Project Acronym: MuG

Project title: Multi-Scale Complex Genomics (MuG)

Call: H2020-EINFRA-2015-1

Topic: EINFRA-9-2015

Project Number: 676556

Project Coordinator: Institute for Research in Biomedicine (IRB Barcelona)

Project start date: 1/11/2015

Duration: 36 months

Milestone 26: Comparison of different methodologies to add protein-induced perturbations into predictions of DNA sequence-flexibility relationship

Lead beneficiary: University of Nottingham

Dissemination level: PUBLIC

Due date: 01/11/2017

Actual submission date: 26/04/2018

Table of contents

Introduction

The deformability, spatial organisation, and sequence-dependent structure of DNA are important modulators of key biological processes such as drug-DNA recognition, protein-DNA interactions, and promoter identification [1]. At the same time, the organisation of long genomes in the cell requires special facilitating mechanisms. Architectural proteins such as the bacterial heat-unstable (HU) protein play an important role in promoting DNA spatial organisation and biological processing of bacterial DNA [2]. HU essentially steers the organisation and regulation of the Escherichia coli (E. coli) genome [3,4]. More specifically, HU helps to condense DNA by introducing a sharp bend along its pathway which promotes the closure of short linear molecules into circular arrangements [4], with loss of the protein decidedly disabling DNA looping in vivo [5].

The contribution of HU to biological function is strongly connected to protein-induced DNA looping, and one of those loop-inducing entities is the deformable Lac repressor (LacR) protein assembly, as can be seen in Figure 1 [6]. The LacR protein represses the expression of the *E. coli* lactose (*lac*) operon by simultaneously binding to two spatially separated operator sites in the vicinity of the nucleotides where transcription starts, thus looping the intervening DNA [7]. More specifically, the primary site of LacR attachment on DNA is called O1 and lies 401 and 92 bp away from the two secondary sites, called O2 and O3 respectively [8]. The efficacy of repression of the *E. coli lac* genes is significantly enhanced with the presence of HU, implying that both proteins promote DNA looping. The LacR tetramer can be found in two different conformations: either a rigid V-shaped structure, or an extended conformation where the two dimer LacR subunits are connected by a flexible hinge [8]. Although the genetic and biochemical aspects of expression in the *lac* system have been studied extensively [9], there is little detail available on the exact configuration of the LacR–DNA loop assembly.
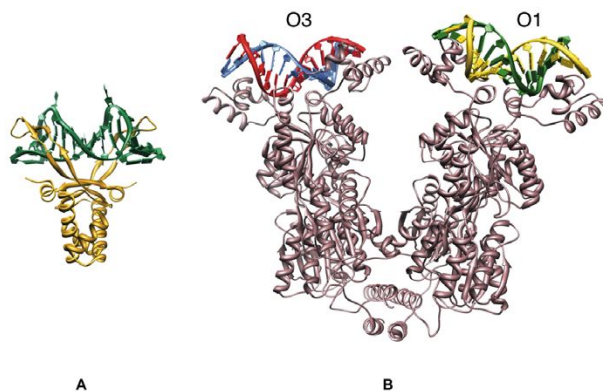

Figure 1. (A) The non-specific architectural protein HU (yellow)introduces a sharp bend in the DNA loop (green) upon binding. (B) The LacR assembly attaches on O1 (primary site) and O3 (secondary site). Modified from [6].

HU-driven modulation of DNA organisation

Extensive research [**refs**] has been conducted on the effect of proteins on DNA flexibility and on the interplay of protein and DNA structure. A molecular simulation study on the *lac* operon [6] has investigated two likely structural contributions towards the *in vivo* looping of bacterial DNA, namely (i) the HU-induced double helix distortions, and (ii) the large-scale LacR deformations. This study [6] utilises a computationally efficient base-pair-level Monte-Carlo (MC) treatment [10] of DNA cyclisation of the likelihood of LacR-mediated loop formation, and incorporates the three-dimensional structural effects of both HU and LacR on DNA. Representative configurations of DNA chains were obtained by direct MC enumeration using a standard Gaussian random-number generator [11] and a modification of the Alexandrowicz half-chain pairwise-combination technique [12], combining only the pairs of half-segments that satisfy the desired end-to-end spacing of the full-length molecule [10]. Proteins are treated implicitly and the DNA model used is the rigid base-pair (RBP) model. The results confirm that the opening of LacR and the presence of HU, at levels approximating those found *in vivo*, promote DNA loop formation. The looping and cyclisation probability is measured by computing the Jacobson-stockmeyer *J* factor [13], that in turn reveals a complex chain-length dependence. In the case of the HU-bound loops, the *J* factor oscillations with *N* chain length are dampened so that LacR-mediated loops are easily formed at most chain lengths, whereas in the absence of HU the chances of the chain ends coming into appropriate contact increase with *N* between 75 and 150 bp, as can be seen in Figure 2. The HU-induced modulation of the chain-length dependence of LacR mediated looping could not have been predicted if an ideal elastic model of DNA had been utilised. In the same work, *N* is also shown to determine HU uptake.

