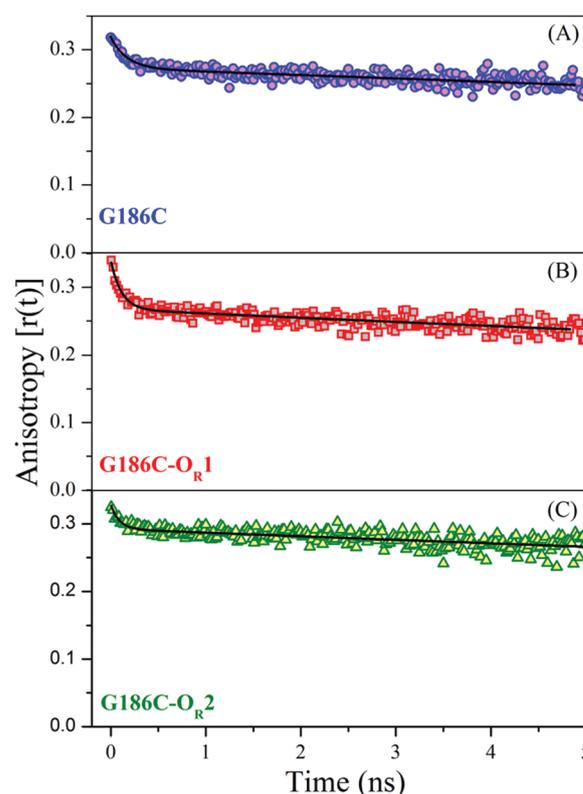


Fig. 4 Fluorescence anisotropy decay, $r(t)$ profiles of acrylodan in (A) acrylodan-G186C- λ -CI, (B) the acrylodan-G186C- λ -CI- O_{R1} DNA complex and (C) the acrylodan-G186C- λ -CI- O_{R2} DNA complex. The data were fitted to a bi-exponential decay equation using Origin 8.5.

constants of the decay is counterintuitive as binding of the DNA target sites to the repressor should result in the formation of a larger complex with a higher hydrodynamic radius and consequently a longer rotational correlation time. To compare the observed rotational correlation time to the expected rotational correlation time, the rotational correlation times were calculated with the aid of the program HYDROPRO from the crystal structures.⁶³ The calculated rotational correlation times were 46 and 53 ns for the free protein and the protein- O_{R1} complex, respectively.

The calculated value for the free protein agrees reasonably well with the observed value, suggesting that the free protein tumbles as a whole in solution and may have only relatively small segmental motions in this region. Interestingly, the observed rotational correlation time of the protein- O_{R1} complex is significantly lower than that of the corresponding calculated rotational correlation time, suggesting a very significant increase in segmental fluctuations upon O_{R1} binding. This conclusion is also in agreement with the reduction of the steady-state anisotropy value upon O_{R1} binding. Taken together, the results of the anisotropy decay experiments indicate that significant segmental motions are induced in the C-terminal domain of the λ -CI upon binding of O_{R1} in the N-terminal domain, although the average structure (as reflected by conformation) remains largely unchanged, suggesting transmission of the allosteric effect through changes in the protein dynamics. In a previous article, we reached a similar conclusion with a probe, dansyl, whose site-specificity was not known.⁵²

O_{R2} bound λ -CI shows distinctly different dynamic character

In the phage genome, λ -CI binds to six target sites, of which O_{R2} is a crucial site for repressive complex formation (see Fig. 1A). O_{R2} differing from O_{R1} at seven base pairs out of seventeen has significantly lower affinity and distinctly different thermodynamics of binding.⁶⁴ Since the dynamics of the protein are significantly changed upon binding to O_{R1} , we decided to explore whether binding to different target DNA sequences induces different dynamic character in the protein. Anisotropy decay of G186C-acrylodan- λ -CI in complex with O_{R2} was recorded (Fig. 4C).

When the fast part of the decay was fitted to the wobbling-in-cone model, the time constant was found to be 80 ps and the semi-cone angle was found to be 15.9 degrees. The semi-cone angle was somewhat smaller than that of the free λ -CI, whereas the time constant indicated a somewhat faster motion. When the slower part of the decay was analyzed, it yielded a rotational correlation time of 53 ± 2.4 ns, not different from the HYDRO-PRO predicted value of 53 ns, indicating a lack of enhanced segmental motions like that seen in the O_{R1} - λ -CI complex. Clearly, the segmental dynamics that resulted in the rotational correlation time of 41 ± 1.5 ns in the O_{R1} bound λ -CI are largely absent in the O_{R2} bound form.

