Simulating chromatin: Gene prediction at the structural level?

Steve Dennis

Genetic analysis of DNA has traditionally been done in one dimension (i.e., the DNA sequence), while 3D analysis of molecular structure is usually reserved for proteins. The actual transcribing of DNA into proteins however, involves a 3D structure - the chromatin complex. Since chromatin is composed of proteins as well as DNA, questions could be asked that are similar to those asked for proteins: Given a DNA sequence at a particular chromosomal location from a cell at a specific cycle or stage of differentiation, can we predict the structure of the corresponding chromatin segment, particularly the positions of the nucleosomes? Next, can we simulate conformational changes in the chromatin segment during transcription, and can that knowledge be used to then find new genes or predict the function of known genes? And finally, is all this computationally feasible at the present time?

Introduction

Gene promoter and regulatory regions are often located far upstream of the actual coding region on a DNA strand. These regions can sometimes be identified by sequence analysis, but such methods can't answer why certain distant sequences are necessary for transcription to take place. It isn't until relevant molecular structures or models are available that the process starts to make sense.

This paper is a review of research related to chromatin, the cell nucleus DNA packaging system for eukaryotic organisms. Chromatin, along with its role in transcription and gene expression, is a part of a hugely complex research area, so the focus here will be on chromatin in the phase just prior to transcription, and in particular the details of the nucleosome and nucleosomal DNA. Included in this report are summarized current experimental findings, some new hypotheses, and an overview of current bioinformatical and computational methods. A more detailed look at the nucleosome is followed by some proposed computational methods for producing a biologically relevant chromatin structure from a sequence of DNA.

Current knowledge of chromatin

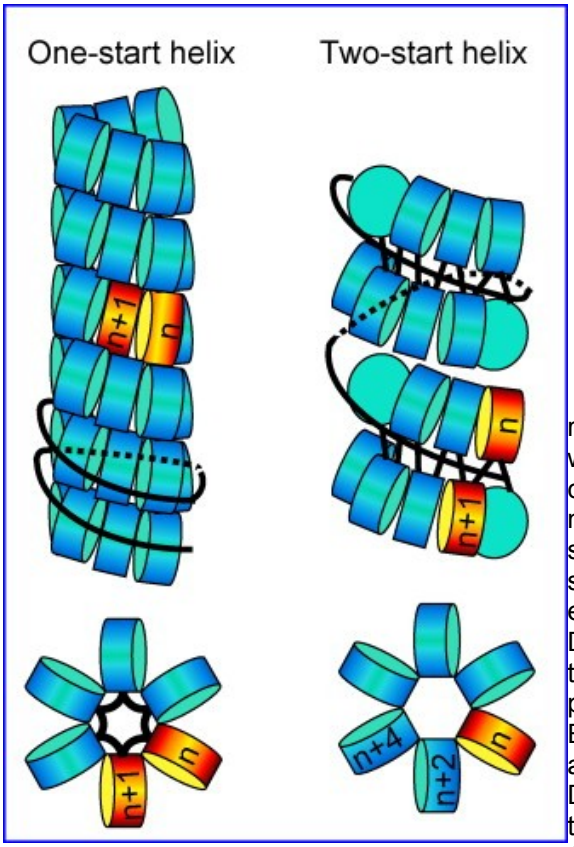

Fig. 1. [1]

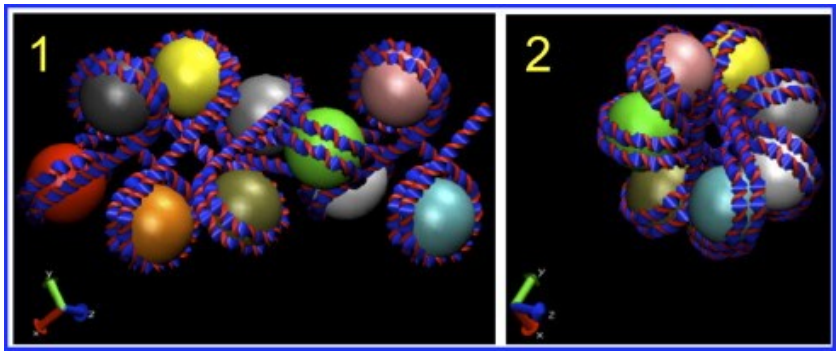

Fig. 2. Two-angle model for chromatin [1]

The chromatin being studied is the “30 nm” (diameter) transcription-ready fiber that exists in interphase cells. Double-stranded DNA is wrapped around clustered nucleosomes, with each nucleosome consisting of a core of four pairs (an octamer) of histone proteins – the nucleosome core particle (NCP). More specifically, a 147 base pair-long strand is wrapped 1.67 times around the core (although there can be some variation on this). Each of the histones has an N-terminus tail which extends outside the nucleosome, probably helps secure the wrapped DNA, and can be chemically activated for transcription-related changes in the nucleosome. The core histones are so highly conserved that for many purposes they can be treated as identical across all eukaryotic organisms. Even so, there can be subtype variants based on cell differentiation. An additional H1 linker histone seals the DNA to the core by binding with the DNA-enveloping “tails” of the H2A histones. H1 is also thought to pack the nucleosomes together. Nonhistone proteins bind to the linker DNA (~10 – 90 bp: much of that variation is across different organisms and

tissues) between the nucleosomes. Although these proteins comprise ~10% of chromatin, they have thousands of varieties and their specific functions are not well understood, although some may participate in forming loops in the fiber and others in unpacking nucleosomes for transcriptional access [2], [3].

