

ICM Web: the interactive chromatin modeling web server

Richard C. Stolz^{1,2} and Thomas C. Bishop^{2,*}

¹Department of Biostatistics, Tulane University, 1440 Canal Street, Suite 2001, New Orleans, LA 70112 and ²Centre for Computational Science, Lindy Boggs Centre, Tulane University, Suite 500, New Orleans, LA 70118, USA

Received February 2, 2010; Revised May 7, 2010; Accepted May 18, 2010

ABSTRACT

The interactive chromatin modeling web server (ICM Web) is an interactive tool that allows users to rapidly assess nucleosome stability and fold sequences of DNA into putative chromatin templates. ICM Web takes a sequence composed of As, Cs, Gs, and Ts as input and generates (i) a nucleosome energy level diagram, (ii) coarse-grained representations of free DNA and chromatin and (iii) plots of the helical parameters (Tilt, Roll, Twist, Shift, Slide and Rise) as a function of position. The user can select from several different energy models, nucleosome structures and methods for placing nucleosomes in the energy landscape. Alternatively, if nucleosome footprints are known from experiment, ICM Web can use these positions to create a nucleosome array. The default energy model achieves a correlation coefficient of 0.7 with 100 experimentally determined values of stability and properly predicts the location of six positioned nucleosomes in the mouse mammary tumor virus (MMTV) promoter. ICM Web is suitable for interactively investigating nucleosome stability and chromatin folding for sequences up to tens of kilobases in length. No login is required to use ICM Web.

INTRODUCTION

Experimentally determined genome wide maps of nucleosome positions are becoming readily available (1–3). These data have prompted renewed interest in models of nucleosome positioning (4). Most models of nucleosome positioning rely on the idea that nucleosome stability is dependent upon the ability of a sequence of DNA to assume the structure observed in the nucleosome, as summarized in ref. (5). Low-energy, stable nucleosomes are formed when the DNA sequence exhibits a flexibility that enables it to more readily assume the nucleosome conformation or when the sequence exhibits deformations

that match the nucleosome's geometry. Little attention has been devoted to determining if the experimentally determined irregularly preferred positions can actually be mapped into three-dimensional structures. This is likely due to the fact that sequence-dependent nucleosome positioning is not the sole determinant of chromatin organization and that not all preferred nucleosome positions will be occupied simultaneously. It is well known that histone tails (6–8), linker histones (9) and many other factors play a role in determining the organization of chromatin. Models that account for such extra-nucleosomal interactions are being developed (10–12). The structure of chromatin is further complicated by the fact that nucleosomes exhibit a family of structures (13).

Here, we introduce a web server that can rapidly fold tens of thousands of base pairs into a nucleosome array or putative chromatin template using either an automated placement procedure or a list of defined nucleosome positions. The automated procedure uses a nucleosome stability model to predict the lowest energy conformation, while the list-based approach allows direct input of experimentally determined positions. By comparing the two methods of chromatin folding, one can assess the role of sequence-dependent nucleosome stability in determining chromatin architecture.

The web server offers a choice between several different nucleosome stability models. All models assume elastic deformations of individual base pairs in the nucleosome. The default provides a correlation of 0.7 with experimentally determined values of nucleosome stability for over 100 sequences of DNA (14). The mouse mammary tumor virus promoter sequence (MMTV) is used here as an example to demonstrate how users can employ these tools to investigate promoter function. The default settings properly predict the known positioning in the MMTV.

Below we follow the workflow in ICM Web as a means of introducing the underlying methods while also describing the server functionality. We conclude with an application of these tools to the MMTV.

*To whom correspondence should be addressed. Tel: +1 504 862 3370; Fax: +1 504 862 8392; Email: bishop@tulane.edu

WORKFLOW

Workflow in ICM Web begins at <http://dna.ccs.tulane.edu/icm>. The user inputs a sequence, selects an energy model, specifies the occupancy and provides a temperature (see Supplementary Figure S1). Clicking 'Go' produces the output: a 2^D energy level diagram, 3^D structures of free DNA and the nucleosome array and a 6^D helix parameter model (see Supplementary Figure S2).

Inputs

Sequence. The sequence of interest is entered into the primary text box on the main page as a string of As, Cs, Gs or Ts. Any other characters will generate an error code when the 'Go' button is activated. A set of more than 100 DNA sequences that have been extensively studied by our group (14) and others (15–17) is provided for convenience. A link to PubMed's Entrez Nucleotide Database is also provided. FASTA formatted sequences without the header lines are suitable input.