Fluctuations increase in many regions of the protein upon DNA binding

In order to understand the nature of the dynamical changes that occur upon target site binding, we explored protein dynamics in the DNA-protein complex as well as in the free protein dimer using molecular dynamics. The starting point of the molecular dynamics study is the O_{L1} - λ -CI complex. This is the only λ -CI/DNA complex whose crystal structure is available. The operator site used in the solution experiments mentioned above used O_{R1} , which differs from O_{L1} by only one base pair in the non-consensus half and has a similar affinity.^{64,65} O_{R1} was used experimentally as most of the solution experiments were previously done with O_{R1} . We assume that for the purpose of this study, O_{R1} and O_{L1} behave similarly. λ -CI is an asymmetric homodimer in which the two subunits have different conformations,

although they have the same primary structure.^{50,66} The main difference between the two subunits in the λ -CI dimer lies in the conformation of the hinge, the peptide segment (approximately, residues 93–132) that connects the N-terminal DNA binding domain (1–92) with the C-terminal protein-protein interaction domain (133–236). In one subunit, the hinge is significantly more compact than in the other. We call this subunit SH or S (for short hinge; Chain B in 3bdn; coloured green in Fig. 1B) and the other LH or L (for long hinge; Chain A; coloured purple in Fig. 1B). The short hinged subunit interacts with the consensus half-site of O_{L1} , the half-site that is most highly conserved among all the binding sites in the lambda genome.

Initially, the structure was fully equilibrated and then the fluctuations of atoms were studied through 100 ns of molecular dynamics simulations. To understand the effect of target DNA binding on the protein dynamics, the O_{L1} DNA was removed from the complex structure and the free protein dimer was subjected to molecular dynamics simulations. Fig. 5 shows the comparison of root-mean-square-fluctuations of all the protein residues of the two subunits considering all non-hydrogen atoms in the presence and absence of DNA. Noticeably, many C-terminal domain residues show enhanced fluctuations in the subunit S (for example, around residues 185, 200 and 215 in the subunit S (Chain B) and in the hinge region of subunit L (Chain A)).

Some of these residues are in regions thought to be involved in protein-protein interactions.⁵⁵ We then explored other measures of disorder along the chain using side-chain parameters. $T\Delta S$ values of each residue were calculated using all non-hydrogen atoms. Fig. S1 (ESI[†]) shows the difference of $T\Delta S$ values for each residue between the DNA bound and the free state. Most of the N-terminal residues in Chain A (subunit L or LH) show a reduction in entropy due to additional interaction with the DNA, but many residues in the C-terminal domain, including the previously mentioned residues around the protein-protein interaction site, showed augmented entropy values. Chain B (subunit S or SH) shows a similar picture in the C-terminal domain, but interestingly it shows less motional quenching in the N-terminal domain. In the ESI,[†] Fig. S2, $T\Delta S$ values of each residue were calculated from side chain torsion angle (χ_1) distributions. A similar enhanced propensity for disorder was observed in the C-terminal domains of both the subunits upon DNA binding. Thus, an important conclusion may be drawn that many regions of the C-terminal domain, including the residues at or near the protein-protein interface, attain enhanced mobility upon DNA binding. A similar change in protein dynamics was observed with another force field and a longer simulation time (ESI,[†] Fig. S3–S5). Interestingly, the propensity for higher disorder was more pronounced in the C-terminal domain of the S-subunit.

Quenching of fluorescence by acrylamide is widely used for measuring the accessibility of tryptophan residues in proteins. For buried tryptophan residues, this accessibility is mostly due to transient opening of the protein structure, thus creating access for the quenching molecule.^{67,68} Since all the tryptophan residues in the protein are situated in the C-terminal domain, distant from the DNA binding interface and found to be mostly

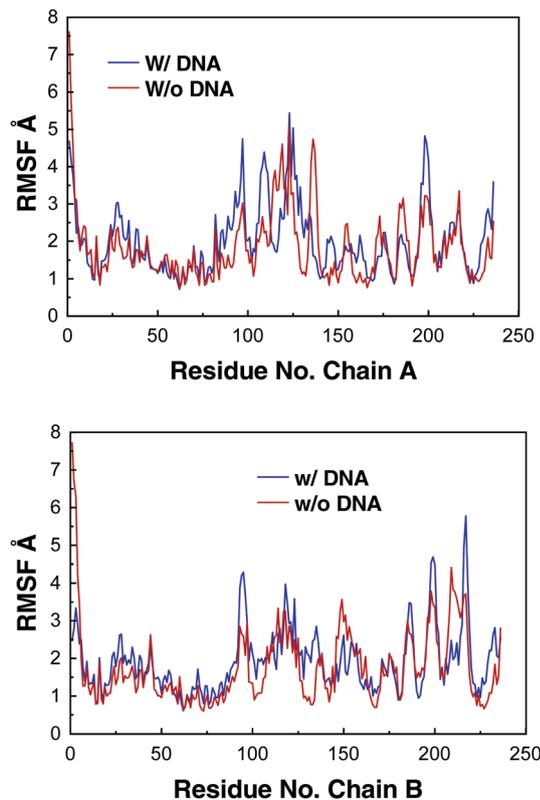


Fig. 5 Root-mean-square fluctuations of residues with and without O_L1 in the L-subunit (chain A) and S-subunit (chain B) (simulation corresponding to the ff14SB force field).