In their recent review paper, van Holde and Zlatanova [1] summarize what is currently known about chromatin structure and what remains unresolved. New technologies and techniques have led to two different proposed structures: a one-start helix of nucleosomes, and a two-start or zig-zag helix (Fig. 1). However, some of the useful *in vitro* tricks (such as DNA cloning) tend to skew the results in favor of a regular structure by introducing uniform base repeats and constant linker lengths. As other research (detailed later in this paper) shows, there are good reasons why nucleosomes may need to be irregularly spaced along the DNA helix. No current techniques can map chromatin under true *in vivo* conditions, and trying to assign a regular structure to such chromatin may not be practical or realistic. Instead, native chromatin may likely be a dynamic structure that varies its configuration according to DNA linker lengths between nucleosomes, as depicted in the "random walk" structure shown in Fig. 2. The only real boundary conditions seem to be that the nucleosomes must be tightly compacted and facing outward, and that the fiber must have a diameter of approximately 30 nm. Assigning a regular structure to chromatin could still be important however; such a structure could be the minimal energy / maximum stability ideal against which generated structures could be RMS measured.

Finally, there is the question of the path shape of the 30 nm fiber itself. Even though interphase chromatin needs to be in a uncondensed state in order to facilitate transcription, space constraints in the nucleus require some form of compaction, such as the fiber looping back upon itself. Such higher-level structure could explain how some transcription regulatory factors can be thousands of bases upstream from the gene [1],[2].

The dynamic model

It would seem reasonable that one of the requirements for the structure of a section of chromatin fiber should be that any promoter regions are easily recognizable and accessible to regulatory proteins. The problem with this assumption is that evidence suggests that many promoters, when inactive, have their transcription factor binding sites buried in the nucleosomes. Chromatin remodeling proteins can make these sites accessible, but it's still an open question as to exactly how the sites are recognized in the first place.

In another review paper, van Holde and Zlatanova [4] detail the model of a dynamic chromatin which can occasionally expose its binding sites to freely diffusing nuclear proteins. Experiments and evidence indicate that one of the mechanisms for this involves DNA periodically unwrapping from the histone core (Fig 3). This behavior is not the result of the external ATP-dependent remodeling complexes but rather the result of inherent instability in the nucleosome (even though the core octamer does not unfold or change shape to cause or respond to this unwrapping). The amount of time the particular nucleosome stays open and / or the rate at which it opens and closes affects the odds of whether it will be seen and identified by one or more copies of a particular transcription factor.

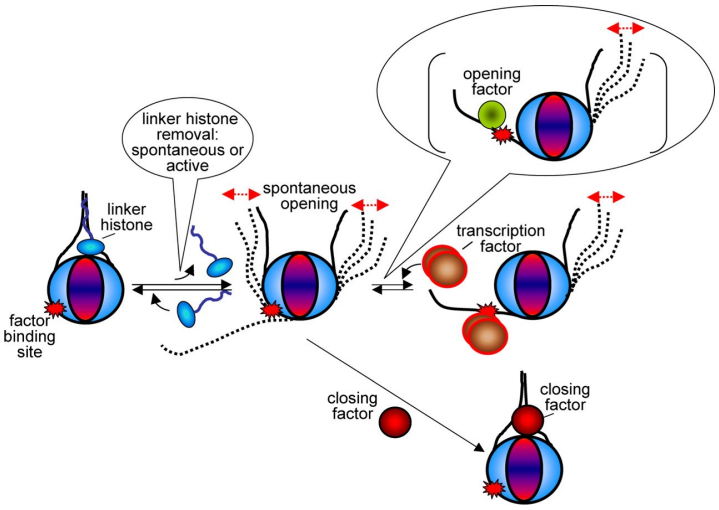

Fig. 3. Model of spontaneous opening /closing in nucleosome [4]

An earlier concept which also ties into this model is the epigenetic (outside the genome) "histone code". In this process, transcription and even identification of genes is determined not (at least directly) by the DNA sequence but by a complex chain of feedback targeted reactions centered around post-translational covalent modifications in the core histones. Modifications to the linker histone accompanied by the binding of the one or more of the many varieties of nonhistone proteins could also contribute to this "code". Furthermore, these proteins are somehow duplicated in their modified states along with the rest of chromatin and can carry epigenetic information for many generations. Much more experimental data (such as genome-wide microarray databases of recorded chromatin-associated factors and modifications) is needed before any attempts can be made to crack the code or computationally model/predict the process [5].

