Histone Structure

II. A model which places DNA in the N-terminal region of the octamer

By Ken Biegeleisen, M.D., Ph.D.

Transcript of narration

SLIDE 1
Title Page
(no audio)

SLIDE 2
Acknowledgments
(no audio)

SLIDE 3
(Flash Player Controls)

Before we proceed, a word about the screen. If you haven't downloaded this PowerPoint presentation to your computer, then you're still online, and you're looking at a Flash export of the PowerPoint, courtesy of a lovely little company called iSpring.

There are a number of toggle switches in the iSpring Flash Player you're looking at now, but I only know about three of them.

First of all, if you click this little thing in the upper right-hand corner, you'll toggle back-and-forth to-and-from full-screen mode. If you want an even larger image, you can get rid of the slide strip on the left by clicking this thing in the lower right. If you do this, you'll still be able to navigate, at least from section-to-section, because I've inserted a "Table of Contents" hyperlink on nearly every slide in this presentation.

And finally, you can adjust the audio volume with this volume button down here. There are several other controls on the Flash Player, but I have no idea what they do. If you've got lots of time on your hands, you can experiment with them.

SLIDE 4
Table of Contents
(No audio)

SLIDE 5
Title slide
(No audio)

Introduction

SLIDE 6

Hello. I'm Dr. Ken Biegeleisen, and I'm here so show you a modified structure for the nucleosome. The modification shall involve only the N-termini of the 8 subunits of the histone octamer; the rest of the structure will be entirely preserved. Well, almost entirely – a few atoms will have to be moved a few angstroms each, but even an experienced histone scientist will not readily notice the difference. For the most part, I have totally respected the octamer core, and modified only the N-termini.

This slide shows only the octamer core, excluding the N-termini. If you're not quite familiar with this structure, please go back and look at Part I of this series, which reviews our current concepts of nucleosome and histone structure. The following 1-minute review is all we're going to have time for here, and, by itself, it will not be sufficient background for you to understand the arguments which follow, unless you're a histone scientist already.

As this slide shows, the histone core structure consists of 4 subunit types. These have the following names, which shall always be given in upper case letters: H3, H4, H2A and H2B. There are two instances of each subunit type in the octamer, and they are distinguished in two different ways. In the first place, the 8 chains are uniquely colored, and the color scheme shown here is carefully maintained in nearly every slide shown on this web site. Secondly, the subunits are *uniquely* identified by chain names, which we give here in lower case letters, which will hopefully be of at least *some* help in distinguishing them from the upper case subunit type names. The lower case chain letters are a-h, following the convention established by Luger *et al* in their Protein Data Bank entry, which reference we shall be giving shortly.

Each of the two groupings of the 4 subunit types can be construed to be what I call a "logical tetramer", meaning a collection of 4 subunits which do not coalesce spontaneously *in vitro*, but which do appear to constitute a logical structure when in the octamer setting. Here's the first:

SLIDE 7

This is the a-b-c-d tetramer, that is H3[a], H4[b], H2A[c] and H2B[d]...

SLIDE 8

And this is the e-f-g-h tetramer, that is, H3[e], H4[f], H2A[g] and H2B[h]. If you put these two logical tetramers together, you get the complete octamer core:

SLIDE 9

And here it is again. Next, let's look at the entire histone structure, by adding back the N- and C-termini.

SLIDE 10

This is the entire histone octamer structure.

Here are the subunit names and chain letters again.

Here are labels identifying the 8 unordered C-terminal strands, which are discussed at more length in Part I of this slide series.

And last, but not at all least, are the N-terminal strands. Note how unruly they look. We're going to try to establish some order here.

SLIDE 11

Since the publication of Luger *et al's* high-resolution crystal structure of the nucleosome in 1997, the reference for which appears below*, the N-termini of the 8 histone subunits have been considered to be essentially random coils. This is startling, since the bulk of the positively-charged basic amino acid residues in histone, which I say ***must*** be presumed to be involved in the electrostatic binding of DNA, are located in those same N-termini."

*Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J.
Crystal structure of the nucleosome core particle at 2.8 A resolution.
Nature **389:** 251-260, 1997 . **PubMed:** 9305837 . **DOI:** 10.1038/38444

SLIDE 12

Please take a look again at those unstructured N-terminal arms. In the published histone structure, those arms are portrayed as meandering aimlessly in the peri-nucleosomal space. Their lengths are variable because this structure was determined by x-ray crystallography, the preparatory process for which results in variable degradation of the 8 N-termini, which therefore appear to be random coils of unpredictable length.

And yet these unstructured N-termini are unequivocally the locations of the highest concentrations of positively-charged basic amino acid residues to be found anywhere in the histone octamer. Odd, is it not? Everyone understands that DNA, in the cell nucleus, seeks out positively-charged proteins with which to interact, yet in this structure, the DNA is commonly held to have no interaction whatsoever, with the part of the protein where the highest density of positive charges lies. When we add the DNA...

SLIDE 13

...we get this picture, now found in almost every textbook of molecular biology on earth. You can download the virtual structure file for this, by clicking the Protein Data Bank link at the top of the screen.

You cannot see it from this flat, 2-dimensional perspective, but the DNA is *not* interacting with the unstructured N-terminal arms of the octamer, it's merely obscuring our view of them.

As we showed in exhausting detail in Part I of this series, this, our current concept of nucleosome structure, calls for the DNA to lie in a so-called "Superhelical Ramp" or "Superhelical Groove", as it's also called; a region of the octamer which is, in fact, well-endowed with positively-charged basic residues to hold DNA. But... this Superhelical Ramp has substantially *fewer* basic residues than the N-termini, and, perhaps even more importantly, the charge-charge *alignment* between DNA phosphate groups and basic amino acid residue side-chains, in the superhelical ramp, is surprisingly poor. This was also discussed in great detail in Part I of this series, and if you haven't seen that, I again strongly recommend that you do so before proceeding. For now, however, let's look at the ***number*** of potentially-DNA-binding salt bridges in the N-termini, vs. the number in the Superhelical Ramp:

SLIDE 14

Here's a table showing the positive charge counts. First, some definitions. The superhelical ramp consists of Helix I of each of the 8 histone subunits, plus the so-called β-bridges which form between β-strand 1 of each subunit & β-strand 2 of its adjacent partner subunit.

Percentage of DNA phosphate groups bound by Lys/Arg residues

Structure	DNA bp	Phos. Groups[1]	Total aa's	Basic aa's[2]	% Phos. bound
Superhelical ramp[3]	146	292	212	60	21%
N-termini, octamer[4]	124	248	202	80	32%

1. This is merely [bp] x 2.
2. Arg and Lys only; His not counted.
3. The spiral path of the [Helix I]+[β-strand 1]+[β-strand 2] regions contributed by each of the 8 subunits.
4. The sum of the contributions of the 8 octamer N-termini, proximal to Helix I.

TEXT BOX (NOT NARRATIVE): Conclusions: The N-termini, collectively, have substantially more basic residues than the superhelical groove, and they have the theoretical capacity to bind a commensurately greater percentage of DNA phosphate groups. Furthermore, the quality of the binding (with respect to length of the salt bridges) is far superior in the N-termini (data to be presented later).

SLIDE 15

We'd better take a quick time-out and look back at these structures. Because I myself find it hard to visualize these things, even after a year and a half of living with them night and day, I presume a quick review would be useful to you also.

Here, as an illustrative example, is a quadrant of the histone core structure containing subunits H3 & H4. These are the H3 [a] chain, which is colored white in *all* of our slides, and the H4 [b] chain, which we always color red. The small arrows show the Helix I α-helices of these subunits, and the two β-bridges formed by their respective β-strands. The large yellow arrow indicates the direction that DNA travels around this quadrant of the Superhelical Ramp. It passes over Helix I and its associated β-bridges, but steers clear of Helices II and III, which are well-recessed into the octamer core.

SLIDE 16

Back to our chart: Now that we've refreshed our memories concerning the structure of the superhelical ramp, let's compare its DNA-binding power to that of the N-terminal region of the histone octamer. This latter region is defined simply as everything N-ward of Helix I; namely those aimlessly-meandering 8 strands we just looked at a few minutes ago.