Energy options. The energy model is an elastic energy model that utilizes the inter-base-pair DNA helical parameters (Tilt, Roll, Twist, Shift, Slide and Rise) (18). Let values for the six helical parameters of free DNA be denoted by the list X_{DNA} . Since there are 16 canonical dinucleotide steps AA, AC, ... TG, TT, the lists are indexed as $X_{DNA}(j)$ $1 < j < 16$. Similarly, the helical parameters describing the conformation of nucleosomal DNA with n_{nuc} base pairs is denoted by the list $X_{nuc}(i)$ where $1 < i < n_{nuc} - 1$. For a sequence of length $n_{seq} > n_{nuc}$ base pairs, the energy associated with deforming any subset of n_{nuc} adjacent base pairs into the geometry required for a nucleosome is modeled as:

$$E_{nuc}(s) = \sum_{i=s}^{s+n_{nuc}-1} E_i \quad 1 \leq s \leq n_{seq} - n_{nuc} \quad (1)$$

where

$$E_i = [X_{nuc}(i) - X_{DNA}(j_i)]^T K(j_i) [X_{nuc}(i) - X_{DNA}(j_i)] \quad (2)$$

Here $K(j_i)$ denotes a 6×6 matrix of base pair step specific stiffness constants, and j_i is the sequence identity at step i . In the simplest model of DNA elasticity K is diagonal and reduces to independent linear constitutive relations for each helical parameter. For a more realistic description of DNA, K is not such a sparse matrix; there exist couplings between deformations of the various helical parameters. Sequence-specific values for K are available from refs. 19 and 20. For a sequence of length n_{seq} there are $n_{seq} - n_{nuc}$ values of $E_{nuc}(s)$. Each one represents the energy associated with a footprint of the histone octamer on the DNA that begins at s and extends to $s + n_{nuc}$.

DNA Stiffness and Conformation (K, X_{DNA}). Currently, ICM Web allows four different choices for the conformation of free DNA, X_{DNA} , and an accompanying set of stiffnesses, K . These are labeled **XrayB**, the B-DNA values from ref. 19; **XrayP**, the P-DNA (protein-DNA) values from ref. 19; **MD-B**, the values in ref. 20; and **IdealB**, an

idealized set of B-DNA values. The values from ref. 19 are based on analysis of X-ray crystallographic structures of B-form DNA or of protein-bound DNA, while the values from ref. 20 were obtained from analysis of molecular dynamics simulations of free DNA in B-form conformation.

The values in **IdealB** represent an idealized conformation of free DNA with a Twist of 35.4°/bp and Rise of 3.32 Å. The other helical parameters in the **IdealB** set are zero; thus, the intrinsic conformation of B-form DNA is straight for this set. It has no intrinsic bend or shear deformations. The matrix of stiffness constants for this idealized case corresponds to gross measures of stiffness that are sequence independent and exhibit no coupling, i.e. K is diagonal. This set is determined by averaging the sequence dependent values in ref. 20. Isotropic bend and shear values are determined as harmonic averages of the anisotropic values. The **IdealB** set (K and X_{DNA}) will not capture any of the known sequence-specific properties of DNA or sequence-specific nucleosome stability measures. The energy landscape is flat (i.e. constant) so it cannot be used to predict nucleosome positioning. However, the **IdealB** parameters sets, K or X_{DNA} , are useful for determining the role of intrinsic conformation, nucleosome conformation, X_{nuc} or sequence specificity, K , when mixed with other parameter sets.

Nucleosome conformation (X_{nuc}). The choices for conformation of the nucleosome, X_{nuc} , are more numerous. There are currently some 30 X-ray structures of the nucleosomes. Helical parameter sets for 20 of these are provided as possible choices. The choices on the web site correspond to the protein databank id tags (<http://www.rcsb.org>). Recently, Bishop conducted a study of all available X-ray structures to identify the necessary and sufficient distribution of helical parameter values required to achieve high-resolution models of nucleosomal DNA geometry (21). This was accomplished using Fourier filtering of the helical parameters to create so-called knockouts and knockins. The knockout and knockin series obtained from the crystal structure 1kx3 are provided. These lists of helical parameters are indicated as ko.NN.par or ki.NN.par where NN represents the knockout wave number or the total number of knockins (see ref. 21 for complete details). These structures do not necessarily provide realistic models of the nucleosome's 3^D structure. For some X-ray structures a 'min' selection is also offered. These structures contain only the necessary and sufficient distribution of helical parameter values required to achieve a high-resolution reconstruction of the original X-ray structure, i.e. within 3 Å RMSD, as determined in ref. 21. Since the helical parameters in the sets denoted with a min have been smoothed by Fourier filtering, the energy landscape predicted with these data sets is also smooth compared to the corresponding non-filtered data set. Additionally, the energy values are lower.

The default settings for K , X_{nuc} and X_{DNA} provide energy values that have a 0.7 correlation with over 100 experimentally determined values of nucleosome stability (14), 84 from refs. 15 and 22 and 33 from ref. 16.

Thermal fluctuations. For DNA free in solution, random thermal fluctuations cause its conformation to deviate from the intrinsic helical parameters, X_{DNA} . In nucleosomes, the gross super helical structure of DNA must be preserved. Therefore, thermal motion of nucleosomal DNA is not entirely random. ICM Web allows the user to introduce thermal fluctuations into the model of free DNA and the linker DNA between nucleosomes. For this purpose, a temperature must be given. Temperature serves as a scale factor for the Gaussian noise added to each base pair step parameter in the free DNA or linker DNA regions as the 3^D models are constructed. For each helical parameter a value is determined by $X_{\text{temp}} = \text{Norm}(X_{\text{DNA}}, \sqrt{T/298}\sigma_{\text{ABC}})$ where $\text{Norm}(\mu, \sigma)$ is a routine that provides a Gaussian distribution with mean value μ and deviation σ . Here, T represents the temperature in Kelvin specified on the web form and σ_{ABC} represents the sequence independent fluctuation values for Tilt, Roll, Twist, Shift, Slide and Rise obtained from molecular dynamics simulations of B-form DNA free in solution (23). A temperature of zero thus provides no variation from the X_{DNA} set of helical parameter values selected as input. The random seed is fixed so that the results are reproducible.