buried in the crystal structure (Fig. 6A), acrylamide quenching can be used in this situation to obtain information about the dynamical character of the regions harbouring the tryptophan residues, that is the C-terminal domain. Previous studies have indicated a distinctly different pattern of quenching of tryptophan fluorescence in the O_{R2} bound form when compared to the O_{R1} bound form.⁶⁹

To elucidate whether a change of DNA sequence correlates with the dynamic character of the distant C-terminal domain, we systematically introduced substitutions in the O_{R1} sequence—one on top of another—to move towards the O_{R2} sequence. These operator site sequences were studied for their effect on the quenchability of tryptophan fluorescence using acrylamide. Fig. 6B shows a plot of Stern–Volmer constants (obtained from the initial slopes of the quenching curves) as we move from O_{R1} towards the O_{R2} sequence. The K_{sv} values fall as we cumulatively introduce substitutions and make the sequence more and more O_{R2} -like. The fall is more pronounced with the first few substitutions starting from the non-consensus end. Apparently, the bases situated towards the end of the non-consensus site have a more profound influence on the dynamic character of the C-terminal domain than the central ones. Thus, we may conclude that the C-terminal domain in the O_{R2} bound form is dynamically “colder” than that in the O_{R1} bound form and two protein molecules with different dynamical signatures produce a more enhanced interaction than other combinations.

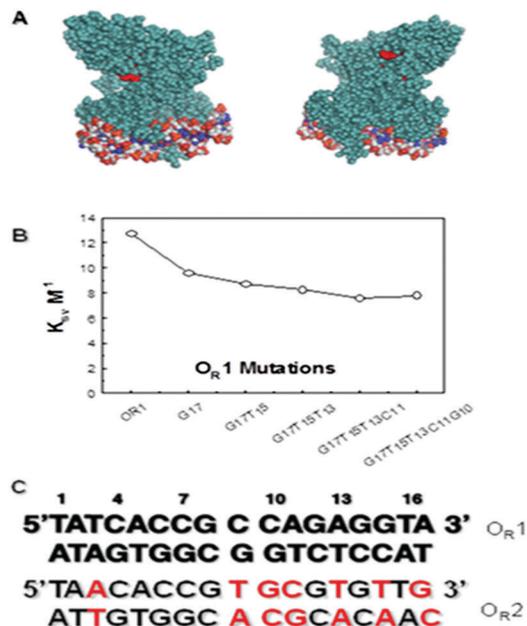


Fig. 6 (A) Space filling model of two faces of the λ -CI/ O_L1 complex (3bdn) with tryptophan residues coloured red. Out of six, only two are partially visible. (B) Stern–Volmer constants obtained from acrylamide quenching of tryptophan fluorescence of λ -CI/ O_{R1} as a function of progressive basepair substitution of O_{R1} . Only the quenching constants at the lower concentrations of acrylamide (initial part of the plot) were plotted. (C) Sequences of O_{R1} and O_{R2} . Differences are marked in red on O_{R2} .

Dynamic correlation analysis shows a pathway for the allosteric transmission of information

It was observed clearly from anisotropy decay studies that the nature of the DNA sequence influences the dynamic nature of the residues in the distant C-terminal domain at or around the protein–protein interaction site. To elucidate how this information transfer occurs, we analyzed how motion in each residue was correlated with that of other residues. A dynamic Cross-Correlation Map has been used extensively to trace the allosteric transmission pathway.^{70,71} Fig. 7A shows the Dynamic Cross-Correlation Map (DCCM), derived from the MD data of both the DNA bound subunits. Among other correlations, the map shows a pathway (circled) in which a strong correlation path may be observed from residues near the DNA interface to residues near the protein–protein interface. The residues that are in proximity to the DNA interface are residues around K39 in the L-subunit. Fig. 7B maps the highly dynamically correlated residues starting from the K39: L-subunit onto the structure of the λ -CI- O_L1 complex, clearly indicating a pathway of information transfer from the protein–DNA interface to the protein–protein interface. How this alteration of dynamics affects protein–protein interaction is not clear. One possibility is that certain “hotspot” residues require flexibility to interact, and a change in dynamics imparts such flexibility.⁷²

Discussion

When transcription factors bind to their target sequence in the regulatory regions of a gene, this initiates subsequent processes,