The model of dynamic chromatin effectively changes transcription from a deterministic process to a probabilistic sequence of events. At its most extreme, it could make chromatin structure and nucleosome positioning largely irrelevant to gene transcription, since in just a few minutes multiple copies of a transcription factor can theoretically scan the entire genome, all of which will be exposed for at least part of that time [4]. It seems unlikely, though, that a transcription factor would use such a brute-force method to find its binding site, and a constantly churning chromosome would seemingly use a lot of energy. A compromise view could be that the histone code allows only the promoter and enhancer sites to be periodically opened up. That however makes chromatin a passive agent which can be changed only from the outside, a doctrine at odds with the idea of dynamic chromatin [4]. A third option could be that the behavior of periodic nucleosome opening somehow occurs on a core wrap which has a certain kind of DNA sequence/property (such as a high energy state). That could provide one explanation of nucleosome positioning and linker DNA lengths. More importantly, it puts DNA back in charge of its own genes and provides a potential genetic bridge to the epigenetic histone code.

Genome-wide structural mapping

Technologies such as microarrays and high-throughput sequencing have been used to identify the coregulation patterns (and hence the transcriptional regulation) of a large number of genes on a given genome. Increasingly these techniques are being used to provide genome-wide data on chromosome structure. High-throughput sequencing can provide single-base resolution, but it's still relatively slow and expensive. Limited single-base resolution can be obtained in microarrays by exploiting the cytosine methylation of most CpG dinucleotides. Chromatin structure can be determined from the following genomics approaches:

Chromatin primary structure:

1. *Nucleosome positioning and occupancy*: Digestive enzymes can be used to remove linker DNA and show the nucleosome locations, while the microarray sampling shows if the locations are the same across cells (which in turn shows if the locations are determined by DNA sequence instead of "histone code" reactions). During transcription, histones get temporarily removed to free the DNA, and nucleosome depletion is a sign a region is undergoing transcription.
2. *Histone variant localization*: Alternate histone forms can also be markers of a region being actively transcribed.
3. *Histone modification*: Acetylation, methylation, and phosphorylation patterns on histones provide clues to the epigenetic code governing chromatin and transcription.

Chromatin secondary structure: Currently limited to simple measurements such as packing density (regions without genes tend to have higher density), but new techniques such as chromosome conformation capture could potentially create high-resolution detail [6].

With the collection of enough expression data, sequence motifs of nucleosomal DNA regions can be statistically inferred [7]. Without single-base resolution, however, this information cannot be used as a gold standard for finding new, unknown regions. The main benefit of genomic data is to corroborate and direct other chromatin modeling methods.

Current efforts to model and simulate chromatin

Because of the size and complexity of the problem, computationally modeling 30 nm chromatin fiber with all-atom methods such as molecular dynamics is currently out of the question. Such methods wouldn't make much sense anyway, because chromatin has many repeating structures and multiple levels of organization. Instead molecular subunits can be treated as single objects or entities that have either single point or aggregate surface properties. Many such simplifications could probably be used without any loss of overall physical accuracy, but apparently such a model is still too large and complex to attract any interest in tackling it – and perhaps, the code from current MD packages can't easily be adapted to this model and instead new software would have to be written. The models that seem to be currently used might instead be described as completely generalized and simplified to the point of being unrealistic. Still, these efforts are yielding some useful insights into chromatin properties and behavior.

Langowski & Heermann cover and implement a good cross-section of these methods in their recent review paper [8]. DNA is approximated as a flexible worm-like chain, with short molecular sections making up rigid segments connected by elastic joints. The properties are mostly described by conventional physics and constant values are provided from experimental data on DNA. The chain model is applied to chromatin fiber with less success; just as it's proving difficult to measure and map *in vivo* chromatin structure, the fiber seems to elude measurement of its stretching and bending properties as well.

The total energy of the DNA chain forms the basis for two different simulations. The first uses a Monte Carlo method to produce a collection of configurations that are at minimal free energy. The second uses Brownian Dynamics to compute a time-dependent conformation, which is then used to predict kinetics of supercoiling (nucleosome wrapping) and model the intramolecular reactions.