Occupancy. There are two placement options available. One defines the starting coordinate of each nucleosome footprint. The other allows ICM Web to automatically place nucleosomes in the energy landscape based on occupancy criteria.

We define occupancy as the number and location of nucleosomes on a segment of DNA. Free DNA corresponds to an occupancy of zero. The maximum number of nucleosomes that can be placed on a segment of DNA is $n_{\text{max}} = n_{\text{seq}}/(n_{\text{nuc}} + n_{\text{lk}})$ where n_{seq} is the length of the input sequence, n_{nuc} is the number of base pair in the nucleosome and n_{lk} is the minimum allowed length of linker between nucleosomes. Occupancy is controlled on the server by specifying n_{occ} and n_{lk} . If $n_{\text{occ}} > 1$ an attempt will be made to place this many nucleosomes on the sequence of DNA. If $0 < n_{\text{occ}} < 1$, n_{occ} is interpreted as a percent of n_{max} . In practice, we find that 70–80% occupancies can be readily achieved with our simple population method described below. Higher occupancies cannot usually be realized with this method. To achieve occupancies approaching 100%, the user has to specify nucleosome start sites.

The automated placement option utilizes the energy landscape to populate the landscape as follows: the global minimum of the energy landscape is determined by sorting all possible locations and a nucleosome is located at this position. This nucleosome's footprint occludes n_{nuc} base pair from the landscape. The sort and locate process repeats until the desired occupancy is achieved. The specification of a nonzero linker length extends the nucleosome footprint by $n_{\text{lk}}/2$. This is an *ad hoc* method of avoiding steric overlap between neighboring nucleosomes without including an explicit nucleosome-nucleosome interaction potential. This simple occupancy algorithm does not necessarily guarantee the lowest total energy conformation for the

nucleosome array. More sophisticated algorithms are being considered, but for web functionality they must also be fast.

Since $n_{\text{occ}} < 1$ is a non-physical occupancy we utilize such values to trigger alternate methods of populating the segment of DNA with nucleosomes. If $n_{\text{occ}} = 0$ then nucleosomes are located at regular intervals spaced by n_{lk} , i.e. 0, $n_{\text{nuc}} + n_{\text{lk}}$, $n_{\text{nuc}} + 2 n_{\text{lk}}$, ... If $n_{\text{occ}} < 0$ the nucleosomes are spaced at regular intervals as in the $n_{\text{occ}} = 0$ case but the positions of all nucleosomes are shifted by n_{occ} base pairs. For example, if the occupancy is -5 the footprints are 5, $n_{\text{nuc}} + n_{\text{lk}} + 5$, ...

Instead of using our automated procedure, a list of nucleosome positions can be specified directly. In this case, an energy landscape is calculated and reported, but only the specified positions are used to determine nucleosome start sites. This allows users to utilize their own algorithms for determining nucleosome positioning or to use experimentally determined values as compiled in ref. 3 to generate a putative chromatin fold. On the ICM input page links are provided to other web sites that predict nucleosome positions.

Outputs

The 'Reset' button clears all forms and sets all variables to default values. The 'Go' button initiates three calculations and generates the results page. The first calculation determines the energy landscape. The second determines nucleosome footprints via the iterative search criteria mentioned above or checks the validity of the user defined start positions before placing nucleosomes at the indicated positions. Two sets of helical parameters are then generated: one describes free DNA; the other describes the putative chromatin template. The third calculation converts the helical parameter representations to Cartesian coordinate representations of the DNA using the algorithm defined in ref. 24.

The data from these calculations is presented on the results page (Supplementary Figure S2), which contains an energy landscape and footprints displayed as a 2^D nucleosome energy level diagram; DNA helical parameters displayed a set of six plots, a 6^D model; and two 3^D models, one for free DNA and one for the nucleosome array. Since nucleosome positions are calculated based only on nucleosome-DNA interactions (i.e. in helical parameter space) or obtained as a list of start sites, there is no *a priori* check to identify unfavorable steric interactions. For this reason, a distance calculation is conducted on the 3^D nucleosome array to identify such steric clashes. The minimal nucleosome-nucleosome, nucleosome-DNA and DNA-DNA distances are reported.

The 3^D models are 4 points-per-base pair models of DNA in xyz format. The four atoms define the centerline (atom name CA) and the local helical axis frame (atom names H1, H2 and H3). For the chromatin model, an additional atom is located at the center of geometry of each nucleosome (atom name OC for octasome center). The 3^D models are displayed using Jmol: an open-source Java viewer for chemical structures in 3^D (<http://www.jmol.org/>) (25). Jmol displays a sphere of diameter 2 nm

every 5 bp along the DNA and a sphere of diameter 8nm for each histone octamer. Spheres in close contact are colored red. Close contacts are defined as $< 110 \text{ \AA}$ for nucleosome-nucleosome spacing, $< 65 \text{ \AA}$ for nucleosome-DNA spacing, and $< 20 \text{ \AA}$ for DNA-DNA spacing. The user has full control over the display in Jmol and can color base pairs to display biological information (Figure 3). A 'Save Data' link allows for download of all data. The helical parameter files can be converted into all-atom models by 3DNA (26) or a coarse-grain representation with a VMD plug-in that we developed, VDNA (27).