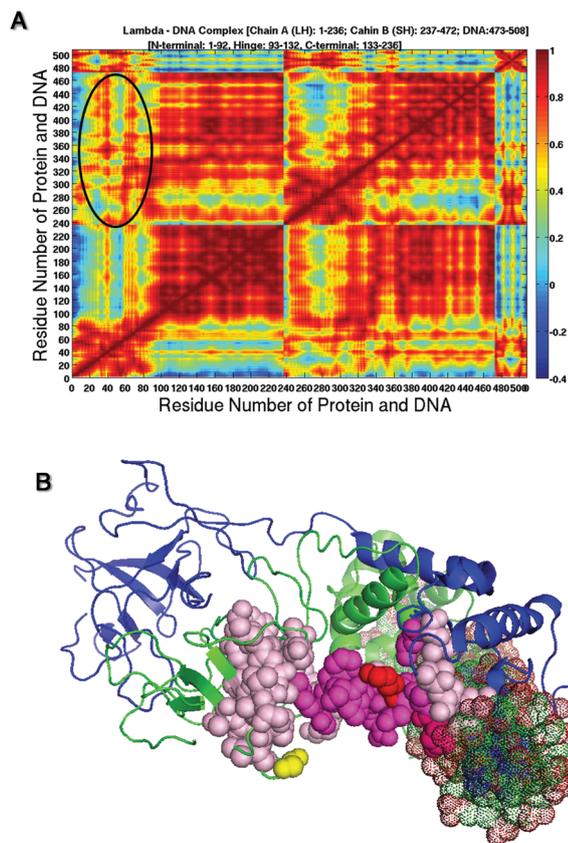


Fig. 7 (A) Dynamic cross correlation map of O_{L1} bound λ -CI. The circled part shows a series of highly dynamically cross-correlated residues from around residue 39 of the L-subunit (Chain A) to the residues in the C-terminal domain of the S-subunit (Chain B). (B) These highly cross-correlated residues are mapped onto the crystal structure of the λ -CI/ O_{L1} (3bdn). The cross-correlated residues in this path are shown as spheres. DNA is shown as translucent spheres (dotted). The L-subunit is coloured blue, whereas the S-subunit is coloured green. The red spheres represent residue 39 of the L-subunit, the magenta spheres represent residues with a cross-correlation coefficient of greater than 0.9, and the light pink coloured residues have cross-correlation coefficients between 0.8 and 0.9. The yellow coloured residue is G186 of the S-subunit.

which lead to regulation of the expression of that gene. The canonical model of such regulation is that the target DNA sequence plays an anchoring role for the localization of the transcription factor and initiation of site-specific processes. However, increasing evidence points towards additional roles of the DNA, *e.g.*, changing transcription factor conformation and, as a consequence, some down-stream effects.^{73,74} This allosteric change exerted by DNA sequences and its functional consequence is neither widely appreciated nor mechanistically well understood.

To understand the functional significance of the allosteric change in the transcription factor induced by the DNA sequence, we need to explore quantitatively its effect on the specificity of recognition of the target site. A transcription factor has to bind a target site in the genome in face of competition from a large number of non-target sites. The differential binding energy between the target and the non-target sites is a crucial factor in site discrimination, which depends on the number of base pairs

in the target site.⁷⁵ Annexure I (ESI^\dagger)^{76,77} presents a quantitative approach to study these effects.

An observed phenomenon that may be important in enhancing specificity is the multi-partite nature of many transcription factor binding sites in prokaryotes.⁷⁸ Multi-partite binding of many transcription factors in prokaryotes leads to a protein–protein interaction with the formation of a DNA loop, thus, energetically coupling two separate binding events. However, unless such loops form only on the specific sites, enhancement of specificity becomes non-existent, because if such loops form on non-specific sequences, an additional adverse effect will become operative. The number of possible loops that can form for each non-specific site is large. Hence, it would create an unfavourable entropic factor and diminish the specificity. For example, for each non-specific site bound λ -CI, there should be many geometric arrangements of loop formation with another non-specific site bound λ -CI in the genome, whereas there is only one possible arrangement of loop formation with two λ -CI molecules bound to specific sites (ESI^\dagger , Fig. S6). It is hard to estimate the exact number of non-specific loop formation possibilities. Record and co-workers have shown that the lac repressor is capable of forming loops over large intervening distances,⁷⁹ suggesting that the number of looping possibilities could be large. Apart from this entropic factor, other factors may also favour binding to non-target sites. For example, quasi-specific sites distributed throughout the genome may also act as stronger attractants than non-specific sites.⁷⁷ Hence, it is entirely possible that even with binding to multi-partite target sites, some additional specificity enhancing mechanisms may be necessary to generate high enough occupation probability of the target site.

