Supplementary Materials for: Mechanism of DNA compaction by yeast mitochondrial protein Abf2p

Raymond W. Friddle,*† Jennifer E. Klare,* Shelley S. Martin,† Michelle Corzett,§ Rod Balhorn,§ Enoch P. Baldwin,† Ronald J. Baskin,† and Aleksandr Noy*

*Biosecurity and Nanoscience Laboratory, Chemistry and Materials Science Directorate, Lawrence Livermore National Laboratory, 7000 East Ave, Livemore, CA 94550 †Department of Molecular and Cellular Biology, University of California at Davis, Davis, CA 95616 §Biology and Biotechnology Program, Lawrence Livermore National Laboratory, 7000 East Ave, Livermore, CA 94550

The mean-squared end-to-end distance for a Worm-Like Chain with Bends

Consider a polymer chain of n segments of length a. This chain contains an arbitrary number of identical pieces of p segments each. Each of the pieces is a freely rotating polymer chain (FRC)(Flory, 1969; Yamakawa, 1971). These pieces are joined by the angle ϕ, while all other joint angles within the piece are equal to θ, and all the dihedral angles set unrestricted. The mean squared end-to-end distance of the overall chain is given by,

$$\begin{aligned}\langle \mathbf{R}^2 \rangle &= \sum_{i=1}^{n}\sum_{j=1}^{n}\langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle \\ &= \sum_{i=1}^{n}\langle \mathbf{r}_i^2 \rangle + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n}\langle \mathbf{r}_i \cdot \mathbf{r}_j \rangle \end{aligned} \tag{1}$$

The first sum in Eq. 1 is simply the sum over a^2. The second sum is more involved. The factor of 2 ensures that we account for all terms in the first part of Eq. 1. Consider the case of $i = 1$, as we look at one piece from segments $j = 2, \ldots, p$.

$$\sum_{j=2}^{p}\langle \mathbf{r}_1 \cdot \mathbf{r}_j \rangle = \sum_{j=2}^{p}\alpha^{j-1}$$

Summing over the next piece should include the fixed angle ϕ in place of a θ. Let $\beta = -\cos\phi$, then for the next piece ($j = p+1, \ldots, 2p$) we have:

$$\sum_{j=p+2}^{2p}\frac{\beta}{\alpha}\alpha^{j-1} = \alpha^p\frac{\beta}{\alpha}\sum_{j=2}^{p}\alpha^{j-1}$$

Likewise, the sum over the third piece of the chain must include two factors of β and remove two factors of α,

$$\sum_{j=2p+2}^{3p}\frac{\beta^2}{\alpha^2}\alpha^{j-1} = \alpha^{2p}\left(\frac{\beta}{\alpha}\right)^2\sum_{j=2}^{p}\alpha^{j-1}$$

Correspondence and requests for materials should be addressed to A.N.: noy1@llnl.gov

Symbol	Definition
$\mathbf{R}$	End-to-end vector
$\mathbf{r}_i$	i^{th} Segment vector in the FRC
a	Segment length in the FRC
n	Total number of segments in the FRC
p	Number of segments per piece in the FRC ($\frac{n}{p}$ = number of pieces per chain)
θ	Segment bond angle in the FRC
ϕ	Bend angle
A	Persistence length
L	Contour length
P	Number of bends of angle ϕ in the bWLC
T	Length between adjacent bends in the bWLC
$\langle\beta\rangle$	Average excluded volume between two Kuhn segments
β_s	Excluded volume between unperturbed DNA Kuhn segments
β_l	Excluded volume between Kuhn segments of looped DNA (treated as cylinders)
n_o	Number of *effective* Kuhn segments per molecule
n_s	Number of Kuhn segments in an unperturbed DNA molecule
n_l	Number of Kuhn segments for DNA converted completely to solenoidal loops

Table 1: Definition of Parameters

We can then see that the $(m+1)^{th}$ piece contribution is given by,

$$\alpha^{mp}\left(\frac{\beta}{\alpha}\right)^m \sum_{j=2}^{p} \alpha^{j-1}$$

A chain of n segments divided into $\frac{n}{p}$ pieces contains $\frac{n}{p}-1$ bends. The entire sum of segment projections of $j > i$ for $i = 1$ is therefore given by,

$$\sum_{j=2}^{n} \langle \mathbf{r}_1 \cdot \mathbf{r}_j \rangle = a^2 \sum_{m=0}^{\frac{n}{p}-1} \alpha^{mp}\left(\frac{\beta}{\alpha}\right)^m \sum_{j=2}^{p} \alpha^{j-1} \tag{2}$$