INTERPRETATION OF RESULTS

As noted in methods, if X_{DNA} and K are chosen to be **IdealB** then the total energy is constant, regardless of sequence. The only determinant of energy in this case is X_{nuc} , the conformation of the nucleosome. If X_{nuc} is set to 1kx5.min the energy required to deform DNA into this nucleosome is 151 kcal/mol-nuc $\sim 1.7 \text{ k}_b\text{T}/\text{bps}$. For $X_{\text{nuc}} = 1\text{kx}5$ the value is 417 kcal/mol-nuc $\sim 4.8 \text{ k}_b\text{T}/\text{bps}$. The corresponding results for pdb entry 1kx3 are 173 kcal/mol-nuc $\sim 2.0 \text{ k}_b\text{T}/\text{bps}$ and 502 kcal/mol-nuc $\sim 5.8 \text{ k}_b\text{T}/\text{bps}$ (1kx3 has 146 instead of 147 bp). These results agree well with physical expectations. For a simple

homogenous model of DNA, $\sim 2 \text{ k}_b\text{T}/\text{bps}$ is needed to wrap DNA around the histone octamer (28). Our smoothed nucleosomal DNA super-helices, 1kx5.min or 1kx3.min, yield this result. The helical parameters obtained directly from X-ray structures, 1kx5 or 1kx3, have not been smoothed so the energy is higher. Since there are 6 degrees of freedom in our model of DNA we expect the effects of thermal motion to contribute an extra $3 \text{ k}_b\text{T}/\text{bps}$ over the minimum energy conformation. Again, our results agree with this expectation. Similar energy values are obtained if the sequence specific parameter set denoted, **MD-B**, is used.

Using the sequences in refs. 16 and 17 as a test set we have demonstrated that the default energy options ($K = \text{MD-B.dat}$, $X_{\text{nuc}} = 1\text{kx}5.\text{min.par}$, and $X_{\text{DNA}} = \text{MD-B.par}$) achieve a correlation of 0.7 with the experimentally determined values of $\Delta\Delta G$ associated with these sequences of DNA. Results using the **XrayB** and **XrayP** sets are not so well correlated. A complete analysis has been reported elsewhere (14).

Thermal fluctuations strongly affect the 3^D conformations of the free DNA model and nucleosome array (Figure 1). For the free DNA conformation thermal motion introduces a random distribution about the intrinsic conformation of each base pair step. For nucleosome arrays, thermal fluctuations are only introduced into the

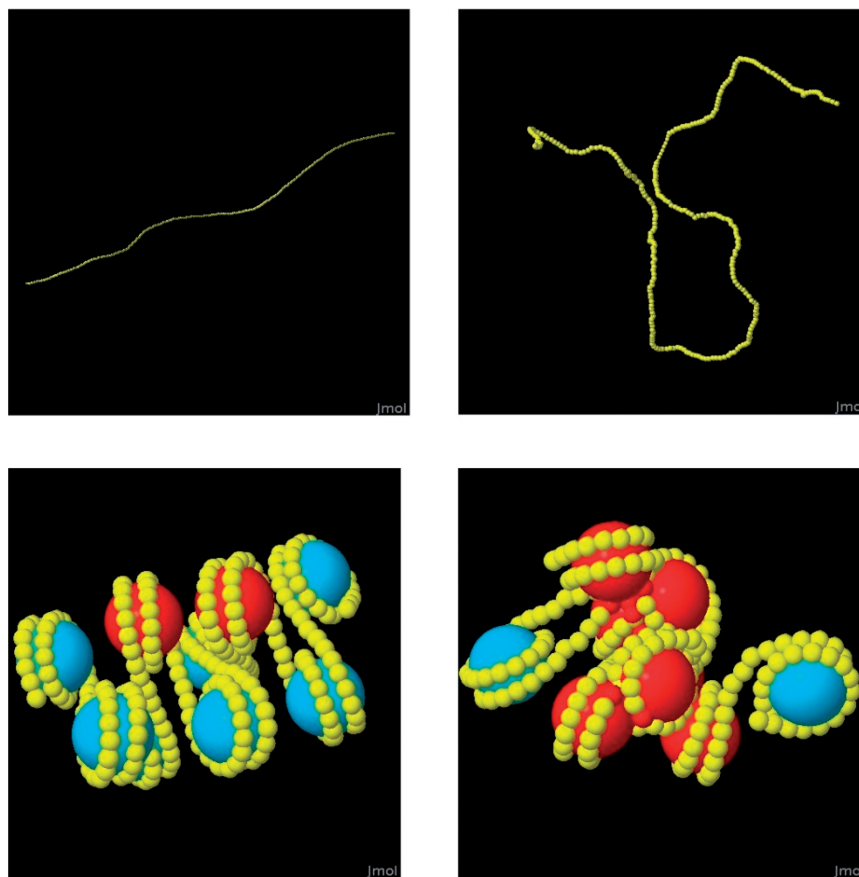


Figure 1. Effect of thermal fluctuations. Thermal fluctuations (left $T = 0 \text{ K}$; right $T = 298 \text{ K}$) affect both the conformation of free DNA (top images) and the nucleosome arrays (bottom images). For DNA each base pair step allows random thermal motion. In the nucleosome arrays, only the linker DNA allows thermal fluctuations. This can have a significant effect on the overall geometry.

linker regions. If the temperature is set to 0 K, then the equilibrium conformation is defined only by the choice of helical parameters X_{DNA} . A zero temperature model is useful for identifying intrinsic bends or other deformations in the free DNA and linker regions. A nonzero temperature is useful for investigating effects of flexibility on overall conformation.