λ -CI is a well-studied prokaryotic transcription factor, which binds to several target sites or operator sites in the phage λ genome, and protein–protein interactions between λ -CI dimers play crucial roles in the life cycle of the phage. Ackers and co-workers obtained accurate binding free energies towards specific and non-specific target sites.^{64,80} The ratio of the specific to the non-specific association constants appears to be around 10^5 (a $\Delta\Delta G$ value of ~ 7 kcal mol⁻¹) for O_{R1} , which may be insufficient alone for saturation of the target site in the context of the whole *E. coli* genome.⁷⁷ However, in the lysogenic state, a cooperative complex forms between adjacent O_{R1} and O_{R2} bound λ -CI, coupling the two binding events. In the absence of any other factor, this should lead to an increase in the $\Delta\Delta G$ value and discrimination ability (see Annexure I, ESI^\dagger).

In a previous article,⁵² we have shown that in contrast to two O_{R1} bound λ -CI dimers, a stronger protein–protein interaction occurs when two dimers bound to O_{R1} and O_{R2} interact in *trans* (bound to two separate pieces of DNA).⁵² This additional binding energy originating from the interaction of the transcription factor with the target site is also operative when the two transcription factors bind in *cis* (one the same piece of DNA; as in the natural context), but only for the correct combination of target sequences. As a consequence, this DNA sequence induced allostery modulated additional binding energy would impart additional stabilization on the interacting molecular complex formed only on target sites. This results in significantly enhanced occupancy of the target site (Fig. 8).

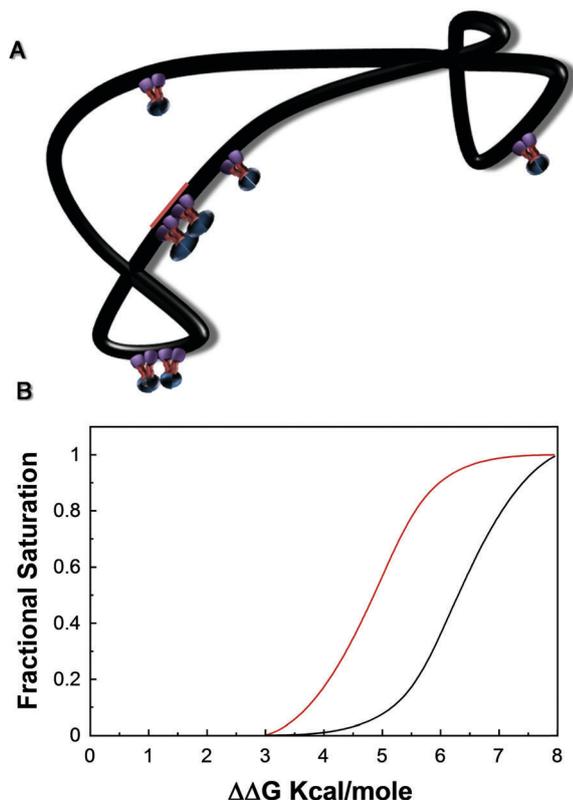


Fig. 8 (A) A cartoon diagram of different possible states of a transcription factor bound to the genome. Dimers are bound to non-specific sites, tetramers can either bind to non-specific sites without any change in the protein–protein interaction interface or to target sites (red line) with a change in the protein–protein interface. (B) Simulation of probability of occupancy of the target site by a transcription factor (y-axis) as a function of free energy difference of specific and non-specific binding (x-axis; $\Delta\Delta G$ (kcal mol⁻¹)). The red line represents a genome size of approximately 4×10^6 (the same as the number of non-specific sites) basepairs and a transcription factor concentration of 100 per cell with a cooperative interaction free energy of 3 kcal mol⁻¹. The black line represents the same genome size with a transcription factor concentration of 100 molecules per cell without any cooperativity. The simulation was generated by the freeware WINPLOT.

Gene regulation in eukaryotes involves the participation of many transcription factors. This is highlighted by the recent discovery of super-enhancers. Individually, eukaryotic transcription factors apparently behave very differently with respect to sequence discrimination than prokaryotic transcription factors in regulating gene expression. Eukaryotic genomes are much larger; thus they ultimately require much higher discrimination capability. It is not well understood how this high degree of specificity is achieved. Mirny and co-workers pointed out that eukaryotic transcription factors in general possess much lower specificity than their prokaryotic counterparts in isolation, even though they operate in the context of a much larger genome.⁸¹ The upstream regulatory regions of many, if not all, genes in eukaryotes have multiple transcription factor binding sites. It has been pointed out that in general, in eukaryotes, the binding of a single transcription factor at a regulatory site does not correlate well with the regulation of expression of that gene.^{82,83} Binding of

multiple transcription factors in a single gene regulatory region may be mandatory to switch the gene regulation on. The binding of multiple transcription factors results in protein–protein interactions—either directly or mediated by another adaptor protein or DNA—between the bound transcription factors, coupling multiple binding events. This should result in enhancement of discriminatory ability by energetic coupling of multiple binding events as implied above. This would also explain why individual transcription factor binding is not well correlated with the regulation of gene expression, as only simultaneous binding of multiple factors switches on gene regulation.