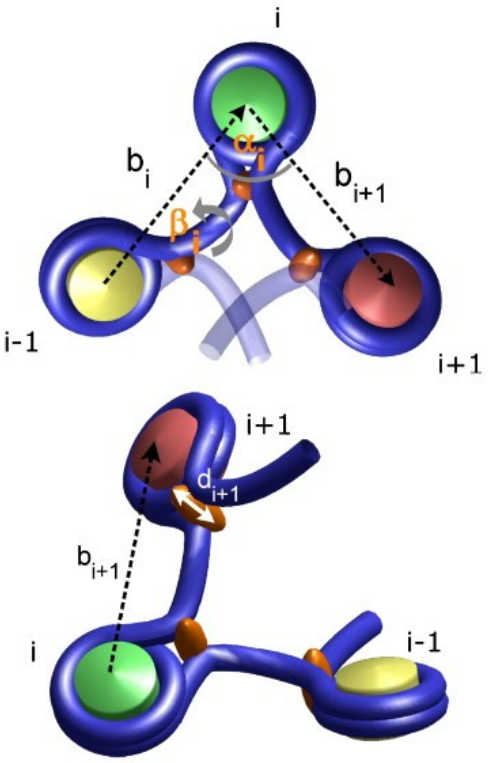

Fig. 4. Two-angle model (small ellipsoid is H1) [8]

The two-angle model of chromatin is part of a common computational method (Fig. 4) as well as a hypothesis for *in vivo* structure (see Fig. 2). Nucleosome core particles are modeled as flat disks (11 nm dia., 5mm thick) with an attractive potential between them, while linker histone H1 is responsible for the pitch and the stability of the nucleosome. The simulations show that the fiber is more resistant to stretching than bending and therefore tends to want to both tightly compact and form tight loops or hairpin structures, which could open up the fiber to nuclear proteins. Additional refinements are still needed dealing with the effect of salt concentrations on the histone tails.

Analyzing nucleosomal DNA

Since computational modeling and simulation of chromatin seems to only offer general answers at this point, perhaps the solution is to refocus at a more elemental level. Any structural prediction of chromatin must begin with the molecule that is the reason for its existance: the double-helix DNA strand. Overall, DNA is fairly rigid and brittle (or at least that was the thinking until recently). This is not a problem for prokaryotes, which can fit their entire genome into one gently arcing circle of DNA. The much larger genomes of eukaryotes, however, necessitate the compressed form of DNA found in chromatin, and the histone is the protein structure which evolved to enable the DNA strand to double back on itself. Fortunately (or not, since this makes things considerably more complicated) the properties of DNA are not constant along its length but rather depend on the specific type of base pair and its orientation, as well as on the interactions with surrounding proteins [9]. A variety of approaches can help determine the likelihood of a DNA segment being wrapped around a nucleosome core and provide details of the resulting conformation:

Using the molecular geometry of the helix and histone core: Getting the helix to bend around the histone core at first seems like a tall order, since the diameter of the octamer is only four times larger than the helical diameter itself and nearly all of the curvature is done with a DNA length of only 129 bp. However, recent studies and measurements of nucleosomal DNA are showing structural differences from the traditional models generated from *in vitro* oligonucleotides [9]. The conformation of a Watson-Crick base-pair or dinucleotide step (i.e., the mutual positions of base pairs) is defined by by three translations (rise, shift, and slide) and three rotations (twist, tilt, and roll) (Fig. 5) [10]. The geometric factors involved can be broken down into at least five areas:

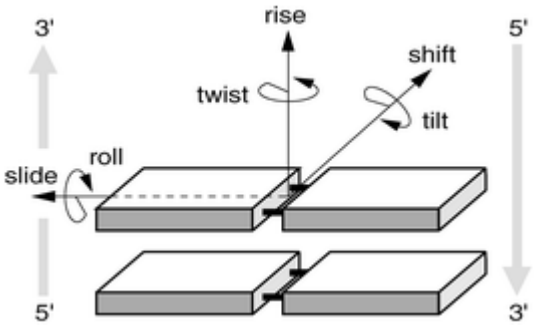

Fig. 5 [10]

1. *Steric clashes* – The maximum possible degree of curvature in a helical section is often limited by whether molecular shapes will collide. Most dinucleotide sequence combinations have a conflict in one or the other direction involving the outer ring of a purine. In the 5'-3' direction (either strand), a purine (A or G) followed by a pyrimidine (C or T) eliminates the ring-ring overlap that would normally prevent bending, making this combination the most potentially flexible [3].
2. *Helix twisting / untwisting* – At any point the helix can have a higher or lower turn density or twist angle than the standard model, and this can affect groove orientation and the direction of bending (see #4) [9].
3. *Helix stretching* – The helix can stretch as much as two base-pairs per nucleosome, shifting the base pair locations of expected histone contact, and even altering the overall twist angle. Furthermore, this stretching seems to occur more often than not, suggesting that any energetic costs of keeping DNA in a stretched state are more than outweighed by the energetic advantages of optimal nucleosome / base-pair positioning and binding. [11].
4. *Major groove / minor groove orientation and anisotropic bending* – The two helices of DNA are slightly offset relative to one another so that between the helices are major (wide) grooves alternated with minor (narrow) grooves. The major groove exposes one edge of a given base pair and the minor groove exposes the opposite edge; this has implications for binding or contact with the histone proteins [3]. The major groove has more room to bend, and is more likely to be facing the core during smooth bending. Minor-groove bending usually occurs as the result of kinking (see #5) at a single CA=TG base-pair step [9].

