The Effect of the Nucleoid Protein HU on the Structure, Flexibility, and Ring-closure Properties of DNA

THE STATE UNIVERSITY OF NEW JERSEY RUTGERS

Luke Czapla[a], David Swigon[b], and Wilma K. Olson[a]

[a]Department of Chemistry and Chemical Biology, Rutgers University, and [b]Department of Mathematics, University of Pittsburgh

Abstract

HU is a non-sequence-specific DNA binding protein that distorts double-helical structure and induces a large (~110-140 degrees) bend in the bound DNA region. We extend our previous DNA Monte Carlo methods[1] to consider the presence of *HU* and its binding based on two parameters: the probability *Pr* of *HU* binding at any position along the DNA - based on the average number of bound *HU* proteins - and the known structure of *HU*-bound DNA, from one of the four currently known crystal structures found in the Protein Data Bank (PDB).

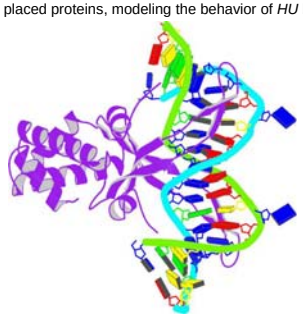
Applying our novel Monte Carlo methods, we have found that the ring-closure probability, measured as the *J* factor, is greatly enhanced in comparison to naked DNA, and is essentially constant over a wide range of short chain lengths (90-300 bp), indicating a role of *HU* in stabilizing DNA loops *in-vivo*. Moreover, the preference of *HU* in binding negatively supercoiled DNA is consistent with the observed negatively supercoiled circles captured in simulation of DNA with *HU*.

The features of *HU*-bound DNA (bending and untwisting the double helix) are common among other non-specific binding proteins, such as the *HMGB* family of eukaryotic proteins, known to substitute for *HU* in looping assays and in condensing the bacterial chromosome into the negatively supercoiled nucleoid and restoring normal morphology to *E. coli* cells.

Introduction

HU ("heat unstable") protein is an *E. coli* DNA-binding protein with about equal affinity for any DNA sequence (non-sequence-specific). Similar to the High Mobility Group (ex. *HMGB*) proteins found in eukaryotes, this protein strongly bends the DNA, as well as plays other regulatory roles in the cell beyond the scope of DNA looping. There are about ~30,000 *HU* dimers present in *E. coli* during its exponential growth phase, corresponding to 1 *HU* per 150 bp of genomic DNA.

Our Monte Carlo method can generate a Boltzmann ensemble of ~10^{17} configurations using a novel selective half-chain method and a Gaussian sampling technique[1]. By systematically placing the base-pair step parameters for DNA bound to *HU* (deduced from crystallographic data) at all base-pair positions with probability *Pr*, we can extend the method to investigate the properties of DNA with randomly placed proteins, modeling the behavior of *HU* binding.

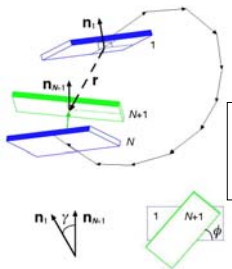


HU-DNA structure with NDB ID PD0430 (PDB ID 1P71), one of four experimentally deduced X-ray structures of *HU* complexed with DNA. The protein is shown in purple and contacts a ~15 base-pair region of the DNA. Regions of flipped out thymine bases are approximated by virtual steps in our base-pair step model. (Illustration courtesy of NDB server)

Methods

Cyclization of DNA in the presence of HU

Calculating the *J* factor from terminal base-pairs:

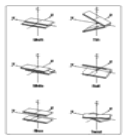


r = **0**

cos γ=1

ϕ=0

In testing for ring closure, an extra joining step *N*+1 is added to the sequence. Configurations within chosen bounds of *r*=0, cos(γ)=1, ϕ=0 are used to approximate the *J* factor, which is equal to the value -exp($\Delta G/kT$), and hence is well-suited to approximate the free energy of looping.



For modeling DNA, we apply a simple elastic, inextensible base-pair step model:

$$E_{step} = \frac{1}{2}A(\Theta_{bend}^2 + \Theta_{twist}^2) + \frac{1}{2}C(\Theta_{twist} - 34.3°)^2$$

Proteins are then added to the chains:

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

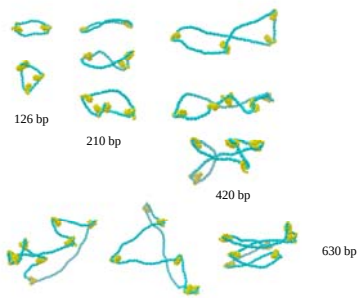
Representative 126 bp chains with an average of 1 HU/100 bp: some have no bound HU, most have one bound, and some have more than one bound (2, 3 bound).