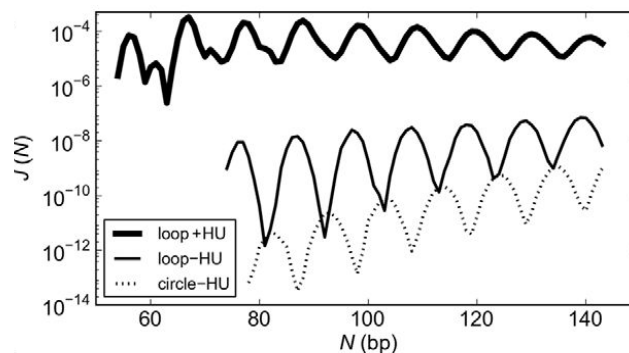Figure 2. The computed *J* factors of DNA fragments of different *N* chain length, in the presence or absence of HU, measure the looping and cyclisation probability between the headpieces of the LacR assembly. Modified from [6].

When bound on V-shaped LacR, the DNA loop can adopt four possible orientations, as can be seen in Figure 3, namely an antiparallel (A1, A2) and a parallel (P1, P2) orientation.
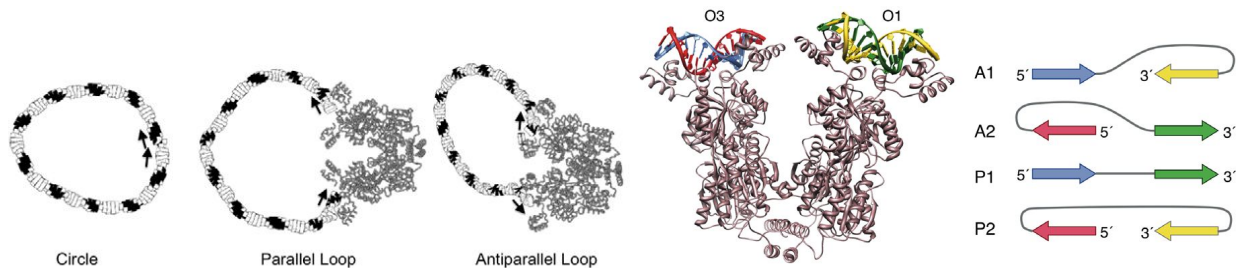
Figure 3. The DNA loop can adopt four possible orientations (A1, A2 - antiparallel, P1, P2 - parallel) when bound with V-shaped LacR. Modified from [6].

HU was also shown to increase the mix of different DNA loops that are formed on V-shaped LacR, as can be seen in a schematic representation of $J(N)$ in the presence (Figure 4A) and absence (Figure 4B) of HU. The wider range of N in the (A) compared to (B) reflects the greater ease of forming short LacR-mediated loops in the presence of HU.
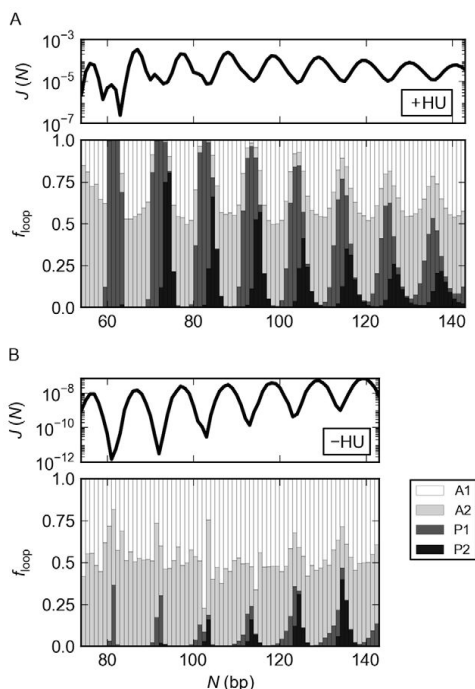

Figure 4. HU increases the mix of DNA loops formed on the V-shaped LacR assembly, which underlies the chain-length-dependent variation in the J factor. (A) More complex plot of $J(N)$ and greater diversity of loops (f_{loop}) in the presence, compared to (B) the absence (B) of HU. From [6].

Another molecular simulation study [4] utilising the same computationally efficient base-pair-level Monte-Carlo (MC) treatment [10] and sampling technique [10] as discussed earlier [6] has investigated the effects of non-specific HU binding on the configurational properties of short fragments of ideal B DNA, particularly focusing on cyclisation propensities are modulated. The DNA base pairs were represented by rectangular slabs and the arrangement of each base pair was described by six independent parameters (angular variables called tilt, roll, twist, and variables with dimensions of distance called shift, slide, and rise). The deformable DNA was subject to standard fluctuations in

bending and twisting about the canonical B-type structure. As can be seen in Figure 5, the random binding of HU transformed the deformed and relatively stiff DNA into a flexible chain molecule. The generated configurations demonstrated a degree of bending similar to what was observed experimentally in atomic force microscopy (AFM) images collected under similar binding conditions [14]. The same study illustrated that the likelihood of end-to-end contact increases with the number of bound HU molecules. It was also found that HU presence approximating levels of those found *in vivo* reduced the persistence length (L_p) by roughly threefold compared with that of naked DNA.
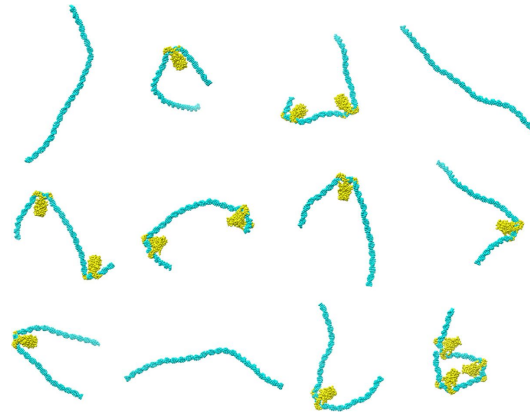

Figure 5. Representative configurations of a 126-bp ideal B-DNA fragment in the presence (five configurations with one HU, three with two HUs, and one with three HUs) or absence (three protein-free configurations) of HU protein. From [4].