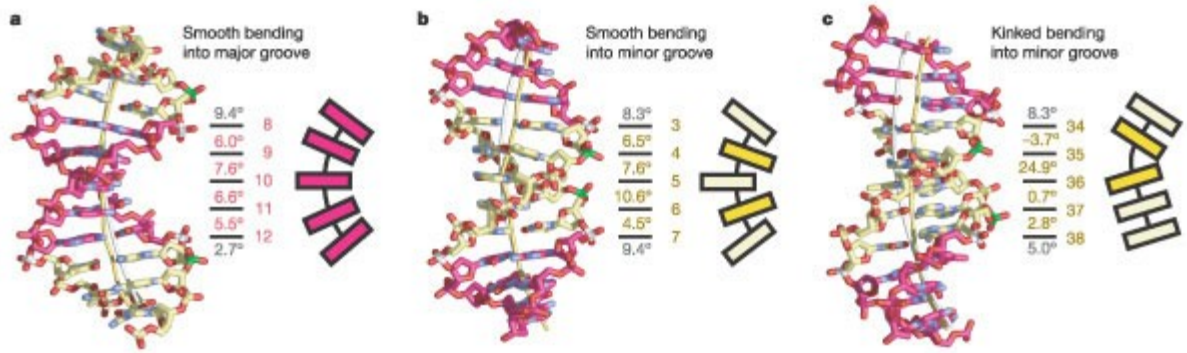

Fig. 6. Some types of bending for short sections of DNA [9]

5. *Kinks in the chain* – Bend a semi-rigid hollow tube too sharply and a sharp, inelastic kink will form. A similar phenomenon can occur in nucleosomal DNA, and the distorted forms may actually attract binding of certain nuclear proteins. Kinking is essentially a dislocation resulting from an extreme roll angle [11].