Let's start our comparison by looking at the superhelical ramp: This is said to bind about 146 bps of DNA, which means twice that many phosphate groups, *i.e.*, 292 phosphate groups in all. There are about 212 amino acids in the superhelical ramp structures, of which 60 are basic.

By the way, if you watched Part I of this series, you may recall I gave this number as "62" there, but only "60" here. That's because we're not counting His here. Histidine is rather short compared to Lys and Arg, and furthermore, at physiological pH its positive charge is negligible.

As for the longer and more powerfully-charged Lys and Arg residues in the Superhelical ramp, their numbers are sufficient to neutralize only 21% of the phosphate group negative charges in DNA. Moreover, as we showed in Part I, the alignment between these positive and negative charges is surprisingly poor.

SLIDE 17

Now let's look at the comparable statistics for the N-terminal region of the octamer. In the new histone model I'm going to show you today, there are slightly fewer DNA bps, namely 124, than in the current model, and, of course, two times that number of phosphate groups, *i.e.*, 248 phosphate groups. Note, however, that of the total of 202 amino acid residues in the eight N-termini, *basic* amino acid residues account for fully *80* of them, meaning that 32% of the phosphate negative charges are neutralized.

That's a lot more than the cognate 21% figure for the superhelical ramp structure and, moreover, as we shall see shortly, the *quality* of the salt bridges in our new N-terminal DNA-binding model is going to be extremely high, each one being nearly perfect, at about 3 angstroms. You'd be hard pressed to find any large number of 3-angstrom salt bridges in the superhelical ramp; in my 2006 protamine slide show I glibly asserted that there were *none*, which, as it turns out, is not entirely correct, but the number is, nevertheless, rather small. Most of the so-called 'salt bridges' in the superhelical ramp are in the 5-10 Å range, hardly a 'salt bridge' at all.

Ask yourselves this anthropomorphic question: "What reason would DNA have, to be entropically-content to dwell in the Superhelical Ramp, where only 21% of its negative charges are neutralized, and only partially so at best, when it can dwell instead in the N-termini of the octamer, where *32%* of its negative charges can be neutralized, and *perfectly* so?

SLIDE 18

My answer to that question is: "No reason". I feel quite confident that, under normal circumstances, DNA does indeed dwell in the N-terminal region of the histone octamer. Although I have not at all ruled out, in my own mind, that DNA might ***occasionally*** be found

coiled about the superhelical ramp, I believe that that would only be under some set of special and relatively extraordinary conditions.

For me, then, the next step is to deduce the structure of DNA in the N-terminal histone regions, and its manner of binding to the basic residues there.

SLIDE 19
Title slide
(No audio)

<u>Preview of structure</u>
<u>(without explanation of its origin)</u>

SLIDE 20

This is a summary of the steps necessary to create the new model. The first step, determination of octamer sites of binding of DNA, is simply a matter of primary structure. We are going to assume that the existence of highly-basic regions of nucleoprotein implies, in the absence of evidence to the contrary, that those regions bind DNA. The second step, a search for key protamine-DNA spacings, will be based upon the presumptions that (a) the protamine model is correct, and that (b) the particular spacings between DNA-binding β-strands in protamine will be found in histone as well. With respect to the question of orientation of DNA relative to the histone octamer core, however, there was no clue up front, and I count it as good fortune that I came upon a structure I refer to as the “Histone Equatorial Plane” which provided a logical answer to that question. The next step, correction of the length mismatch between H3 and the other histone subunits, proved to be the most difficult problem of all, and the last to be solved.

The final three steps, creation of protein templates, DNA templates, and rotamers, were fairly routine modeling chores.

We’re going to proceed backwards from here on in. That is, we’ll start by showing the finished model, then go back and explain how it was derived.