The choice of occupancies also strongly affects the overall conformation of the nucleosome array. We provide a series of images to demonstrate various occupancies (Figure 2). All of the images have been assigned a temperature of 298 K. The images include uniform nucleosome spacing (i.e. constant linker) with different linker lengths and irregular nucleosome spacing determined by our minimum energy criteria with different percent occupancies.

Sample application: the MMTV promoter

We utilize the MMTV long terminal repeat promoter as a sample sequence because it is widely used as an experimental construct to regulate gene expression. It has been extensively characterized (29–33). The sequence GR-MMTV (34) used here can be obtained from PubMed as entry V01175. The C3H strain of the MMTV, J02274 (35) is also available from PubMed. Differences between these sequences are discussed in ref. 36.

This promoter is steroid activated but some basal transcription occurs even in the inactive state. This promoter is known to position six nucleosomes, denoted [A–F], that assemble into an organized chromatin structure (30,37). The location of these nucleosomes are indicated in Figure 3 with A +70, B –127, C –415, D –672, E –918 and F –1139. (The reported positions are for the nucleosome dyad position relative to the transcription start. Our E_{nuc} uses the footprint start site so subtract 73 from these values.) The promoter contains four binding sites for the glucocorticoid receptor (GR), two octamer transcription factor 1 binding sites (Oct) and a nuclear factor 1 (NF1) binding site (34). Nucleosomes A and B are of particular

interest because nucleosome A overlaps with the TATA element, while nucleosome B overlaps with the NF1 and four GR binding sites. The rotational phasing of B leaves the GR binding sites open to binding (38). Steroid activation alters the organization of nucleosomes B and C (39,40), and establishes a specific translational phasing that does not exist in the inactive state (41). There also appears to be a boundary effect resulting from the stability of nucleosomes A and D, while the positioning of C is influenced by the more proximal region of the promoter (41). Recent models for MMTV structure and mechanism of action are presented in (42) and (43), respectively. A bent secondary chromatin structure is proposed in figure 7 of ref. 42. Given this information, we can readily interpret the nucleosome energy level diagram and the models predicted by ICM Web.

The energy landscape predicted with the default energy options produces the locations of six positioned nucleosomes in the MMTV (37,40) with a high degree of accuracy. Nucleosome D is the most stable and is separated from A, B and C by a region of high energy where nucleosomes are least likely to form. We propose that this feature corresponds to the observed boundary effect. The models returned in the Jmol viewer clearly indicate that the MMTV with six positioned nucleosomes is a much more open structure than depicted in the literature (42). For this array, the minimum nucleosome-nucleosome distance is nearly 24 nm, thus the histone tails from different nucleosomes cannot interact. The TATA box (green) and most distal GR binding site (red) are just outside the nucleosome footprint and are in close proximity to one another as the linker enters and exits nucleosome B. The Oct 1 binding site (orange) closest to the TATA box has a GR binding site adjacent to it, while the NF1 binding site (blue) is directed away from the entire nucleosome array.

It is not likely that the length of free DNA between nucleosomes will assume the extended conformation indicated by ICM Web (Left image in Figure 3); rather,