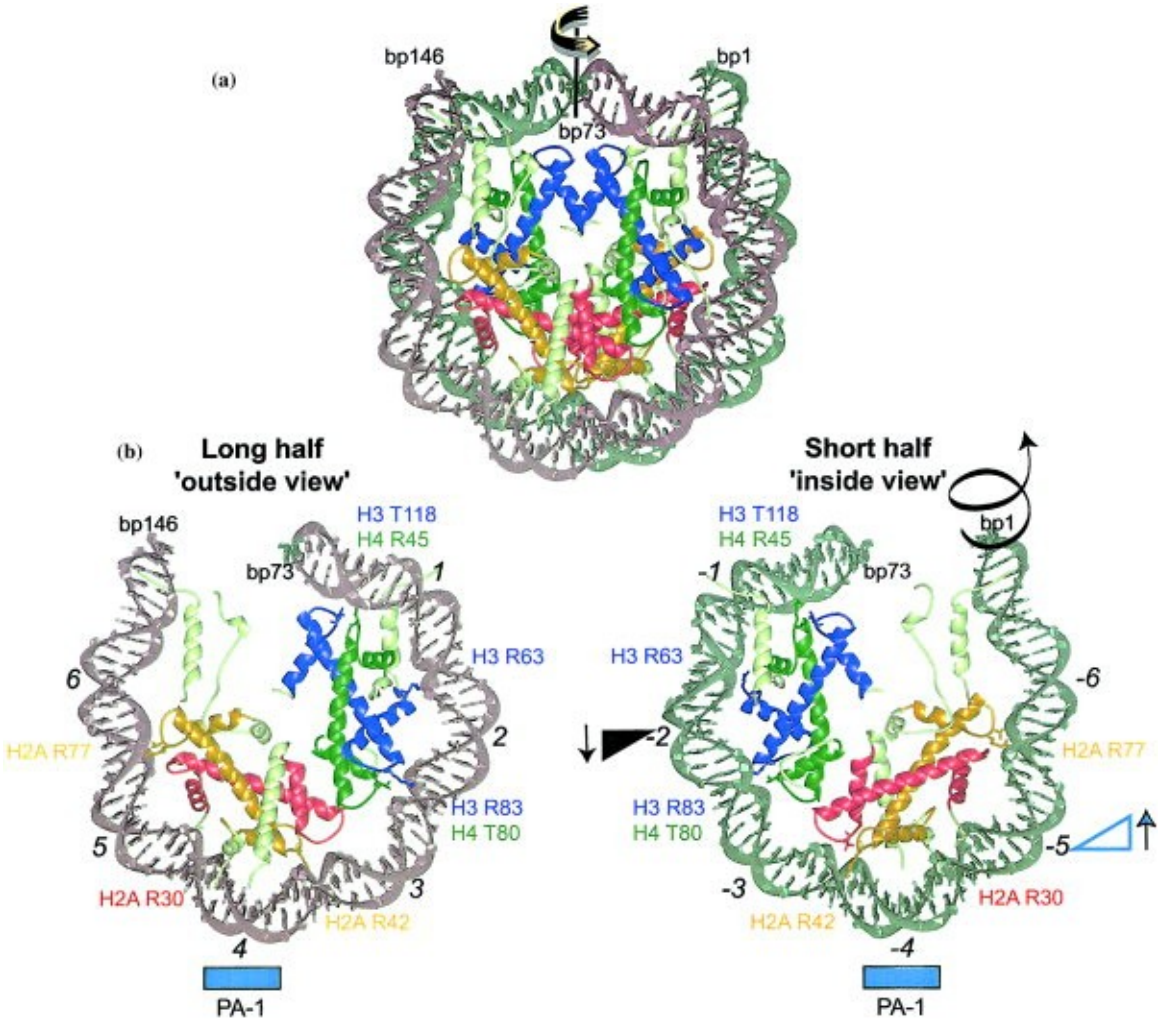

Fig. 7. Nucleosome core particle with the DNA/histone binding sites shown [12]

Identifying nucleosome core particle constraints and nucleotide binding sites: The curvature of the wrapped helix is dictated not only by the general shape of the core but also by histone secondary structures. In extreme cases (such as with a short helical section in contact with two different secondary structures), there may be steric clashes with the core that require the DNA to take on a contorted (kinked) shape [11]. Even more significant are the specific locations on the histone proteins where DNA either must bind or prefers to bind (see above) [12].

Experimental / statistical sampling: If a statistically significant sample of sequenced nucleosomal DNAs can be obtained, probabilistic methods can be used to find motifs that prefer to be in specific nucleosomal locations. These motifs can in turn be used to find nucleosomal DNA candidates. Segal et al. use such an approach with yeast and combine it with parts of the geometric approach, using probability distributions for dinucleotides (to capture bending preferences) and applying dynamic programming to find all sterically allowed sequence combinations. The end results are nucleosome location predictions that are correct about half the time – a good start, but the authors admit that three-dimensional energetics need to included [13].

Microarrays, data mining and analysis of the annotated genome: (see previous section on genome-wide structural mapping)

Measuring / simulating DNA flexing: The less force that is required to wrap a section of DNA around a core particle, the more stable that sequence is likely to be for nucleosome positioning. The flex properties of nucleosomic DNA have been extensively studied, with many of the efforts summarized in a 2004 review paper by Beveridge et al. Experimental techniques have so far not been successful in mapping the structural details of a short section of bending DNA, so molecular dynamics simulations have been used to fill in the gaps [14].

Simulated energetics of DNA with its surrounding proteins: The flexing properties of DNA probably can't be considered in isolation; DNA likely becomes more flexible/stable in response to binding / attraction to proteins – this is likely what happens in the case of kinks [11]. Nor is it safe to assume that the effects of these chemical reactions will be the same for

all rotations/translations, or even that the nucleosome will remain static and stable. The question of the optimal way to wrap a sequence of DNA around a histone core then becomes part of a broader molecular dynamics problem.

Most of the current efforts to do MD simulations of nucleosomal DNA seem to take the same approach as that for proteins: calculate all of the molecular interactions at each time step until the DNA settles upon a new conformation. Among the most recent efforts on this front is the work by Ruscio and Onufriev [15], who use MD to confirm the x-ray structure of the nucleosomal DNA mapped by Richmond and Davey [9]. This approach has been useful for gaining greater insight into the general behavior of nucleosomal DNA and validating hypotheses about its structure, but isn't practical to attempt for every prospective sequence.

Conformation of nucleosomal DNA: A proposed scoring method

Given some ~147 bp segment, is it possible to systematically calculate / assign a score or measurement of that segment wrapping around a histone core at specific base pair locations? The basic question is this: Given a sequence of DNA that has been deformed into a specific core-required conformation, what is its potential energy with respect to the same sequence's reference structure? Potential energy in this case is defined as the all-atom sum of five energies: bond stretching, bond angle bending, bond twisting, electrostatic forces and VDW interactions [16]. The problem here is that the number of potential conformations is nearly endless, which is why computationally-intensive MD simulations are used in an attempt to settle upon something. But what if the problem space can be limited in size? After all, this isn't as open-ended a problem as protein folding; DNA is a repeating structure that varies only by its bases and it has to conform around a fixed (at least prior to transcription [4]) core structure with constant properties, an electrostatically uniform surface [15], and fixed locations required for DNA binding. Furthermore, a DNA segment structure can be defined by the sum of the dinucleotide conformations, potentially dividing up the problem even more.