Notice in Eq. 2 that the starting index in the sum over j is $j = 2$. Therefore we lose the projection of segment $\mathbf{r}_{p+1}$ on to $\mathbf{r}_1$ for each piece in the chain (i.e. the first segment of each piece is not included in the sum over j). Likewise, when $i = 2$ we lose the first and second segments of each piece. We will account for these lost projections later. Let Eq. 2, inserted into the second term of Eq. 1, be called $\mathbb{A}$ (not to be confused with the persistence length A). If we multiply every segment in the first piece with the rest of the $\frac{n}{p}-1$ pieces (and still neglect the lost terms) we obtain:

$$\mathbb{A}_0 = \sum_{m=0}^{\frac{n}{p}-1} \alpha^{mp}\left(\frac{\beta}{\alpha}\right)^m \sum_{i=1}^{p-1} \sum_{j>i}^{p} \alpha^{j-1}$$

The second piece is identical to the first except we now need only sum over $\frac{n}{p}-2$ pieces,

$$\mathbb{A}_1 = \sum_{m=0}^{\frac{n}{p}-2} \alpha^{mp}\left(\frac{\beta}{\alpha}\right)^m \sum_{i=1}^{p-1} \sum_{j>i}^{p} \alpha^{j-1}$$

and so on up until we reach the last piece in which it simply sums over itself and we only need $m = 0$. Therefore we write the entire sum as,

$$\mathbb{A} = \sum_{s=0}^{\frac{n}{p}-1} \mathbb{A}_s = \sum_{s=0}^{\frac{n}{p}-1} \sum_{m=0}^{\frac{n}{p}-1-s} \alpha^{mp} \left(\frac{\beta}{\alpha}\right)^m \sum_{i=1}^{p-1} \sum_{j>i}^{p} \alpha^{j-i} \tag{3}$$

Eq. 3 is just a combination of geometric series that can be easily solved. The result is given by,

$$\mathbb{A} = \left[\frac{\frac{n}{p} - \left(\frac{n}{p}+1\right)Q + Q^{\frac{n}{p}+1}}{(1-Q)^2}\right]\left[\frac{p\alpha}{1-\alpha} - \frac{\alpha - \alpha^{n+1}}{(1-\alpha)^2}\right] \tag{4}$$

where $Q = \alpha^{p-1}\beta$. Now we need to account for the lost terms in the sum of Eq. 3. As mentioned before, the construction of $\mathbb{A}$ periodically skips over segments as it sums over i. If we inspect Eq. 3, we can see that:

for $i =$	we miss $j =$
1	$mp + 1$
2	$mp + 1$, $mp + 2$
$\vdots$	$\vdots$
$p - 1$	$mp + 1$, $mp + 2$, $\ldots$, $mp + p - 1$

where $m = 1, 2, \ldots, \frac{n}{p} - 1$. We therefore construct the following sum which contains the above terms,

$$\mathbb{B} = \sum_{s=0}^{\frac{n}{p}-2} \sum_{m=1}^{\frac{n}{p}-1-s} \alpha^{mp} \left(\frac{\beta}{\alpha}\right)^m \sum_{i=1}^{p-1} \sum_{j=1}^{i} \alpha^{j-i} \tag{5}$$

The solution can again be simplified into geometric series, giving:

$$\mathbb{B} = \left[\frac{(\frac{n}{p}-1)Q - \frac{n}{p}Q^2 + Q^{2\frac{n}{p}}}{(1-Q)^2}\right]\left[-\frac{\alpha^2}{(1-\alpha)^2}(1-\alpha^{-p}) - p\frac{\alpha}{1-\alpha}\right] \tag{6}$$

A third and final term is needed to make up for another missed product in the sum in Eq. 3, which accounts for the dot product of the p^{th} segment of each piece with all higher number segments. This sum is simply given by,

$$\mathbb{C} = \sum_{s=0}^{\frac{n}{p}-2} \sum_{m=0}^{\frac{n}{p}-2-s} \alpha^{mp} \frac{\beta^{m+1}}{\alpha^m} \sum_{j=1}^{p} \alpha^{j-1} \tag{7}$$

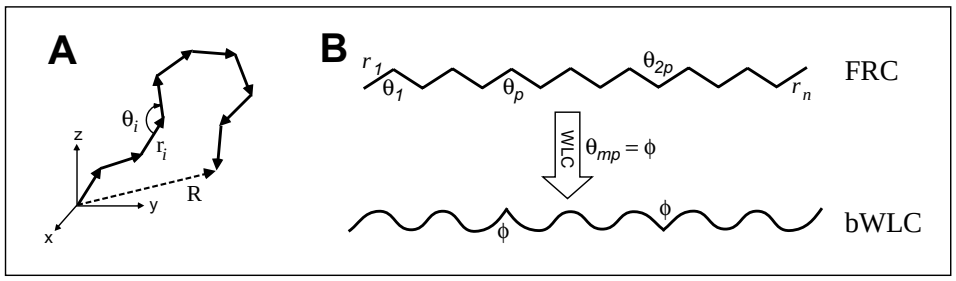

Figure 1: **A**. Diagram of a freely rotating chain (FRC) showing vector segments and bond angle assignments. **B**. Schematic representation of the relationship between the b-WLC model and the FRC model. We start with a FRC model that contains abnormal bend angles at regular intervals. The final b-WLC model retains the discontinuous bends of angle ϕ, yet the polymer chain is smooth everywhere else.