The number of HU molecules in successfully closed duplexes was shown to exhibit a periodic dependence on chain length, as can be seen in Figure 6 (right). When comparing the left (*J* vs *N*) and right (number of bound HU molecules vs *N*) plots of Figure 6, it can be observed that there are fewer proteins bound to minicircles of chain lengths corresponding to local peaks in the plot of *J* versus *N*, and more HU molecules associated with minicircles in valleys. DNA chains that are more likely to cyclise (i.e., chains of length corresponding to local maxima in the computed dependence of the *J* factor on *N*) tend to bind two to three HU molecules on average, and DNA chains that are less likely to cyclise tend to bind an average of three or more proteins.
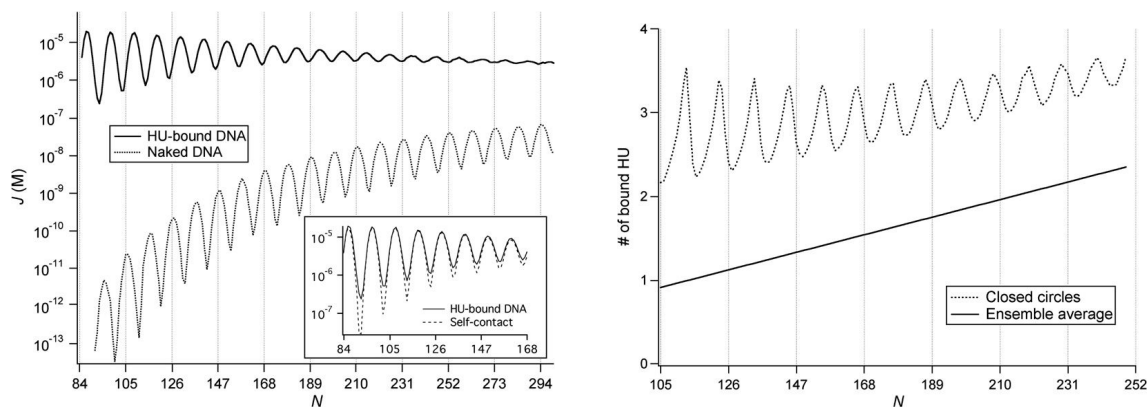
Figure 6. Left: MC estimates of the dependence of J factors on N chain length of naked (dashed line) and HU-bound (continuous line) DNA. Inset: dependence of J on N when configurations with protein–protein, protein–DNA, and DNA–DNA overlaps are excluded in small minicircles. Right: Average number of HU molecules as a function of chain length N. The dashed curve represents successfully cyclised DNA chains, whereas the continuous line is the simulated degree of binding in the ensembles of linear and cyclic HU-bound chains. Modified from [4].

The nonspecificity of HU-binding allows HU to accumulate at different levels in closed chains of specific length N, with Table 1 demonstrating that HU is preferentially found in cyclised over linear DNA structures. Longer N chains can accommodate more HU molecules under the chosen binding conditions, altering the fraction of sampled DNA chains with a particular number (0, 1, 2, 3, etc.) of HU molecules (Table 1). As can be extrapolated from the data in Table 1, doubling the binding fraction of HU from 1 per 100 to 1 per 50 bp (with a HU binding probability of w_{HU}=2.70 x 10^{-2}) decreases the fraction of linear 126-bp chains with 0 or 1 HU. At the same time it increases the fraction of chains with $\geq$ 2 HU's compared with chains with 1 HU per 100 bp. Decreasing the HU binding fraction to 1 per 200 bp (with a HU binding probability of w_{HU}=5.35 x 10^{-3}) increases the fraction of naked linear DNA molecules.

Table 1. Protein populations, as a function of simulated binding level w_{HU} and chain length N, on linear and circular HU-bound DNA chains.

DNA chain	w_{HU}	N (bp)	0 HU	1 HU	2 HU's	$\geq$3 HU's
Linear	5.35×10^{-3}	126	**0.546**	**0.354**	0.089	0.011
		210	**0.347**	**0.395**	**0.194**	0.064
	1.15×10^{-2}	126	**0.277**	**0.412**	**0.236**	0.075
		210	0.105	**0.274**	**0.312**	**0.309**
	2.70×10^{-2}	126	0.049	**0.212**	**0.347**	**0.392**
		210	0.005	0.038	**0.124**	**0.833**
Circular	5.35×10^{-3}	126	0	0.002	**0.897**	**0.101**
		210	0.004	0.041	**0.615**	**0.340**
	1.15×10^{-2}	126	0	0.002	**0.730**	**0.268**
		210	0	0.005	**0.301**	**0.694**
	2.70×10^{-2}	126	0	0	**0.326**	**0.674**
		210	0	0	0.037	**0.963**

HU binding probabilities w_{HU}=5.35×10^{-3}, 1.15×10^{-2}, and 2.70×10^{-2} correspond to one bound protein per 200, 100, and 50 bp, respectively, in the simulated ensemble of open chain configurations (note the larger HU binding fraction in closed molecules). Dominant HU–DNA stoichiometric ratios are highlighted in boldface.