SLIDE 21

This is an axial view of the structure we shall propose. Because this is also the view of the octamer core shown in the Luger *et al* Protein Databank entry (and that, again, is accession #1aoi), this is therefore a view from which many deliberations will begin, wherefore I have taken to referring to this particular view as “root position” (which is actually a term I have borrowed from music). In part I of this slide series, I generally referred to this same position as being a "frontal view", but in the current presentation, I have come to prefer the non-directional term "root position", because this view will generally be treated here as more of a "top view" than a "frontal view".

The core structure, taken directly from the Luger *et al* PDB file, has been left almost entirely untouched, but the eight N-terminal strands, formerly unstructured, are, in this modified model, *highly*-structured. No longer random coils, here, each one is now a β-strand, potentially

able to form a β-sheet with a neighboring nucleosome. Each β-strand complexes with DNA through charge interactions between the positively-charged basic Lys/Arg residues of the protein, and the negatively-charged phosphate groups of DNA.

The problem in modeling the DNA is exactly the same one we had with protamine: there is simply no way to model the N-termini as DNA-binding domains if we assume the Watson-Crick structure pertains to the human somatic cell nucleus. If, however, we assume the structure for DNA that almost surely exists in the human *sperm* cell nucleus, where the only significant DNA-binding protein is protamine, then the task here, of moving the DNA out of the superhelical ramp, and into the N-terminal region, is reduced from one of unimaginable difficulty, to one of relative simplicity, requiring only the ironing out of a few details. This slide shows the final result. The remainder of this presentation will merely show how this structure was derived.

I should say at the outset that this is not a perfect model, but it's as far as I have been able to take it, in accordance with the limitations placed upon us by our current concepts of the histone core structure.

Let us begin to look at some of the details. The N-termini are those curved structures, one of which is now indicated by an arrow. Note that there are two such β-strands in each of the four quadrants of the nucleosome; 8 in all; corresponding to the 8 subunits of the histone octamer.

The DNA is the "X"-shaped figures, between the paired β-strands. This species of tetraplex structure is described in great detail in the accompanying slide show entitled "The Probable Structure of the Protamine-DNA Complex", right here on this web site, but we shall nevertheless review its basic features again here, in a few moments, for the benefit of those who can't, or *won't* watch the protamine presentation.

The picture shown in the current slide is, obviously, a composite picture. The histone core is portrayed only as ribbons, because the atomic details of the core were fully dealt with in Part I, and need not be repeated here. The N-terminal β-strands and their associated DNA, however, are portrayed as atomic models, because that will reveal the details which will be necessary, if we are to *comprehend* this structure.

The four quadrants of the model are shown in what is called "non-perspective" view, courtesy of Schrödinger Maestro, a virtual modeling program which has the ability to portray vertical columns as if they were cross-sections through the middle of the column (and you'll see how this *non*-perspective view differs from *perspective* view a few slides down).

Later on I'll show you exactly and precisely how this structure was derived. For now, however, let's just look at its general features, so you'll have at least some preliminary orientation for the discussion that follows. Let's start by isolating one of the four β-strand/DNA complexes, and rotating that one, so we can comprehend what it might look like in 3 dimensions. The corner quadrant which is labeled here is the dwelling place of histone subunits H2A and H2B, specifically the [c] and [d] chains:

SLIDE 22

First, we remove everything except the N-termini of the H2A and H2B subunits, and their associated DNA, and move them to the center of the screen.

SLIDE 23

Now let's rotate it to a longitudinal view:

SLIDE 24

[Pseudo-movie of tilt from axial-to-longitudinal]
(No audio)

SLIDE 25

Now we have a longitudinal view showing the complete H2A-H2B protein-DNA complex. With respect to the protein, we show only the N-terminal β-strands, having removed not only the entire nucleosome core structure, but also the stands connecting to it. H2A extends to the right, and H2B to the left. As we shall see in a moment, they are on opposite sides of the DNA. The reason H2B is so much longer is that it has a peculiar 10-amino-acid N-terminal sequence, which is almost 50% proline, that is not found in the other histone subunits. We'll be discussing that presently.

When we rotate this structure ...

SLIDE 26

...you'll see that the two β-strands are on opposite sides of the DNA.

If you're experienced with virtual molecular models, you'll be able to see the lysine and arginine R-groups, as well as the 4 proline residues at the N-terminus of H2B on the left.