What is the evidence for this model? A number of DNA bound ternary complexes with two transcription factors have been described in which the two bound transcription factors participate in a protein–protein interaction. These include Ets-1/SRF, Sox2-Oct1 and ATF-2/IRF-3 complexes.^{84–86} In a very recent paper, Xie and co-workers demonstrated that even without protein–protein contact, DNA mediated allostery can provide cooperative binding, that is, provide additional binding energy to the ternary complex.⁸⁷ Thus, it is clear that simultaneous binding of several transcription factors to a gene regulatory region can be cooperatively coupled creating a very high discriminatory ability. DNA sequence-specific conformation has also been demonstrated for eukaryotic transcription factors.⁷⁴

Conclusions

This work suggests that the allosteric effect created by a specific DNA sequence may be transmitted by a change in protein dynamics, rather than a change in conformation. The classical model of allostery involves ligand induced conformational change with little understanding of the role of protein dynamics. Kalodimos and co-workers have shown that the well-known allosteric effect of cAMP on DNA binding of CRP is mediated by the change in protein dynamics.^{9,54} How protein–protein interaction energy may be modulated by altered dynamics at the interface is not well understood. Engel and coworkers proposed that hot spot residues at the protein–protein interaction interface must remain relatively static, whereas some residues surrounding them may possess high dynamic character.^{88,89} Thus, allosteric alteration of dynamic character of the interface may produce a desired dynamical fingerprint in the presence of a proper DNA sequence.

We thus conclude that the specificity of transcription factor binding and gene regulatory complex formation at its target site is dictated by specific protein–DNA interaction energy, coupling of multiple protein–protein interactions and allosteric modulation of protein–protein interactions by the DNA sequence. The latter effect may be mediated by induced conformational changes and altered dynamics.

Acknowledgements

We wish to thank Prof. Gautam Basu for his very useful suggestions.