Modelling the Lac repressor - operator assembly

As it has already been discussed in the Introduction of this review, the LacR tetramer can be found in either a rigid V-shaped structure, or an extended conformation [8]. Further, LacR binds to a primary site on DNA, which is called O1, and lies 401 and 92 bp away from the two secondary sites, called O2 and O3 respectively [8]. Interestingly, various studies have demonstrated that the angle between the dimer subunits, in other words, the angle of aperture α (Figure 7, iii), can vary. A recent study [9] has focused on the loops formed between the primary site O1 and the secondary site O3. There are four possible DNA loop types, namely antiparallel (A1, A2) and parallel (P1, P2) as can be seen in Figure 3 (right) and Figure 7 (ii). This work took account of the influence of DNA sequence and protein flexibility on the configuration of the O3-O1 loop.
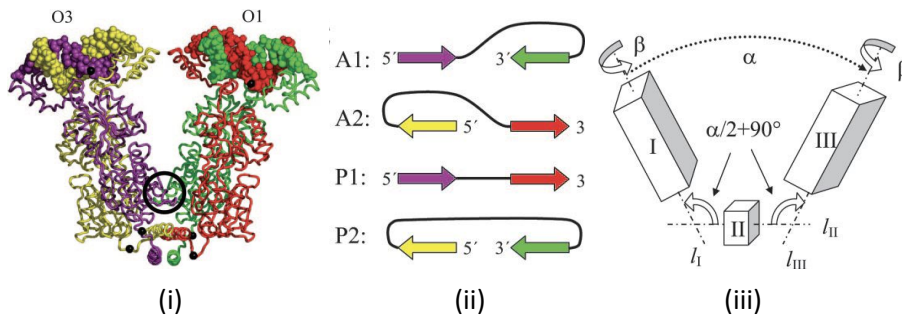
Figure 7. **(i)** structure of the LacR protein. The purple and yellow chains form one dimer unit, and the red and green chains form the second dimer unit. **(ii)** the possible DNA loop types. The colour-coded arrows illustrate the direction of the loops corresponding to the structure seen in (i) in the same figure. **(iii)** schematic representation of the LacR opening and the angle α changing. The transition from V-shape to extended configuration is treated as a motion of three rigid domains (I, II, and III) about two hinges. Modified from [9].

It was shown [9] that a DNA loop in the LacR–DNA complex can be one of the five types shown in Figure 8. For four of these, (A1, A2, P1, and P2) the LacR entity is in its V-shaped conformation, whereas for P1^E the LacR assumes its extended form. A base-pair-level theory of sequence-dependent DNA elasticity was employed. Linear DNA segments of short-to-medium chain length (50–180 bp) tend to form loops with the extended form (P1^E) of LacR, whereas loops formed within negatively supercoiled plasmids give rise to the V-shaped structure observed in the crystal.
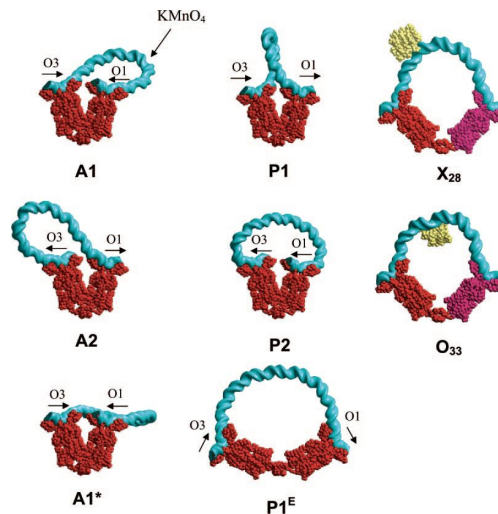
Figure 8. Calculated minimum energy configurations of the wild type O3-O1 loop complexed with the LacR tetramer, for various combinations of orientation, linking number, LacR conformation, and nucleotide sequence. The loop labeled A1* has the same orientation and DNA sequence as that labeled A1 but differs in linking number and therefore has a different minimum energy configuration and free energy. Loops X$_{28}$ and O$_{33}$ are examples of minimum energy configurations of the loop with DNase I bound at favourable and unfavourable sites, respectively. From [9].

Cyclisation of short DNA

The cyclisation properties of short, realistically modelled DNA chains with significantly increased sample sizes, i.e. $O(10^{14})$ configurations, have been investigated [10]. The same work [10] has generated the fast and computationally efficient MC treatment and the modified Alexandrowicz half-chain pairwise-combination technique [12] also used in more recent works [4,6], as those have been discussed earlier in this review. The double helix was treated at the level of base-pair steps, also accounting for sequence-dependent variability, coupling of conformational variables, and anisotropy of bending. The ring-closure properties of an ideal, inextensible, naturally straight DNA molecule, over the range of N chain lengths between 90 and 450 bp was initially determined. In this model, deformations are allowed in tilt, roll, and twist, while the translational parameters shift, slide, and rise are fixed near their intrinsic values. The computed J factors appeared generally unaffected by the selected boundary conditions.

While the dependence of J on N satisfactorily reflected similar findings in literature for short chains (105-130 bp) and for chains of 240 bp or more, it wildly underestimated J for very short chains (89-105 bp), which encouraged the inclusion of more realistic DNA models in the investigation. An idealised, naturally curved DNA model was used, where the 150 bp long molecule adopted a stress-free nearly circular configuration. This time, the computed J factors were many orders of magnitude greater than those of a naturally straight molecule, which agreed remarkably well with values predicted for a naturally curved DNA molecule with the same ratio of curvature and L_p. The curvature was also shown to enhance the ring closure propensities of the molecule.