Note the little x-y axis I've added in the lower left-hand corner. As I've mentioned in several other places on this web page, the atoms of these rotating models do not shrink in size when they move into the background, which will frequently give rise to a serious perspective ambiguity, manifested by the illusion that the model is stopping and reversing its direction of rotation every so often. Well, that's not really happening, but your eye will probably tell you that it is, wherefore I have added the x-y axis to remind your eye that the correct direction of rotation in this model is the positive-x sense. If your eye fools you into thinking otherwise, the only problem will be that the directions of helical windings of structures will appear reversed when the eye is fooled into thinking that the direction of rotation has reversed, *i.e.*, right-handed helices suddenly appear to have become left-handed, etc. So in slides where comprehending the direction of helical winding is of paramount importance, this problem can be *very* serious, but in most other rotating model slides, the problem, if it arises at all, may not be so serious.

Whenever your eye perceives the direction of rotation to be contrary to the little coordinate axis at the bottom, the only thing you can do is to just keep staring until the perceived direction of rotation corrects itself, which will eventually happen. Blinking once or twice might encourage that.

By the way, in case you're wondering, the "bare space" in the middle of the DNA strand, where there appears to be no protein, is actually the place where these straight N-terminal β-strands connect to the octamer core, but we've removed the connecting strands to make the view more clear.

SLIDE 27
Title slide
(No audio)

How does the new histone structure differ from the currently-accepted model?

SLIDE 28

How does the new histone structure differ from the currently-accepted model? The answer is "Not very much". To understand how the new structure was derived from the old, let's put them side-by-side and look at the N-termini, because everything else is about the same in both. The first thing to note is that in our proposed new structure, on the right, the eight N-termini of the 8 subunits, while not particularly orderly-looking in this view, nevertheless all appear to be about the same length. In the "currently-accepted structure" on the left, however, the N-termini seem to vary considerably in length. That's because they do vary – in the currently-accepted structure, every one of the N-termini was, to a greater or lesser extent, degraded by the processes of isolation and purification, *in vitro* reassembly, and crystallization. The longest-looking N-terminal strand in the crystal structure is the H3 strand, chain [e], which is colored yellow. It really stands out, does it not? Let's compare the N-termini of H3[e] in the two models:

SLIDE 29

Let's remove everything except H3[e]. I think I should emphasize that we're looking here at the entirety of these H3 subunits; the N-termini as well as the complete subunit bodies, isolated from the octamer core. In this position, the differences between the old and new model are not particularly striking. In order to fully-comprehend the differences, we must rotate them both.

SLIDE 30

Let's move them into a single column, so we can better compare them...

SLIDE 31

They still look about the same in this view, but we shall see that they have one major difference, namely that the N-terminus of the new structure is orderly, and that of the current structure is random.

SLIDE 32

We now rotate these structures, in the positive-Y sense, as indicated by the x-y coordinate axis in the lower left hand corner. You should see the long N-terminus of the New Structure sweeping *in front* of the plane of the screen. If, perchance, the long N-terminus of the New

Structure seemed to you to be going *behind* the plane of the screen, then your eyes are playing a perspective trick on you, and, if the spirit seizes you, you should perhaps run the slide again; hopefully your eye will see it right the second time.

When we rotate 90º thusly, the currently-accepted crystal structure, on top of the slide, *still* looks unordered, but we see now that our new structure, on the bottom, is *highly*-ordered, having the perfectly-straight β-sheet we saw previously, which was illustrated in the atomic model shown in the slides above. This is essentially the only difference between the old model and the new one – we've remodeled the N-termini.

By the way, you might have also noticed that our N-terminus includes a short, additional α-helix not found in the crystal structure. I refer to this as "Helix Sub-Zero", for reasons I'll give later. By way of preview, I'll say that this additional helix solved a length mismatch problem which threatened to destroy the entire modeling project.