An alternative approach to this problem could be as follows: Divide up the DNA segment into ~147 overlapping, equal size dinucleotide sections and then sum the potential energies of each to get the total energy value. This of course would not be the true potential energy, but then it doesn't need to be; it just needs to be a metric by which DNA segments can be compared and ranked – a pseudo-energy if you will. Even computing an all-atom energy sum of a single dinucleotide section is not a trivial task if it has to be performed a large number of times, however. Instead, can all of the conformational and energetic possibilities for any section be constrained enough to precompute the energies and store the results in a lookup table or database? As outlined in the molecular geometry section above, a dinucleotide section has six degrees of freedom (three rotations and three translations – see Fig. 4) and 10 unique dinucleotide combinations. Even with a somewhat crude approximation of around 10 step values per translation/rotation (10^6), the resulting database/table would have several million entries – not small by any means, but doable on a personal computer.

The relative energy value of a dinucleotide section likely won't be decided by just these parameters, however, and the number of database entries would be multiplied by any additional parameters. These could include the orientation of the helical section (whether the major groove or minor groove is facing the core surface) or a reaction at a binding location – any factor which does not equally affect the energies of all conformational dinucleotide combinations. For the factors which do apply to all energies (such as core electrostatics), a small energy surface could be introduced into each of the energy computations. In the case of DNA binding to a core protein, the particular type of bound side chain could be combined into the all-atom computation for that dinucleotide conformation, which would also be adjusted for experimentally determined changes to bound DNA. Overall these additional parameters should not increase the database size to an unmanageable level; the main problem will be the difficulty and complexity in figuring out how to accurately compute these relative energies.

A proposed algorithm to find best nucleosomal DNA conformations

The method in the previous section provides a way to build a database which can quickly give an energy value for a 147 bp DNA sequence of a specific nucleotide-wrapping conformation. Now there needs to be way to sort through all of the possible conformations and find the best (lowest energy) one. This appears to be a problem suited for a Hidden Markov Model and some modified version of the Viterbi algorithm [17].

<u>HMM definition</u>
Alphabet $\sum$ = {set of 10 unique dinucleotide combinations}
Set of states Q = all possible predefined conformational and energetic combinations (rise, shift, slide, twist, tilt, and roll, major/minor groove, bound/unbound, etc.)
Transition probability: 1 (since the same sequence of dinucleotides will be evaluated each time)
Emission probabilities: The potential energy score, expressed as a proportion of the maximum score (which is the inverse of the minimum energy) in the dinucleotide conformation database.

<u>Viterbi-based algorithm</u>
Iteration (row position i and column position j):

- Parse $V_j(i) = e_j(x_i)(\max_k V_k(i-1))$ where e is the (energy-related) emission probability associated with a possible conformation state – only a relatively small number of states may be possible at a given dinucleotide step (particularly

if base is at a binding location), so each column can be stored in its own vector to save memory. Disqualified states could include conformations that cause steric clashes, break Watson-Crick bonds, or separate too far from core surface
- Back pointer $Ptr_j(i) = argmax_k V_k(i-1)$ also contains reference to DNA conformation up to that point (and therefore the current geometric position)

Termination: Is reached when dinucleotide is at the geometric end position of the core wrapping (which may occur 1-2 bases early depending on DNA stretching)
Traceback: Produces 145-147 bp DNA structure with optimal score and conformation

The next step is repeat this algorithm for each starting base pair step of the sample sequence in order to find a set of candidate locations for nucleosomes. Again, the lowest scores or energies are the best, but the nucleosomes also have to be appropriately and more or less evenly spaced based on known linker lengths. As such, it may not be necessary to evaluate every overlapping 147 bp subsegment; a stochastic sampling method could be used instead.

A proposed algorithm for nucleosome-DNA positioning and chromatin structure prediction

The two competing 30 nm chromatin regular structural models (shown in the first section) may be simplistic and not very realistic for *in vivo* chromatin, but they probably represent a low energy / high stability ideal. In other words, they likely represent the structure chromatin might take if the positioning of DNA along the nucleosomes were not an issue. With this assumption, predicting a structure of chromatin fiber from a specific DNA sequence becomes a question of how to maintain close packing and an approximation of the ideal structure while at the same time maintaining close to optimal DNA positioning around the nucleosomes. This structure could be defined as an optimal alignment of these two competing interests, and could be solved using dynamic programming. For the simplest model (below), the specific structure doesn't matter – just the linker lengths and the histone core locations on the DNA sequence.

<u>HMM definition</u>
Alphabet $\sum$ = {range of energy values for nucleosome}
Set of states Q = the range of ~20 bp possible variation in linker lengths
Transition probability: 1 (since the there will only be a transition to the next sequence)
Start probabilities: ~1/200 - probability of start location from base index 0 to 200 (the approximate length of nucleosomal DNA plus linker DNA)
Emission probabilities: matrix score of previous linker length and current nucleosome energy that would have to be set up either based on experiment, iteratively based on the results, or computationally based on MD simulations.