Multi-protein-DNA assemblies

Aiming to enrich the investigation of DNA loops formed by the binding of LacR [9], a recent study has focused on the structures and free energies of multi-protein-DNA assemblies. More specifically, this work has investigated the effects of the catabolite activator protein (CAP) binding on the configurations of the wild-type 92-bp LacR-mediated O3-O1 DNA loop. CAP is a protein which, like LacR, also bends DNA. Further, CAP paradoxically facilitates the binding of RNA polymerase. Studies on CAP binding have shown that while simultaneous binding of CAP and LacR to the O3 site is possible, the sequence covered by LacR is shifted approximately 5-6bp upstream [15]. The modelling study discussed in this section revealed an additional potential LacR binding site 5-bp upstream of O3, bearing structural similarities with O1 and termed O3*. The representative configurations of the LacR-CAP-DNA loop topoisomers, with LacR bound at O3* were generated and can be seen in Figure 9. The structure with the lowest free energy was A2[f] and was therefore the proposed configuration of the LacR-CAP-DNA loop. In A2[f] the LacR entity binds to the alternative O3* site. Despite the fact that O3* has not been confirmed experimentally yet, it is likely that if LacR binds to the sequence, it does so only if CAP is present.
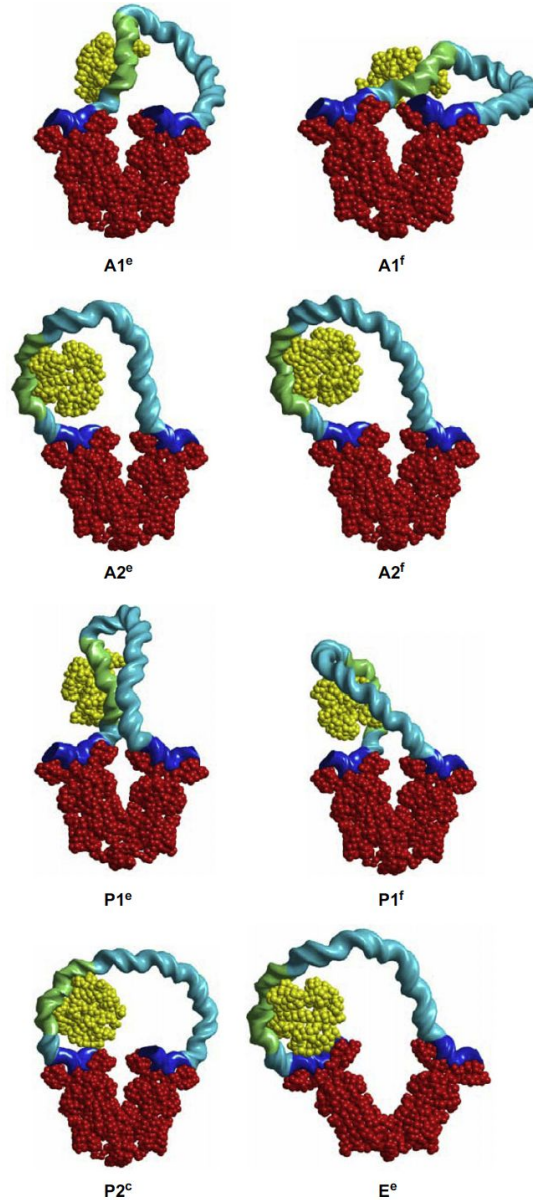
Figure 9. The proposed configuration of the LacR–CAP–DNA loop. In this structure, the LacR entity binds to the alternative O3* site. This is in agreement with experimental evidence [15] demonstrating a 5-bp upstream shift of the LacR binding site after CAP binding.

Protein-induced DNA perturbations: an ongoing scientific question

The 3D organisation of DNA plays a crucial but still not fully understood role in gene regulation [16]. It is therefore clear that the importance of studying the genome from a 3D perspective will only continue to increase [17]. The combined effect of specific and non-specific DNA-binding proteins steering DNA towards loop formation, along with DNA superhelicity, contributes to further DNA compaction in the cell. It is therefore clear that the geometrical concept of DNA loop formation is of significant importance in understanding the structure of the chromosome [7]. The preceding discussion aims to showcase the
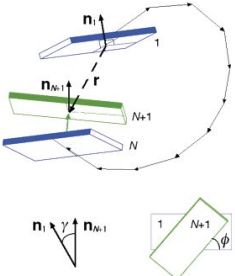
extensive work that has been carried out towards unravelling the intricacies of protein-induced DNA perturbations, cyclisation, and flexibility. The table below (Table 2) offers further detail on the same subject, but this time focusing more on the specific tools and methodologies used to achieve this task. The non-exhaustive list of workflows presented below highlights the plethora of different approaches employed. While the majority of studies presented in Table 2 are mainly computational/theoretical (e.g. [18], [19], [20]), they have significant differences, namely in terms of atomistic detail, DNA model, and protein-body treatment employed. They further differ on the choice of pipelines (e.g. [21]), sampling techniques (e.g. [18]), and analytical tools (e.g. [22], [23]). At the same time, the combination studies employing experimental molecular biology and complementary computational tools (e.g. [24]) introduce another level of variation in this field of study. On the one hand, this makes the successful comparison between different methodologies fairly challenging. However on the other hand, it clearly implies that the study of protein-induced DNA perturbations very much remains an ongoing scientific question.