If we solve it similarly to $\mathbb{A}$ and $\mathbb{B}$, we obtain:

$$\mathbb{C} = \left[\frac{Q^{\frac{n}{p}}\beta - Q\beta + (\frac{n}{p} - 1)\beta}{(1-Q)^2}\right]\left[\frac{1-\alpha^p}{1-\alpha}\right] \tag{8}$$

The terms $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$ constitute the double sum in Eq. 1. Therefore, we obtain:

$$\langle \mathbf{R}^2 \rangle = na^2 + 2a^2\big[\mathbb{A} + \mathbb{B} + \mathbb{C}\big] \tag{9}$$

At this point, Eq. 9 describes a freely rotating chain with a number of *abnormal* bends at an angle ϕ that are positioned equidistant from one another along the chain. We call these bends abnormal because they are different from the normal bends θ found throughout in the regular FRC. (Yamakawa (1971) gives a general formulation for a FRC with periodic structure. Although we did not use this approach in our derviation of Eq. 9, we found that our approach gives an identical expression for $\langle \mathbf{R}^2 \rangle$ to Yamakawa's result in the freely-rotating coil limit.)

Finally, we take the limit to the WLC by the well known procedure employed by Kratky and Porod(Kratky & Porod, 1949; Flory, 1969) to generate a semi-stiff continuous contour chain between the bends (Figure 1). The following constraints are held constant in the limit,

$$\lim_{a\to 0,\ \alpha\to 1} \frac{a}{1-\alpha} \equiv A, \tag{10}$$

$$\lim_{a\to 0,\ n\to\infty} na \equiv L$$

where A is the persistence length derived from the normal FRC and L is the contour length. For the present model we must add another constraint for the length of each piece in the chain, that is:

$$\lim_{a\to 0,\ p\to\infty} pa \equiv T$$

where T is the length between adjacent bends. Making the estimation that as $\alpha \to 1$ then $-\ln\alpha \approx 1-\alpha$ and solving Eq. 10 for α we get:

$$\alpha = e^{-a/A} \tag{11}$$

Now, using the constraints of Eq. 10 and Eq. 11 we can simplify each term in Eq. 9.

$$\begin{aligned}
2a^2\mathbb{A} &= \left[\frac{\frac{L}{T} - (\frac{L}{T}+1)e^{-T/A}\beta + \left(e^{-T/A}\beta\right)^{\frac{L}{T}+1}}{\left(1-e^{-T/A}\beta\right)^2}\right]\left[2TA - 2A^2\left(1-e^{-T/A}\right)\right] \\
2a^2\mathbb{B} &= \left[\frac{(\frac{L}{T}-1)e^{-T/A}\beta - \frac{L}{T}\left(e^{-T/A}\beta\right)^2 + \left(e^{-T/A}\beta\right)^{2\frac{L}{T}}}{\left(1-e^{-T/A}\beta\right)^2}\right]\left[-2TA - 2A^2\left(1-e^{T/A}\right)\right] \\
2a^2\mathbb{C} &= 2aA\left[\frac{e^{-\frac{L}{A}}\beta^2 - e^{-T/A}\beta^2 + (\frac{L}{T}-1)\beta}{\left(1-e^{-T/A}\beta\right)^2}\right]\left(1-e^{-T/A}\right) \\
&\overset{a\to 0}{=} 0
\end{aligned}$$

The first term in Eq. 9 vanishes since $na^2 = La = 0$. The complete function for the ned-to-end distance $\langle \mathbf{R}^2\rangle_{bWLC}$ is given by,

$$\begin{aligned}
\langle \mathbf{R}^2\rangle_{bWLC} = &\left[\frac{\frac{L}{T} - (\frac{L}{T}+1)e^{-T/A}\beta + \left(e^{-T/A}\beta\right)^{\frac{L}{T}+1}}{\left(1-e^{-T/A}\beta\right)^2}\right]\left[2TA - 2A^2\left(1-e^{-T/A}\right)\right] \\
+ &\left[\frac{(\frac{L}{T}-1)e^{-T/A}\beta - \frac{L}{T}\left(e^{-T/A}\beta\right)^2 + \left(e^{-T/A}\beta\right)^{2\frac{L}{T}}}{\left(1-e^{-T/A}\beta\right)^2}\right]\left[-2TA - 2A^2\left(1-e^{T/A}\right)\right]
\end{aligned}$$