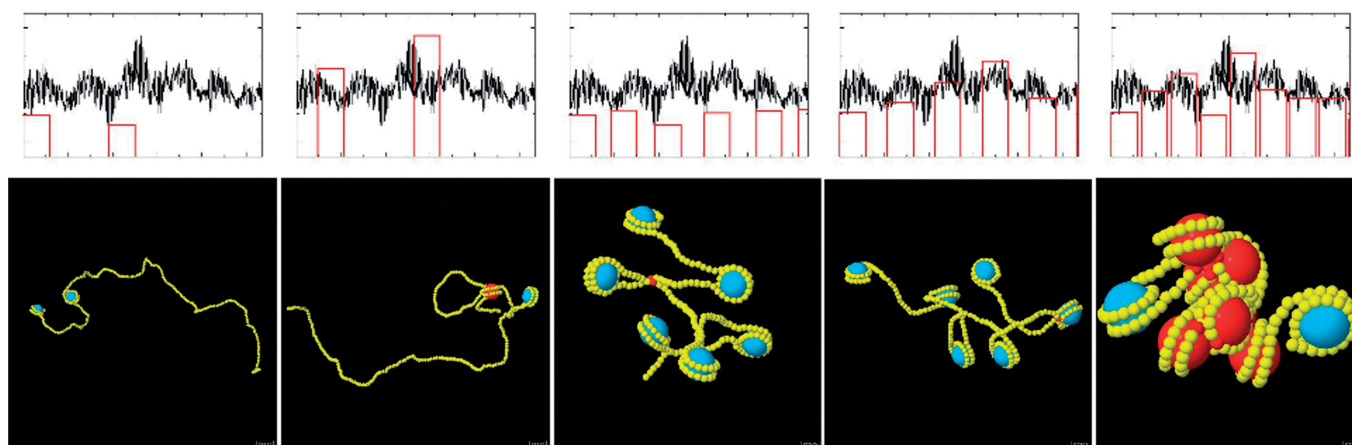


Figure 2. Effect of occupancy. The location of nucleosomes on DNA is a primary determinant of the conformation of the nucleosome array. Here, nucleosome arrays formed with the MMTV sequence and different occupancies are displayed. All have the same energy landscape and a temperature of 298 K. (Left to right) Two nucleosomes automatically placed at the lowest energy positions; two positions specified at high energy locations; 70% occupancy and linker of 20 bp (the default); uniform spacing with a linker of 120 bp; and uniform spacing with a linker of 20 bp.

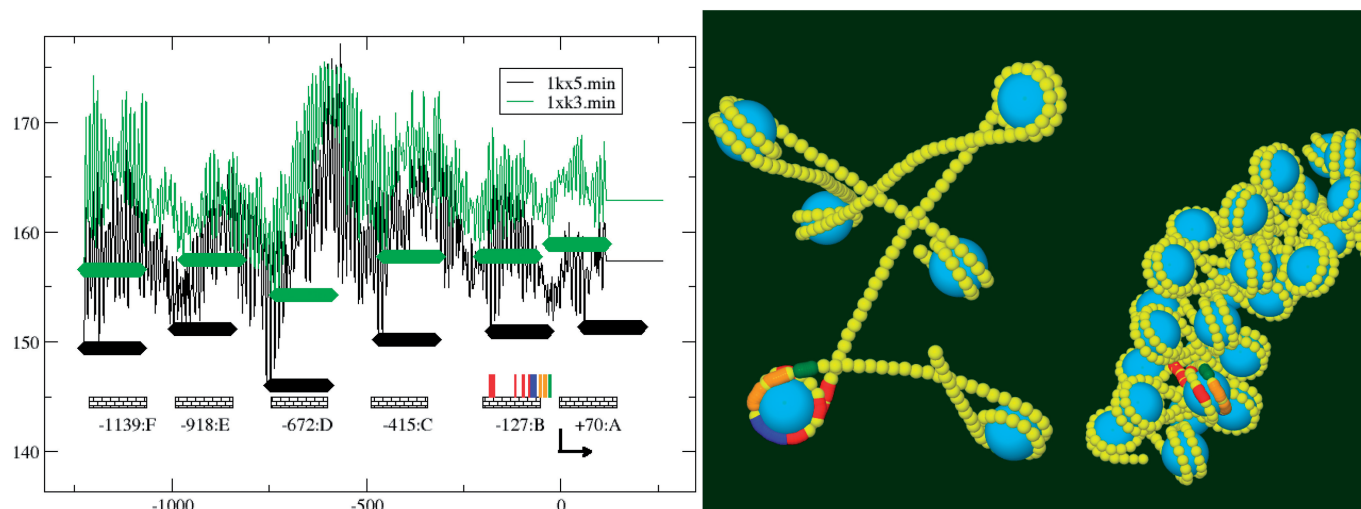


Figure 3. Energy level diagram and conformation of MMTV. (Left) The nucleosome energy level diagram obtained for the MMTV sequence using energy options: 1kx5.min.par and 1kx3.min.par for X_{nuc} (as indicated), MD-B for K and X_{DNA} . Energy values are in kcal/mol–nuc. Solid bars indicate the predicted nucleosome footprints and have been shifted so that the transcription start site corresponds to position 0, as indicated by the arrow. Also included are the known positions of nucleosomes A–F, indicated as tiled boxes, glucocorticoid response elements (red), NF1 (blue) and Oct (orange) binding sites, and the TATA box (green). (Middle) The 3^D structure of the MMTV associated with this energy landscape for a temperature of 0 K. Yellow spheres represent 5 bp. Histone cores as light blue spheres. Here, the glucocorticoid response elements are red, NF1 binding sites are blue, octamer binding sites are orange, and the TATA box is green. Nucleosome A is in the bottom right and B is in the bottom left. (Right) Five tandem repeats of the MMTV with a uniform spacing ($n_{\text{lk}} = 30$) of nucleosomes, and temperature 0 K. The first repeat is shifted to achieve the lowest energies. The bend only arises from deformations present in the linker DNA.

extra-nucleosomal interactions and thermal fluctuations will cause the linker DNA to deviate from its equilibrium conformation. However, the fact that six positioned nucleosomes leaves significant lengths of DNA exposed has not been specifically addressed in the MMTV literature. For highly purified *in vitro* systems, extra-nucleosomal interactions may cause the linker to self-interact or bend and enable the MMTV to further condense. For *in vivo* systems of chromatin, not only are there spatial restrictions on the chromatin, but there are also numerous other factors that come into play.