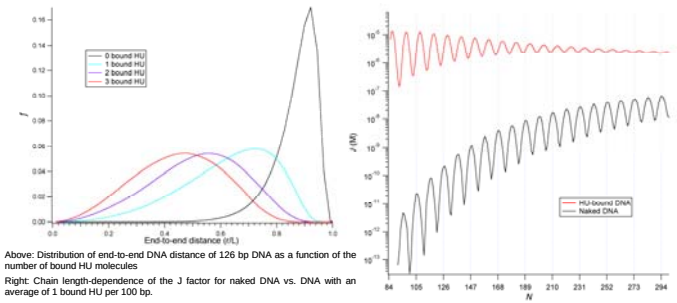
Capturing closed molecules:

Representative closed chains from simulation with an average of 1 HU/100 bp: Columns going down, 126, 210, and 420 bp. Bottom row, 630 bp.

As the chain length increases, the degree of supercoiling and condensation increases. The molecules are characterized by their excess link, total twist, and writhe. For the lengths shown, the intrinsic link Lk_0 is an integer.

$$\Delta Lk = Tw + Wr - Lk_0$$





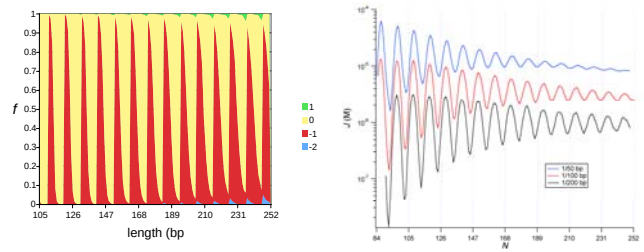
Above: Distribution of end-to-end DNA distance of 126 bp DNA as a function of the number of bound HU molecules

Right: Chain length-dependence of the J factor for naked DNA vs. DNA with an average of 1 bound HU per 100 bp.

Results

The presence of *HU* greatly enhances the cyclization of DNA modeled as isotropically bendable base-pair steps (having the known ~500 Å persistence length of DNA). The effective persistence length decreases greatly in the range of the *in vivo HU* binding probability, while ring-closure becomes essentially independent of length and helical phasing. When peaks in *J* do occur, they are phase-shifted by approximately 3 bp, relative to the intrinsic (10.5 bp) helical repeat of DNA.

For chains of increasing length, the fraction of closed molecules with a negative linking number (I.e., -1, -2, …) becomes greater as the molecules become more untwisted and writhed.



Left: Topoisomer distribution of closed circles of different chain lengths captured in simulation of DNA with an average of 1 bound HU per 100 bp.

Right: Dependence of *J* factor on average HU spacing, 1/50 bp, 1/100 bp, and 1/200 bp.

Conclusion

Consistent with recent experimental measurements of the repression efficiency of the *Lac* operon in the presence or absence of *HU*[(ref 2)], *HU* proteins (like other DNA bending proteins such as *HMGB*) allow DNA to form short tight loops, supported computationally by the dramatic increase in *J*-factors and decreased effective persistence length of DNA, yielding ring-closure probabilities that are nearly constant over a wide range of chain lengths.

Indeed, our approach is promising in deciphering *lac* gene repression levels, without the need to invoke enhanced flexibility of the Lac repressor protein or of naked DNA (Swigon, Czapla, Olson, unpublished work), and is applicable to numerous other non-specific DNA-bending proteins such as *Nhp6A* and *HMGB1* (unpublished results). The combination of modeling protein binding and the sequence-dependent fluctuations of the double helix provides a valuable tool in deciphering gene expression patterns and predicting the looping properties of DNA under a variety of *in-vitro* conditions, enabling a better understanding of which components are essential for efficient looping.

Acknowledgements

I would like to thank Dr. L. James Maher, III for helpful suggestions and discussions. Also Dr. Irwin Tobias for discussions on DNA topology.

Computations were performed on a Rutgers BioMAPS Department cluster, using resources from the Rutgers Center for Computational Chemistry.

This work was funded by U.S.P.H.S. grant GM34809 to W.K.O, and a GAANN fellowship to L.C. is gratefully acknowledged.

[1] Czapla, L., Swigon, D., Olson, W.K., "Sequence-Dependent Effects in the Cyclization of Short DNA", JCTC (2)3, 685-695 (2006).
[2] Becker,N.A., Kahn, J.D., Maher, L.J., "Bacterial Repression Loops Require Enhanced DNA Flexibility", JMB (349) 716-730 (2005).