Different methodologies for studying protein-induced DNA perturbations

A non-exhaustive list of various different workflows used to study and assess protein-induced DNA perturbations is presented in the table below. In most cases, MD or Monte Carlo approaches are employed, with varying levels of atomistic detail. A variety of DNA models are also utilised, such as the rigid base-pair model or the elastic rod. Multiscale approaches are also presented. The ability of multiscale approaches to describe the different parts of the modeled system at different spatial and temporal resolution is ideal, as the sheer size of the simulated systems would make fully atomistic simulations extremely costly computationally.

Table 2. Collated presentation of selected workflows focusing on protein-induced DNA perturbations.

Methods and tools	DNA treatment	Overview	Notes	Ref

Configurational sampling: "Gaussian Sampling" [18] Utilisation of the quadratic form of the DNA elastic energy and usage of a modified standard Gaussian random number generator, to collect a Boltzmann distribution of dimeric states. **Sampling enhancement:** Modified Alexandrowicz [25] technique. The DNA chain is divided into two nearly equal segments and L configurations of each half-chain segment are sampled separately. By taking all pairwise combinations of both halves, L^2 configurations of the full chain are achieved [18] *J* **factor calculation:** To monitor the closure of an N base-pair DNA chain, a virtual $(N+1)_{th}$ base pair of the same type as the 1st base pair is added. The DNA is considered closed when the $(N+1)_{th}$ base pair coincides with the 1st base pair: The *J* factor is expressed as a product of probabilities which describe the contribution of the spatial configuration to the cyclisation equilibrium constant.	Rigid base-pair step parameters, using elastic potentials for the intrinsic structure and elastic moduli of the base-pair steps.	Generation of configurations using a direct Monte-Carlo enumeration approach, and computation of *J* factors.		[18]

3DNA [23]: generated DNA pathway from Monte Carlo-sampled base-pair step parameters. Protein coordinates are obtained by superposition of DNA from the appropriate crystal structure on the simulated pathway. Gaussian Sampling as in [18] Sampling enhancement: modified Alexandrowicz [25] technique, as in [18] Direct Monte-Carlo enumeration as in [18]	Rigid base-pair step parameters, using elastic potentials for the intrinsic structure and elastic moduli.	Dinucleotide step models for the prediction of protein-mediated DNA structure, flexibility, and ring-closure.	The DNA pathway is generated from Monte Carlo-sampled base-pair step parameters using the 3DNA software package [23]. Protein coordinates are obtained by superposition of DNA from the appropriate crystal structure on the simulated pathway. Images are rendered with Chimera.24	[19]
In all computations: - Maple 10 [26] (symbolic algebra software) - Automatic differentiation algorithms [27] The configurations and deformational free energies of multi-protein–DNA assemblies (LacR-CAP-DNA) are determined using a sequence- dependent treatment of DNA [28,29] DNA and proteins represented as elastic bodies Collection of constraints including steric hindrance imposed on DNA configuration Classical mechanics to compute DNA configurations	Base-pair-level theory of sequence-dependent DNA elasticity [28,29]	Computational approach for determining the structures and free energies of multi-protein–DNA assemblies. This method assumes that the DNA deforms in such an assembly, with the protein components providing spatial constraints.		[30]
CURVES [22]: measured the minor groove width of double-stranded DNA in co-crystal structures.		Experimental and computational analysis pipeline, mapping DNA minor groove width and	A region of narrow minor groove width present in naked DNA, is likely to be recognised by a	[35]

DNAshape [31] and ORChID2 [32] (available Perl scripts: http://dna.bu.edu/orchid/): predicted minor groove patterns of naked DNA. RobFinder (https://github.com/rnaplus/RobFinder): Custom MATLAB application; visualises raw electrophoresis data, fits and integrates peaks. ShapeFinder [33] and QuShape [34]: processed capillary electrophoresis data.		comparing naked DNA and DNA-protein complexes.	DNA-binding protein and maintained in the protein–DNA complex.	
Juicer [21]: end-to-end pipeline for converting raw FASTQ reads into Hi-C maps and loop calls Hi-C experiments ChIP-seq analysis		The locations of polycomb (Pc)-mediated chromatin loops in *Drosophila* were determined by Hi-C; Hi-C data were analysed using the Juicer [21] pipeline.	High resolution Hi-C contact maps revealed chromatin loops within PcG domains. The association of loop anchors in *Drosophila* with Pc protein is noteworthy because it points to a role of looping not only in gene activation but in gene repression as well.	[24]
3DNA [23,36] software: computed averages of the local helical axes and generated DNA coordinates from base-pair-step parameters Gaussian sampling as in [18] Direct Monte-Carlo enumeration as in [18] Sampling enhancement: modified Alexandrowicz [25]	Rigid base-pair step parameters; the DNA is treated here as an ideal, inextensible, naturally straight molecule.		The presence of HU at levels approximating those found *in vivo*, enhance the probability of loop formation. HU affects the global organisation of the repressor; the opening of the repressor influences the levels of HU binding to DNA.	[37]