<u>Viterbi-based algorithm</u>
Iteration (row position i and column position j):
- Parse $V_j(i) = e_j(x_i)(max_k V_k(i-1))$ where e is the probability of having a given nucleosome energy following a given linker length
- Back pointer $Ptr_j(i) = argmax_k V_k(i-1)$ also contains reference to absolute position of previous nucleosome so that energy of current nucleosome will be known (based on DP from previous algorithm)

Termination: Is reached when there is no longer room to add new nucleosomes
Traceback: Produces a DNA sequence with indexed nucleosome locations

The main shortcoming of this algorithm is that it probably can't be used to predict tertiary structures such as loops and bends in the fiber, which would be helpful in finding regulatory regions that are thousands of bases upstream of the transcription site. Predicting these structures is probably going to require better knowledge of the epigenetic role of factors such as nucleoplasmin chaperone protein [3] that influence chromatin folding.

Verifying, refining, and interpreting a predicted chromatin structure

Ideally, the finished product of these previous computational steps should be a 30nm chromatin fiber structure derived from a DNA sequence input, and this structure should represent the most stable or lowest energy conformation, particularly for the nucleosomes. According the the dynamic chromatin model however, a completely stable structure is neither possible nor desirable. A search for promoter regions then means (somewhat counter-intuitively) one should look for the nucleosomes with the *highest* energy, since these will be the ones most likely to make their DNA available to transcription factors and other nuclear proteins.

Before such an approach can be used be used to find or predict new genetic regions, a predicted structure of course must be based on and compared against known regions. This information can then be used to adjust the parameters, weights, and constants used in the algorithms. Known regions can include not just sequence motifs but also 3D sites which have been determined by experiment (which would give the best verification of a 3D structure). Of particular interest should be nuclear proteins which have been shown to bind to particular DNA conformations – the conformation of nucleosomal DNA could be the key to learning what type of gene a site codes for and how it functions.

Conclusions

One of the main obstacles to successfully and accurately modeling chromatin (other than computational size and complexity) is the current lack of a clear operational understanding of its details and functions, as well as where it fits in the overall process of gene expression. Better experimental methods are clearly needed, but at the same time computational methods can direct the experimental process and fill in some of the knowledge gaps. More progress on both fronts is needed before the genome can be completely decoded.

One thing that is becoming more apparent however, is that the genome is not the center of a deterministic process but rather a probabilistic one. Molecular structures are just another probabilistic component in this system, and by themselves they won't provide the key to understanding gene function. Accurate modeling of them, however, is crucial to the next step of simulating and mapping all of the interactions involved in the path from gene to protein.

References

1: van Holde, Ken, Zlatanova, Jordanka, Chromatin fiber structure: Where is the problem now?, 2007
2: Geoffrey M. Cooper, Robert E. Hausman, The Cell A Molecular Approach, 2007
3: Voet & Voet, Biochemistry, 2004
4: van Holde K, Zlatanova J., Scanning chromatin: a new paradigm?, 2006
5: Mellor, Jane, Dynamic nucleosomes and gene transcription, 2006
6: Oliver J. Rando, Chromatin structure in the genomics era,
7: Fatih Ozsolak, Jun S Song, X Shirley Liu & David E Fisher, High-throughput mapping of the chromatin structure of human promoters, 2007
8: J¨org Langowski, Dieter W. Heermann, Computational modeling of the chromatin fiber, 2007
9: Richmond, T. J. & Davey, C. A., The structure of DNA in the nucleosome core, 2003
10: Alexander A. Voityuk, Khatcharin Siriwong and Notker Rosch, Charge transfer in DNA. Sensitivity of electronic couplings to conformational changes, 2001
11: Michelle S. Ong, Timothy J. Richmond and Curt A. Davey, DNA Stretching and Extreme Kinking in the Nucleosome Core, 2007
12: Uma M. Muthurajan, Young-Jun Park, R. S. Edayathumangalam, Robert K. Suto, Srinivas Chakravarthy, Pamela N. Dyer, Karolin Luger, Structure and Dynamics of Nucleosomal DNA, 2003
13: Eran Segal, Yvonne Fondufe-Mittendorf, Lingyi Chen, AnnChristine Tha°stro¨m, Yair Field, Irene K. Moore,Ji-Ping Z. Wang & Jonathan Widom, A genomic code for nucleosome positioning, 2006
14: D. L. Beveridge, Surjit B. Dixit, Gabriela Barreiro, Kelly M. Thayer, Molecular Dynamics Simulations of DNA Curvature and Flexibility: Helix Phasing and Premelting, 2004
15: Jory Z. Ruscio and Alexey Onufriev, A Computational Study of Nucleosomal DNA Flexibility, 2006
16: Allen, Michael P., Introduction to Molecular Dynamics Simulation, 2004
17: R. Durbin, S. Eddy, A. Krogh, G. Mitchison, Biological sequence analysis Probabilistic models of proteins and nucleic acids, 1998