For these reasons, we allow the user to override the energy landscape and specify a uniform distribution of nucleosomes or an arbitrary set of nucleosome locations. A model of condensed chromatin can be obtained by selecting the number of nucleosomes, n_{occ} , to zero (uniform spacing) and a linker length, n_{lk} , of 30 bp (Right image in Figure 3). The minimum energy for the condensed array is obtained when all nucleosomes are positioned with a phase shift of 6 ($n_{\text{occ}} = -6$), i.e. the first nucleosome starts at position 6. This positioning locates nucleosome B such that the TATA box (green) is outside of the footprint of nucleosome B (right image in Figure 3). The TATA box also appears on the surface of the fiber. At the other end of nucleosome B, the GR binding sites are shifted into the linker regions and are located deep within the chromatin fiber. However, as the right image in Figure 3 indicates, even when the GR binding sites are located deep within the fiber they are still very accessible. This result agrees well with observations of fast on-off rates for GR *in vivo* (44). ICM Web also predicts a bent chromatin structure. Since all nucleosomes are identical

the bend in chromatin arises only from deformations in the linker DNA.

CONCLUDING REMARKS

We have demonstrated how ICM can be achieved in real time using a simple elastic model to automatically place nucleosomes. We have also introduced the idea of a nucleosome energy level diagram as an intuitive means of tracking energy differences associated with different nucleosome footprint configurations. Our present model of nucleosome stability is overly simplistic. Clearly, a more complete description of chromatin will account for nucleosome–nucleosome, nucleosome–linker, linker–linker interactions and all other types of extra-nucleosomal interactions. Our ‘linker’ length is an *ad hoc* method of providing some of this functionality. There are a number of efforts underway to investigate nucleosome positioning, see for example, ref. (45). Such effects will be included in future versions of our tools.

Since a unique structure for chromatin may not exist, we regard ICM Web, not so much as a predictive tool, but as a tool for rapidly assembling models of chromatin that can be employed to rationalize biophysical data, especially spatial relations. Such a tool is crucial for investigating molecular mechanisms of chromatin. To our knowledge, there is no such tool available. ICM Web provides this functionality.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We wish to acknowledge Gregory Foster for his efforts in creating an initial version of this web page. Carleigh Hebbard, Austin Griffith, and Yuriy Sereda are acknowledged for testing the web page and other contributions to our efforts.

FUNDING

Funding for open access charge: National Institutes of Health (R01GM076356 to T.B.).

Conflict of interest statement. None declared.