technique, as in [18] Implicit proteins				
Sequence-dependent DNA elasticity theory [28,29] to calculate the configuration and free energy of the DNA loop (as a function of various conditions, including end conditions imposed by the LacR tetramer) Comparisons of computed loop configurations with published experimental data of permanganate sensitivities, DNase I cutting patterns, and gel-mobility patterns.	Base-pair-level theory of sequence-dependent DNA elasticity [38,39]			[40]
Elastic rod model Straight-helical-straight (SHS) approximation Computation of the conformation of the looped DNA-LacI complex follows the methods described in [41] Varying the phasing (location of the intrinsic bend relative to the two operator sites) parameters yields a large family of sequences used to evaluate looping behaviors	Computational elastic rod model [42]; sequence-dependentintrinsic curvature of the helical axis is explicitly incorporated in the rod formulation. SHS approximation for highly bent DNA sequences is used for describing intrinsic curvature with the rod model [41]. The rod model makes a direct connection between base-pair-level DNA structure/flexibility parameters and the results to be expected from looping experiments.	Focus on the looping of intrinsically curved DNA by LacI; method and conclusions relevant to any DNA-protein complexes forming short DNA loops.		[43]
NAMD for the MD simulations [44] Solvation using DOWSER [45]	Elastic rod model [42]	Multiresolution approach to modeling complexes between protein and		[20]

and VMD [46] Electrochemical potential calculated using DelPhi [47] COLNEW: FORTRAN library to solve boundary value problems [48] 75-bp-long DNA loop Electrostatic interactions of DNA in the loop were disregarded Proteins and DNA segments in direct contact with proteins are described at full-atom MD level The DNA loop between the protein-bound DNA segments is described by means of the elasticity theory [49], in a coarse-grained (continuum) representation		looped or coiled DNA.		
NAMD for the MD simulations [44] Solvation using VMD [46] Electrochemical potential calculated using DelPhi [47] 75-bp-long DNA loop The MD simulations included the forces accounting for elastic stress and torque due to the looping of the DNA between the operators. The structure and forces of the DNA loop are recomputed every 10 ps, with the boundary conditions derived from the last MD snapshot of the protein-DNA structure, because LacI shifts the operators and termini of	Elastic rod model [42]	Atomistic multiscale MD simulation of a complex between the *lac* repressor protein (LacI) and a 107-bp-long DNA segment	Multiscale simulation: the DNA loop moves toward lower-energy conformations and the force of the LacI–DNA interaction drops; the relaxation of the loop is due to the mobility of the LacI head groups. Opening LacI: the DNA loop adapted a more open configuration and the elastic energy decreased and became more evenly distributed; a significant degree of the loop relaxation can be achieved solely by rotation of	[51]

the loop. The simulations were performed with additional force terms in order to bring the dynamics of the LacI–DNA complex within the time scale available to MD simulations [50] the DNA loop. Forced-opening simulation of the LacI-DNA complex aiming to mimic the *in vivo* opening of the LacI cleft.			the head groups	

References

1. Dršata T, Lankaš F. Theoretical models of DNA flexibility. Wiley Interdiscip Rev Comput Mol Sci. 2013;3: 355–363.

2. Berger M, Farcas A, Geertz M, Zhelyazkova P, Brix K, Travers A, et al. Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU. EMBO Rep. 2010;11: 59–64.

3. Wei J, Czapla L, Grosner MA, Swigon D, Olson WK. DNA topology confers sequence specificity to nonspecific architectural proteins. Proc Natl Acad Sci U S A. 2014;111: 16742–16747.

4. Czapla L, Swigon D, Olson WK. Effects of the nucleoid protein HU on the structure, flexibility, and ring-closure properties of DNA deduced from Monte Carlo simulations. J Mol Biol. 2008;382: 353–370.

5. Bond LM, Peters JP, Becker NA, Kahn JD, Maher LJ 3rd. Gene repression by minimal lac loops in vivo. Nucleic Acids Res. 2010;38: 8072–8082.

6. Czapla L, Grosner MA, Swigon D, Olson WK. Interplay of protein and DNA structure revealed in simulations of the lac operon. PLoS One. 2013;8: e56548.

7. Semsey S, Virnik K, Adhya S. A gamut of loops: meandering DNA. Trends Biochem Sci. 2005;30: 334–341.

8. Swigon D, Olson WK. Mesoscale modeling of multi-protein–DNA assemblies: The role of the catabolic activator protein in Lac-repressor-mediated looping. Int J Non Linear Mech. 2008;43: 1082–1093.

9. Swigon D, Coleman BD, Olson WK. Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. Proc Natl Acad Sci U S A. 2006;103: 9879–9884.

10. Czapla L, Swigon D, Olson WK. Sequence-Dependent Effects in the Cyclization of Short DNA. J Chem Theory Comput. 2006;2: 685–695.

11. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in C. 2nd ed. Cambridge University Press; 1992. pp. 288–290.

12. Alexandrowicz Z. Monte Carlo of Chains with Excluded Volume: a Way to Evade Sample Attrition. J Chem Phys. 1969;51: 561–565.

13. Jacobson H, Stockmayer WH. Intramolecular Reaction in Polycondensations. I. The Theory of Linear Systems. J Chem Phys. 1950;18: 1600–1606.

14. van Noort J, Verbrugge S, Goosen N, Dekker C, Dame RT. Dual architectural roles of HU: Formation of flexible hinges and rigid filaments. Proceedings of the National Academy of Sciences. 2004;101: 6969–6974.

15. Fried MG, Hudson JM. DNA looping and lac repressor-CAP interaction. Science. 1996;274: 1930–1; author reply 1931–2.

16. Travers AA, Muskhelishvili G, Thompson JMT. DNA information: from digital code to analogue structure. Philos Trans A Math Phys Eng Sci. 2012;370: 2960–2986.