where T is the chain length between bends. We can finally derive a more useful form of this function if we notice that the number of bends P, is given by $P = \frac{L}{T} - 1$. We then obtain:

$$\begin{aligned}
\left\langle R^2\right\rangle_{bWLC} &= \left[\frac{1+P-(P+2)\Gamma+\Gamma^{P+2}}{(1-\Gamma)^2}\right]\left[\frac{2AL}{P+1} - 2A^2\left(1-e^{-\frac{L}{A(P+1)}}\right)\right] \\
&- \left[\frac{P\Gamma-(P+1)\Gamma+\Gamma^{2(P+1)}}{(1-\Gamma)^2}\right]\left[\frac{2AL}{P+1} + 2A^2\left(1-e^{\frac{L}{A(P+1)}}\right)\right] \qquad (12) \\
\text{where } \Gamma &= -\cos\phi e^{-\frac{L}{A(P+1)}}
\end{aligned}$$

This is the form given in Eq. 1 in the main manuscript, which gives $\langle \mathbf{R}^2\rangle_{bWLC}$ as a function of the number of bends.

Excluded Volume

Compaction enhances the interactions between DNA segments. We need to relate the parameters in the first order perturbation model, given by Eq. 3 in the main manuscript, to the number of bends. As bends are induced in the chain, some regions are perturbed

more so than others. Only at high coverage of the DNA will the entire chain have uniform interaction parameters describing its state. Therefore for intermediate numbers of bends we will approximate the situation with average parameters linearly related to the number of bends. We now define an average binary cluster integral,

$$\langle\beta\rangle = \frac{P\beta_l + (P_{max} - P)\beta_s}{P_{max}} \tag{13}$$

where β_s and β_l represent the unperturbed *segment* and the superhelical *loop* cluster integrals respectively. As these superhelical loops form, the effective contour length of the chain is reduced. We therefore write the number of segments as a function of increasing bends in the following way:

$$n_o = \frac{n_l - n_s}{P_{max}} P + n_s \tag{14}$$

where n_s and n_l are the number of segments in the free chain and completely covered chain respectively.

Assuming ABF2p binds as a monomer, Diffley and Stillman show the DNA footprint to be approximately 30 bp (Diffley & Stillman, 1991). Therefore, assuming a 30 bp$\times$0.34 nm/bp = 10 nm arc segment we estimate the radius of one superhelical loop as $r = 10\text{nm}/(102^\circ\pi/180^\circ)$ and therefore the diameter of the superhelical Kuhn segments is given by $D_l = 2r = 11.2$ nm. AFM measurements yield an estimate to the superhelical pitch of 15 nm. Using three ABF2 molecules per loop ($3 \times 102^\circ = 306^\circ \sim 360^\circ$) and using pBR322 DNA with 145 binding sites/3 = 48 loops, we have a final effective contour length of $n_l = 48 \times 15$ nm = 725 nm when completely covered in protein. With these calculations we find the following parameters for ABF2p binding to linearized pBR322,

$$\begin{aligned}
\beta_s &= \frac{\pi(100\text{ nm})^2 2\text{ nm}}{2} = 31.3 \times 10^3\text{ nm}^3 \\
\beta_l &= \frac{\pi(100\text{ nm})^2 11.2\text{ nm}}{2} = 175.9 \times 10^3\text{ nm}^3 \\
n_s &= 4361\text{ bp}\frac{0.34\text{ nm/bp}}{100\text{ nm}} = 14.8 \\
n_l &= 48\text{ loops}\frac{15\text{ nm/loop}}{100\text{ nm}} = 7.25
\end{aligned}$$

Finally, the expression for the mean-squared end-to-end distance with excluded volume correction is then given by:

$$\begin{aligned}
\langle R^2\rangle &= \langle R^2\rangle_{bWLC}\left(1 + \frac{4}{3}z\right) \\
z &= \left(\frac{3}{2\pi\langle R^2\rangle_{bWLC}}\right)^{3/2}\langle\beta\rangle n_o^2
\end{aligned}$$

REFERENCES

Diffley, J. F. and Stillman, B. 1991. A close relative of the nuclear, chromosomal high-mobility group protein hmg1 in yeast mitochondria. *Proc. Natl. Acad. Sci. USA*. 88:7864–8.

Flory, P. J. 1969. *Statistical mechanics of chain molecules*. Interscience Publishers, New York,. also reprinted by Hanser publishers, 1989.

Kratky, O. and Porod, G. 1949. Röntgenuntersuchung gelöster fadenmoleküle. *Rec. Trav. Chim. Pays-Bas.* 68:1106–1123.

Yamakawa, H. 1971. *Modern theory of polymer solutions*. Harper & Row, New York,.