Notes and references

- 1 H. E. Arda and A. J. Walhout, *Briefings Funct. Genomics*, 2009, **elp049**.
- 2 Y. Pan, C.-J. Tsai, B. Ma and R. Nussinov, *Trends Genet.*, 2010, **26**, 75–83.
- 3 F. Merino, B. Bouvier and V. Cojocar, *PLoS Comput. Biol.*, 2015, **11**, e1004287.
- 4 G. K. Ackers, M. L. Doyle, D. Myers and M. A. Daugherty, *Science*, 1992, **255**, 54–63.
- 5 J. P. Changeux, J. C. Gerhart and H. K. Schachman, *Biochemistry*, 1968, **7**, 531–538.
- 6 J.-P. Changeux and S. J. Edelman, *Science*, 2005, **308**, 1424–1428.
- 7 A. Cooper and D. Dryden, *Eur. Biophys. J.*, 1984, **11**, 103–109.
- 8 C.-J. Tsai, A. Del Sol and R. Nussinov, *J. Mol. Biol.*, 2008, **378**, 1–11.
- 9 N. Popovych, S. Sun, R. H. Ebright and C. G. Kalodimos, *Nat. Struct. Mol. Biol.*, 2006, **13**, 831–838.
- 10 H. N. Motlagh, J. O. Wrabl, J. Li and V. J. Hilser, *Nature*, 2014, **508**, 331–339.
- 11 S.-R. Tzeng and C. G. Kalodimos, *Curr. Opin. Struct. Biol.*, 2011, **21**, 62–67.
- 12 Z. H. Foda, Y. Shan, E. T. Kim, D. E. Shaw and M. A. Seeliger, *Nat. Commun.*, 2015, **6**, DOI: 10.1038/ncomms6939.
- 13 A. Mazumder, S. Bandyopadhyay, A. Dhar, D. E. Lewis, S. Deb, S. Dey, P. Chakrabarti and S. Roy, *Biophys. J.*, 2012, **102**, 1580–1589.
- 14 U. Banik, R. Saha, N. C. Mandal, B. Bhattacharyya and S. Roy, *Eur. J. Biochem.*, 1992, **206**, 15–21.
- 15 T. Bhattacharyya and S. Roy, *Biochemistry*, 1993, **32**, 9268–9273.
- 16 A. Ortega, D. Amorós and J. G. de La Torre, *Biophys. J.*, 2011, **101**, 892–898.
- 17 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
- 18 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
- 19 D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman, *Comput. Phys. Commun.*, 1995, **91**, 1–41.
- 20 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.
- 21 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
- 22 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712–725.
- 23 S. A. Showalter and R. Brüschweiler, *J. Chem. Theory Comput.*, 2007, **3**, 961–975.
- 24 D. W. Li and R. Brüschweiler, *Angew. Chem., Int. Ed.*, 2010, **122**, 6930–6932.
- 25 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1950–1958.
- 26 D. Petrov and B. Zagrovic, *PLoS Comput. Biol.*, 2014, **10**(5), e1003638.
- 27 J. Zuegg and J. E. Gready, *Glycobiology*, 2000, **10**, 959–974.
- 28 D. J. Price and C. L. Brooks, *J. Comput. Chem.*, 2002, **23**, 1045–1057.
- 29 S. Halder and D. Bhattacharyya, *J. Phys. Chem. B*, 2010, **114**, 14028–14040.
- 30 C. Dong, L. Yong-Zhi, W. Zhi-Chao and L. Bo, *J. Mol. Model.*, 2014, **20**, 1–10.
- 31 M. Mondal, S. Mukherjee and D. Bhattacharyya, *J. Mol. Model.*, 2014, **20**, 1–11.
- 32 D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossvary, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz Jr., F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. L. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman, *AMBER 14*, Reference Manual: 29–31, 2014.
- 33 L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan and K. Schulten, *J. Comput. Phys.*, 1999, **151**, 283–312.
- 34 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
- 35 I. Andricioaei and M. Karplus, *J. Chem. Phys.*, 2001, **115**, 6289–6292.
- 36 N. Trbovic, J.-H. Cho, R. Abel, R. A. Friesner, M. Rance and A. G. Palmer III, *J. Am. Chem. Soc.*, 2008, **131**, 615–622.
- 37 A. Das, J. Chakrabarti and M. Ghosh, *Biophys. J.*, 2013, **104**, 1274–1284.
- 38 B. J. Killian, J. Y. Kravitz and M. K. Gilson, *J. Chem. Phys.*, 2007, **127**, 024107.
- 39 B. M. King and B. Tidor, *Bioinformatics*, 2009, **25**, 1165–1172.
- 40 R. Mendez and U. Bastolla, *Phys. Rev. Lett.*, 2010, **104**, 228103.
- 41 D.-W. Li, D. Meng and R. Brüschweiler, *J. Am. Chem. Soc.*, 2009, **131**, 14610–14611.
- 42 D.-W. Li and R. Brüschweiler, *Phys. Rev. Lett.*, 2009, **102**, 118108.
- 43 D.-W. Li, S. A. Showalter and R. Brüschweiler, *J. Phys. Chem. B*, 2010, **114**, 16036–16044.
- 44 K. H. DuBay and P. L. Geissler, *J. Mol. Biol.*, 2009, **391**, 484–497.
- 45 K. Kinoshita Jr, A. Ikegami and S. Kawato, *Biophys. J.*, 1982, **37**, 461.
- 46 G. Lipari and A. Szabo, *J. Am. Chem. Soc.*, 1982, **104**, 4546–4559.
- 47 K. Kinoshita, S. Kawato and A. Ikegami, *Biophys. J.*, 1977, **20**, 289–305.
- 48 R. Saha, U. Banik, S. Bandyopadhyay, N. C. Mandal, B. Bhattacharyya and S. Roy, *J. Biol. Chem.*, 1992, **267**, 5862–5867.
- 49 M. Ptashne, *A genetic switch: phage lambda revisited*, Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY, 2004.
- 50 S. Stayrook, P. Jaru-Ampornpan, J. Ni, A. Hochschild and M. Lewis, *Nature*, 2008, **452**, 1022–1025.
- 51 C. E. Bell and M. Lewis, *J. Mol. Biol.*, 2001, **314**, 1127–1136.
- 52 T. Mondol, S. Batabyal, A. Mazumder, S. Roy and S. K. Pal, *FEBS Lett.*, 2012, **586**, 258–262.