REFERENCES

- Segal, E., Fondudfe-Mittendorf, Y., Chen, L.Y., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P.Z. and Widom, J. (2006) A genomic code for nucleosome positioning. *Nature*, **442**, 772–778.
- Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G. and Zhao, K. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell*, **132**, 887–898.
- Jiang, C. and Pugh, B. (2009) A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome. *Genome Biol.*, **10**, 161–172.
- Rando, O.J. and Ahmad, K. (2007) Rules and regulation in the primary structure of chromatin. *Curr. Opin. Cell Biol.*, **19**, 250–256.
- Widom, J. (2001) Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.*, **34**, 269–324.
- Garcia-Ramirez, M., Dong, F. and Ausio, J. (1992) Role of the histone “tails” in the folding of oligonucleosomes depleted of histone H1. *J. Biol. Chem.*, **267**, 19587–19595.
- Fletcher, T.M. and Hansen, J.C. (1995) Core histone tail domains mediate oligonucleosome folding and nucleosomal DNA organization through distinct molecular mechanisms. *J. Biol. Chem.*, **270**, 25359–25362.
- Logie, C., Tse, C., Hansen, J.C. and Peterson, C.L. (1999) The core histone N-terminal domains are required for multiple rounds of catalytic chromatin remodeling by the SWI/SNF and RSC complexes. *Biochemistry*, **38**, 2514–2522.
- Georgel, P.T. and Hansen, J.C. (2001) Linker histone function in chromatin: dual mechanisms of action. *Biochem. Cell. Biol.*, **79**, 313–316.
- Stehr, R., Kepper, N., Rippe, K. and Wedemann, G. (2008) The effect of internucleosomal interaction on folding of the chromatin fiber. *Biophys. J.*, **95**, 3677–3691.
- Arya, G., Zhang, Q. and Schlick, T. (2006) Flexible histone tails in a new mesoscopic oligonucleosome model. *Biophys. J.*, **91**, 133–150.
- Li, W., Dou, S.X. and Wang, P.Y. (2004) Brownian dynamics simulation of nucleosome formation and disruption under stretching. *J. Theor. Biol.*, **230**, 375–383.
- Zlatanova, J., Bishop, T.C., Victor, J.M., Jackson, V. and van Holde, K. (2009) The nucleosome family: dynamic and growing. *Structure*, **17**, 160–171.
- Yuriy, S. and Bishop, T.C. (2010) Evaluation of elastic rod models with long range interactions for predicting nucleosome stability. *J. Biomol. Struct.*, **27**, 867–887.
- Anselmi, C., De Santis, R., Paparcone, R., Savino, M. and Scipioni, A. (2002) From the sequence to the superstructural properties of DNAs. *Biophys. Chem.*, **95**, 23–47.
- Fernandez, A.G. and Anderson, J.N. (2007) Nucleosome positioning determinants. *J. Mol. Biol.*, **371**, 649–668.
- Anselmi, C., Bocchinfuso, G., Santis, P.D., Savino, M. and Scipioni, A. (1999) Dual role of DNA intrinsic curvature and flexibility in determining nucleosome stability. *J. Mol. Biol.*, **286**, 1293–1301.
- Dickerson, R.E. (1989) Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.*, **17**, 1797–1803.
- Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. and Zhurkin, V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.
- Lankas, F., Spomer, J., Langowski, J. and Cheatham, T.E. (2003) DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.*, **85**, 2872–2883.
- Bishop, T.C. (2008) Geometry of the nucleosomal DNA superhelix. *Biophys. J.*, **95**, 1007–1017.
- Anselmi, C., Santis, P.D., Paparcone, R., Savino, M. and Scipioni, A. (2004) A possible role of DNA superstructures in genome evolution. *Orig. Life Evol. Biosph.*, **34**, 143–149.
- Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T.C., Case, D.A., Cheatham, T. III, Dixit, S., Jayaram, B., Lankas, F., Laughton, C. et al. (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.*, **38**, 299–313.
- el Hassan, M.A. and Calladine, C.R. (1995) The assessment of the geometry of dinucleotide steps in double-helical DNA: a new local calculation scheme. *J. Mol. Biol.*, **251**, 648–664.
- Herraez, A. (2006) Biomolecules in the computer. *Biochem. Mol. Biol. Educ.*, **34**, 255–261.
- Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
- Bishop, T.C. (2009) VDNA: the virtual DNA plug-in for VMD. *Bioinformatics*, **25**, 3187–3188.
- Langowski, J. (2006) Polymer chain models of DNA and chromatin. *Eur. Phys. J. E. Soft Matter*, **19**, 241–249.
- Perlmann, T. and Wrangé, O. (1988) Specific glucocorticoid receptor binding to DNA reconstituted in a nucleosome. *EMBO J.*, **7**, 3073–3079.
- Richard-Foy, H., Sistare, F.D., Riegel, A.T., Simons, S.S. and Hager, G.L. (1987) Mechanism of dexamethasone 21-mesylate antiglucocorticoid action: II. Receptor-antiglucocorticoid complexes do not interact productively with mouse mammary tumor virus long terminal repeat chromatin. *Mol. Endocrinol.*, **1**, 659–665.
- Richard-Foy, H. and Hager, G.L. (1987) Sequence-specific positioning of nucleosomes over the steroid-inducible MMTV promoter. *EMBO J.*, **6**, 2321–2328.
- Fletcher, T.M., Ryu, B.W., Baumann, C.T., Warren, B.S., Fragoso, G., John, S. and Hager, G.L. (2000) Structure and dynamic properties of a glucocorticoid receptor-induced chromatin transition. *Mol. Cell. Biol.*, **20**, 6466–6475.
- Hager, G.L. (2001) Understanding nuclear receptor function: from DNA to chromatin to the interphase nucleus. *Prog. Nucleic Acid Res. Mol. Biol.*, **66**, 279–305.
- Fasel, N., Pearson, K., Buetti, E. and Diggelmann, H. (1982) The region of mouse mammary tumor virus DNA containing the long terminal repeat includes a long coding sequence and signals for hormonally regulated transcription. *EMBO J.*, **1**, 3–7.
- Donehower, L.A., Huang, A.L. and Hager, G.L. (1981) Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. *J. Virol.*, **37**, 226–238.
- Grimm, S.L. and Nordeen, S.K. (1998) Mouse mammary tumor virus sequences responsible for activating cellular oncogenes. *J. Virol.*, **72**, 9428–9435.
- Fragoso, G., John, S., Roberts, M.S. and Hager, G.L. (1995) Nucleosome positioning on the MMTV LTR results from the frequency-biased occupancy of multiple frames. *Genes Dev.*, **9**, 1933–1947.

38. Flaus,A. and Richmond,T.J. (1998) Positioning and stability of nucleosomes on MMTV 3'LTR sequences. *J. Mol. Biol.*, **275**, 427–441.
39. Archer,T.K., Lefebvre,P., Wolford,R.G. and Hager,G.L. (1992) Transcription factor loading on the MMTV promoter: a bimodal mechanism for promoter activation. *Science*, **255**, 1573–1576.
40. Smith,C.L. and Hager,G.L. (1997) Transcriptional regulation of mammalian genes in vivo. A tale of two templates. *J. Biol. Chem.*, **272**, 27493–27496.
41. Belikov,S., Gelius,B., Almouzni,G. and Wrangé,O. (2000) Hormone activation induces nucleosome positioning in vivo. *EMBO J.*, **19**, 1023–1033.
42. Belikov,S., Holmqvist,P.-H., Astrand,C. and Wrangé,O. (2004) Nuclear factor 1 and octamer transcription factor 1 binding preset the chromatin structure of the mouse mammary tumor virus promoter for hormone induction. *J. Biol. Chem.*, **279**, 49857–49867.
43. Belikov,S., Astrand,C. and Wrangé,O. (2007) Mechanism of histone H1-stimulated glucocorticoid receptor DNA binding in vivo. *Mol. Cell. Biol.*, **27**, 2398–2410.
44. Stavreva,D.A., Miller,W.G., Hager,G.L., Smith,C.L. and McNally,J.G. (2004) Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Mol. Cell. Biol.*, **24**, 2682–2697.
45. Current Perspectives on Nucleosome Positioning (2010) *J. Biomol. Struct. Dyn.*, **27**, 713–894.