SLIDE 33
Title slide
(No audio)

H3[e] atomic model – rotation

SLIDE 34

For now, let's focus our attention on our new H3[e] subunit. It looks like the skeleton of a hand gun, with a very, very narrow barrel! You couldn't shoot anyone with this thing! The thin barrel of the gun is the very simple ribbon representation of the formerly-unordered, now *highly*-ordered N-terminus. First, let's switch to an atomic model.

SLIDE 35

Now let's rotate it to see what residues are in the N-terminal β-strand: I'll put in an x-y axis to confirm the direction of rotation:

SLIDES 36-41
[H3 rotation, 6 slides]
[NO AUDIO]

SLIDE 42

In this view, we can clearly see the Lys/Arg residues which will bind DNA, as well as their connection to the subunit body, which we excluded from our previous slides.

This N-terminal β-strand structure is the entire purpose of our new model; to create a new and improved dwelling place for DNA among the positively-charged basic residues of the N-termini, and to model those N-termini to align with the negative charges on DNA. When we

append the DNA to this structure, isolate it from the subunit body, and rotate it, you'll see essentially the same picture we saw previously with the H2A/H2B DNA complex:

SLIDE 43

Here's the H3 N-terminal β-strand, isolated from its subunit body, but with the DNA added. At this point, there's a minor glitch in our presentation, which was quite unintentional, and you may already have noticed it if you've got a really sharp eye: This N-terminus has been flipped 180º relative to the view in the previous slide; that is, the extreme N-terminal amino acid residue, Ala1, which was on the right in the previous slide, is now on the left. If you look closely you'll see the Ala1 methyl group; at the other end, the ε-amino group of Lys27 is hiding between phosphate groups, but if you look carefully you'll see that too. Had we included the body of the H3 subunit in the slide, it would now be off to the right somewhere.

Sorry about that. OK, let's rotate this structure:

SLIDE 44

This movie shows a view which should hopefully be at least *somewhat* familiar by now, except that it's a bit closer-up and at higher magnification, wherefore it well-demonstrates the mode of electrostatic binding of the Lys and Arg residues to the phosphate groups of DNA. If you say this is NOT 'well-demonstrated' , that probably means that you haven't watched the protamine-DNA slide presentation on this web site, where this structure is examined in exhaustive detail. Do not despair, we shall be reviewing that shortly.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 45
Title slide
(No audio)

Whole octamer rotating models

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 46

We now return to the whole octamer, once again in the hybrid view, with atomic models for the four N-termini only, and ribbons in the core. We're now going to rotate the entire structure, which, however, will be far easier to comprehend with an all-ribbon model, so, in a few seconds, we're going to replace the "non-perspective" atomic modeling at the N-termini, with an all-ribbon *perspective* model, which will make the overall structure of our nucleosome easier to comprehend. I should re-emphasize, before we start, that I'm showing you the logical conclusion of this modeling project, *before* showing you the logic behind the model. That we'll get to later:

SLIDE 47

This is the Accelrys DS Visualizer 'perspective view'. I personally prefer to work with a "non-perspective" view, but the DS Visualizer, in spite of all it's lovely graphics, does not have a "non-perspective" view, so you're now seeing the columns of DNA and protein at the peripheries of the four histone quadrants as you'd see four columns of anything in real life, if you were standing above them. Thus, although they are all perfectly parallel to each other, they seem to diverge in this axial view. That, however, is realistic, and will enable us to demonstrate the entire structure as realistically as possible.

Let's rotate the entire structure about the y-axis:

SLIDE 48

Well, here it is. I'm very confident that this, or something very much like it, is the structure of DNA in cells – in *your* cells. If you're a histone scientist, you'll recognize that the octamer core, as portrayed here, has not been significantly altered. The N-termini, however, are now ordered as β-strands, binding DNA, as we've been discussing all along. You'll note that each DNA tetraplex has a β-strand running up one side, and a second β-strand running *down* the opposite side. If you look at the top of the rotating structure, you'll see the H3 subunits, colored white and yellow, which have that small, additional α-helix I referred to before as "Helix Sub-Zero", which enables the otherwise impossibly-long H3 N-terminus to fit into the model.

SLIDE 49

Let's rewind to our starting position, and rotate the structure longitudinally.