17. Gorkin DU, Leung D, Ren B. The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell. 2014;14: 762–775.

18. Czapla L, Swigon D, Olson WK. Sequence-Dependent Effects in the Cyclization of Short DNA. J Chem Theory Comput. 2006;2: 685–695.

19. Czapla L, Swigon D, Olson WK. Effects of the nucleoid protein HU on the structure, flexibility, and ring-closure properties of DNA deduced from Monte Carlo simulations. J Mol Biol. 2008;382: 353–370.

20. Villa E, Balaeff A, Mahadevan L, Schulten K. Multiscale Method for Simulating Protein-DNA Complexes. Multiscale Model Simul. 2004;2: 527–553.

21. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 2016;3: 95–98.

22. Lavery R, Sklenar H. Defining the structure of irregular nucleic acids: conventions and principles. J Biomol Struct Dyn. 1989;6: 655–667.

23. Lu X-J, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. 2003;31: 5108–5121.

24. Eagen KP, Aiden EL, Kornberg RD. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. Proc Natl Acad Sci U S A. 2017;114: 8764–8769.

25. Alexandrowicz Z. Monte Carlo of Chains with Excluded Volume: a Way to Evade Sample Attrition. J Chem Phys. 1969;51: 561–565.

26. Maple 10. Maplesoft, a division of Waterloo Maple Inc., Waterloo, Ontario.

27. Griewank G, Corliss A. Automatic Differentiation of Algorithms: Theory, Implementation, and Application. Philadelphia: Society for Industrial & Applied Mathematics; 1991.

28. Coleman BD, Olson WK, Swigon D. Theory of sequence-dependent DNA elasticity. J Chem Phys. 2003;118: 7127–7140.

29. Olson WK, Swigon D, Coleman BD. Implications of the dependence of the elastic properties of DNA on nucleotide sequence. Philos Trans A Math Phys Eng Sci. 2004;362: 1403–1422.

30. Swigon D, Olson WK. Mesoscale modeling of multi-protein–DNA assemblies: The role of the catabolic activator protein in Lac-repressor-mediated looping. Int J Non Linear Mech. 2008;43: 1082–1093.

31. Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, et al. DNAshape: a method for the high-throughput prediction of DNA structural features on a genomic scale. Nucleic Acids Res. 2013;41: W56–62.

32. Greenbaum JA, Pang B, Tullius TD. Construction of a genome-scale structural map at

single-nucleotide resolution. Genome Res. 2007;17: 947–953.

33. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA. 2008;14: 1979–1990.

34. Karabiber F, McGinnis JL, Favorov OV, Weeks KM. QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. RNA. 2013;19: 63–73.

35. Azad RN, Zafiropoulos D, Ober D, Jiang Y, Chiu T-P, Sagendorf JM, et al. Experimental maps of DNA structure at nucleotide resolution distinguish intrinsic from protein-induced DNA deformations. Nucleic Acids Res. 2018; doi:10.1093/nar/gky033

36. Lu X-J, Olson WK. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nat Protoc. 2008;3: 1213–1227.

37. Czapla L, Grosner MA, Swigon D, Olson WK. Interplay of protein and DNA structure revealed in simulations of the lac operon. PLoS One. 2013;8: e56548.

38. Coleman BD, Olson WK, Swigon D. Theory of sequence-dependent DNA elasticity. J Chem Phys. 2003;118: 7127–7140.

39. Olson WK, Swigon D, Coleman BD. Implications of the dependence of the elastic properties of DNA on nucleotide sequence. Philos Trans A Math Phys Eng Sci. 2004;362: 1403–1422.

40. Swigon D, Coleman BD, Olson WK. Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. Proc Natl Acad Sci U S A. 2006;103: 9879–9884.

41. Goyal S, Lillian T, Blumberg S, Meiners J-C, Meyhöfer E, Perkins NC. Intrinsic curvature of DNA influences LacR-mediated looping. Biophys J. 2007;93: 4342–4359.

42. Balaeff A, Mahadevan L, Schulten K. Elastic Rod Model of a DNA Loop in theLacOperon. Phys Rev Lett. 1999;83: 4900–4903.

43. Lillian TD, Goyal S, Kahn JD, Meyhöfer E, Perkins NC. Computational analysis of looping of a large family of highly bent DNA by LacI. Biophys J. 2008;95: 5832–5842.

44. Kalé L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, et al. NAMD2: Greater Scalability for Parallel Molecular Dynamics. J Comput Phys. 1999;151: 283–312.

45. Hermans J, Mann G, Wang L, Zhang L. Simulation Studies of Protein-Ligand Interactions. Lecture Notes in Computational Science and Engineering. 1999. pp. 129–148.

46. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996;14: 33–38.

47. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science. 1995;268: 1144–1149.

48. Bader G, Ascher U. A New Basis Implementation for a Mixed Order Boundary Value ODE Solver. SIAM Journal on Scientific and Statistical Computing. 1987;8: 483–500.

49. Olson WK. Simulating DNA at low resolution. Curr Opin Struct Biol. 1996;6: 242–256.

50. Isralewitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. Curr Opin Struct Biol. 2001;11: 224–230.

51. Villa E, Balaeff A, Schulten K. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. Proc Natl Acad Sci U S A. 2005;102: 6783–6788.