- 53 C. Wilson, H. Zhan, L. Swint-Kruse and K. Matthews, *Cell. Mol. Life Sci.*, 2007, **64**, 3–16.
- 54 S.-R. Tzeng and C. G. Kalodimos, *Nature*, 2009, **462**, 368–372.
- 55 F. W. Whipple, E. F. Hou and A. Hochschild, *Genes Dev.*, 1998, **12**, 2791–2802.
- 56 C. E. Bell, P. Frescura, A. Hochschild and M. Lewis, *Cell*, 2000, **101**, 801–811.
- 57 U. Banik, R. Saha, N. C. Mandal, B. Bhattacharyya and S. Roy, *Eur. J. Biochem.*, 1992, **206**, 15–21.
- 58 F. G. Prendergast, M. Meyer, G. L. Carlson, S. Iida and J. D. Potter, *J. Biol. Chem.*, 1983, **258**, 7541–7544.
- 59 R. S. Spolar and M. T. Record Jr, *Science*, 1994, **263**, 777–784.
- 60 D. F. Seneor, T. M. Laue, J. A. Ross, E. Waxman, S. Eaton and E. Rusinova, *Biochemistry*, 1993, **32**, 6179–6189.
- 61 E. L. Quitevis, A. H. Marcus and M. D. Fayer, *J. Phys. Chem.*, 1993, **97**, 5762–5769.
- 62 N. C. Maiti, M. Krishna, P. Britto and N. Periasamy, *J. Phys. Chem. B*, 1997, **101**, 11051–11060.
- 63 J. G. de la Torre, M. L. Huertas and B. Carrasco, *Biophys. J.*, 2000, **78**, 719–730.
- 64 D. F. Seneor, M. Brenowitz, M. A. Shea and G. K. Ackers, *Biochemistry*, 1986, **25**, 7344–7354.
- 65 M. Ptashne, *A genetic switch: Gene control and Phage lambda*, Blackwell Scientific Publications, Palo Alto, CA, USA, 1986.
- 66 S. Bandyopadhyay, S. Deb, S. Bose and S. Roy, *Protein Eng.*, 2002, **15**, 393–401.
- 67 M. Eftink and C. Ghiron, *Proc. Natl. Acad. Sci. U. S. A.*, 1975, **72**, 3290–3294.
- 68 M. Eftink and C. Ghiron, *Biochemistry*, 1977, **16**, 5546–5551.
- 69 S. Bandyopadhyay, U. Banik, B. Bhattacharyya, N. C. Mandal and S. Roy, *Biochemistry*, 1995, **34**, 5090–5097.
- 70 V. A. Feher, J. D. Durrant, A. T. Van Wart and R. E. Amaro, *Curr. Opin. Struct. Biol.*, 2014, **25**, 98–103.
- 71 A. Ghosh, R. Sakaguchi, C. Liu, S. Vishveshwara and Y.-M. Hou, *J. Biol. Chem.*, 2011, **286**, 37721–37731.
- 72 S. Vishwanath, A. Sukhwal, R. Sowdhamini and N. Srinivasan, *Curr. Opin. Struct. Biol.*, 2017, **44**, 77–86.
- 73 S. Tan and T. J. Richmond, *Cell*, 1990, **62**, 367–377.
- 74 S. H. Meijssing, M. A. Pufall, A. Y. So, D. L. Bates, L. Chen and K. R. Yamamoto, *Science*, 2009, **324**, 407–410.
- 75 O. G. Berg and P. H. von Hippel, *J. Mol. Biol.*, 1987, **193**, 723–743.
- 76 L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev and R. Phillips, *Curr. Opin. Genet. Dev.*, 2005, **15**, 116–124.
- 77 J. Chakrabarti, N. Chandra, P. Raha and S. Roy, *Biophys. J.*, 2011, **101**, 1123–1129.
- 78 S. Adhya, *Annu. Rev. Genet.*, 1989, **23**, 227–250.
- 79 S. M. Law, G. R. Bellomy, P. J. Schlx and M. T. Record, *J. Mol. Biol.*, 1993, **230**, 161–173.
- 80 T. R. Pray, D. S. Burz and G. K. Ackers, *J. Mol. Biol.*, 1998, **282**, 947–958.
- 81 Z. Wunderlich and L. A. Mirny, *Trends Genet.*, 2009, **25**, 434–440.
- 82 X. Zhang, D. T. Odom, S.-H. Koo, M. D. Conkright, G. Canettieri, J. Best, H. Chen, R. Jenner, E. Herbolsheimer and E. Jacobsen, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 4459–4464.
- 83 X.-Y. Li, S. MacArthur, R. Bourgon, D. Nix, D. A. Pollard, V. N. Iyer, A. Hechmer, L. Simirenko, M. Stapleton and C. L. Luengo Hendriks, *PLoS Biol.*, 2008, **6**, e27.
- 84 Y. Mo, W. Ho, K. Johnston and R. Marmorstein, *J. Mol. Biol.*, 2001, **314**, 495–506.
- 85 A. Reményi, K. Lins, L. J. Nissen, R. Reinbold, H. R. Schöler and M. Wilmanns, *Genes Dev.*, 2003, **17**, 2048–2059.
- 86 D. Panne, T. Maniatis and S. C. Harrison, *EMBO J.*, 2004, **23**, 4384–4393.
- 87 S. Kim, E. Broströmer, D. Xing, J. Jin, S. Chong, H. Ge, S. Wang, C. Gu, L. Yang and Y. Q. Gao, *Science*, 2013, **339**, 816–819.
- 88 Y. Y. Kuttner and S. Engel, *J. Mol. Biol.*, 2012, **415**, 419–428.
- 89 Y. Y. Kuttner, T. Nagar and S. Engel, *PLoS Comput. Biol.*, 2013, **9**, e1003028.