SLIDE 50
(Movie, no audio)

SLIDE 51
(No audio)

SLIDE 52

This, then, is our new nucleosome structure. You may be wondering, "Do I, Ken Biegeleisen, really believe in this structure?" The answer is, I believe in it enough to have given a year of my life to developing it, and an additional half-year to the making of these three slide presentations, with no reasonable hope or expectation of reward.

I'll say this much: If the nucleosome core structure, which was proposed in 1974 by Kornberg & Thomas, and confirmed by Luger and her associates in 1997 by x-ray crystallography -- if that core structure is accepted as true, and as correct, then I would opine that the structure for the N-termini shown here, or some alternative structure very similar to it, is a near-certainty. It's based upon my 2006 publication on the structure of the protamine-DNA complex in sperm cells; a structure which absolutely requires *non*-helically-twisted DNA. I regard that structure as incontrovertible, and I hasten to remind you that since 1953, the year of the publication of the Watson-Crick double-helix, which was the beginning of nearly all that falls

under the category of "molecular biology", not a single competing protamine structure has ever even been proposed, much less proven.

If, after nearly 60 years, no model for protamine has ever emerged, which can accommodate a double-helical twisted structure for its associated DNA, then it borders on foolhardiness to expect that one will suddenly emerge now. Wherefore I say that the manner of association of DNA with protamine ought to be regarded as prototypical, unless someone can come up with a better idea — and no one has yet.

If, therefore, as I am quite certain is the case, the structure of DNA in association with protamine is all-but-incontrovertible, than the structure of DNA in association with histones cannot possibly stray very far from it.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 53
Title slide
(No audio)

Two Octamers (Accelryt Discovery Studio view)

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 54

This is our new structure, in the same view we employed a few slides up. We're going to spin it about the x-axis again ...

SLIDE 55

... only this time, please watch carefully at the *top* of the frame, as each of the 4 protein-DNA complexes emerges from the rear of the field, comes over the crest, and moves to the forefront. You'll note that for each of the four DNA tetraplexes, there's a point in the rotation where you'll clearly see a protein strand on one side of the DNA which goes to the right, and a second protein strand on the other side of the DNA which goes to the left. This leaves 50% of the DNA tetraplex surface devoid of protein support. How will our model deal with this?

To see how, we've paused the nucleosome rotation on H2A and H2B, the green and blue [c] and [d] chains, as these come to the forefront.

SLIDE 56

Note the directional orientations of the protein strands, and their locations on opposite sides of the DNA. What about the portions of the DNA opposite each histone strand, marked now by white question marks, which have no protein support structure? What will bind to them?

The answer is: "The next nucleosome". Let's see how.

SLIDE 57

To see how two adjacent histone octamers align themselves for the binding of long stretches of DNA, we're going to have to remove the DNA, and start all over again with our "root position" octamer view...

SLIDE 58

...then we rotate 90º in the +y direction.

SLIDE 59

Now we're going to do a bit of a +x rotation, to get a more favorable view for our next scene, which will be to bring in other octamers, so as to bring about a stepwise lengthening of the protein support structure for the theoretically-endlessly-long DNA chain.

SLIDE 60
(Movie, no audio)

SLIDE 61

Next we'll bring in some more histone octamers, which will, unfortunately, turn the screen into a forest of colored lines. In order to comprehend what's happening, please focus your attention on the white and red H3 and H4 subunits, which I've intentionally positioned at the top of the slide.

SLIDE 62

First, we'll have to shrink this picture down, to make room for more octamers.

SLIDE 63

Let's bring in a second octamer; now a third. Notice that the white H3 strand continues indefinitely in a straight line, with only a small gap between adjacent octamers. The red H4 strand does likewise. If the DNA for these two strands was included in this slide, it would lie between them. Let's see how this might look.

SLIDE 64
Title slide
(No audio)

Two Octamers (Schrödinger)

SLIDE 65

I'm going to repeat the above exercise, because I have a rotating model of two complete octamers, including both the protein *and* the DNA, but it's a Schrödinger export, which looks different from an Accelrys export. In particular, the Schrödinger colors differ slightly, and the N-terminal β-strands are portrayed by Schrödinger as wavy lines, whereas Accelrys portrays the N-termini as straight lines.

So I'm now going to show you the Schrödinger version of the very same Accelrys exercise we looked at above, because it will 'morph' directly into the above-referenced rotating movie of two complete octamers, which will demonstrate the position of DNA in the extended model.

Before showing the movie, I want to say a word about the spacing of adjacent octamers. In my current model, the N-terminal strands are all ***about*** 30 amino acid residues in length, which includes the segments that connect the β-strands to their respective subunit bodies, but truthfully, this model, even assuming it's correct in all its broad strokes, is nevertheless certain to undergo many adjustments, as knowledge increases.

It could be, for example, that a perfected model will have the eight N-terminal β-strands *not* approximately, but *exactly* the same lengths. Or, conversely, it might be that the N-termini, for some unknown reason, might have lengths which are not only *not* equal, but are actually even *more* unequal then is shown here. Only time will tell.

I should point out now that a serious spacing issue, concerning adjacent octamers, actually came up during the preparation of this slide show, demonstrating at the outset that the octamer spacing in this model has not, as of the present time, been determined with any certainty at all. I'll discuss this spacing issue later.

In any event, for now, we shall portray adjacent octamers as being as close together as physically possible, with an inter-octamer spacing based upon the length of the longest N-termini, which are the H3 N-termini, colored white and yellow.

SLIDE 66

So, as we did above with the Accelrys model, we once again slide two coaxially-aligned octamers together until the H3 N-terminal strands almost, but not quite, touch the Helix Sub-Zero coils of the adjacent octamer.

It would obviously be desirable to know the exact spacing, since that will determine the stoichiometry of the extended structure, but since we don't, we must, for the time being, proceed by educated guess. As a first approximation, I presume the close spacing shown here.

SLIDE 67

All that remains is to add back the DNA.

SLIDE 68

This slide shows our complete structure, demonstrating both the mode of association of each histone octamer with DNA, and also the manner in which adjacent octamers are aligned and

spaced, so that the negatively-charged DNA has a continuous association with positively-charged histone octamers from one end of the chromosome to the other.

If you are interested in looking more closely, you can stop the nucleosome rotation in this slide any time, with a mouse-click, which acts as a toggle switch to start-and-stop the motion.

If you do, that is, if you look closely, you'll note that each of the 4 strands of tetraplex DNA has histone β-strands of one color only on one side, and another color only on the opposite side. For example, if you focus your attention on the DNA tetraplex associated with H2A/H2B, you'll see only green H2A β-strands on one side of the DNA, and only blue H2B β-strands on the other.

If you follow any one of the histone subunit N-termini to the extremity of its β-strand, you'll see that the same-colored β-strand of the next octamer picks up where the first one leaves off, with only a small gap.

Now, IFF you've been thinking about this structure as we've been going along, you will undoubtedly have been wondering just *what* exactly are those four tetramers of DNA? Are they simply four parallel but otherwise unrelated DNA units? Or are they perhaps ultimately connected, so that they're really one long, folded structure of some sort?

Or, could it be even remotely plausible that they actually include the DNA of ***both*** homologous chromosomes of the diploid pairs?

At one point I seriously considered the last suggestion, namely that DNA from homologous chromosomes might actually be mounted on the same nucleosome. That would greatly facilitate recombination, but it would greatly complicate replication. After long deliberation, I concluded that it was too unlikely to consider further.

My best guess as to the identity of the many DNA strands shown here is that they are indeed all from the same chromosome, and hence that they represent a single duplex of DNA which has merely been elaborately folded. When we get to the subject of higher-order chromatin structure, I'll explain how this might work.

For now, however, the time has come to show how this structure was derived.

SLIDE 69
Title slide
(No audio)

<u>How the new histone structure was derived (introduction)</u>

SLIDE 70

The primary assumption underlying the new histone structure is that the structure of the protamine-DNA complex in sperm cells is essentially as presented in the accompanying slide show, which we shall be reviewing here, but only briefly. If you're interested in histone structure, you owe it to yourself to watch "The Probable Structure of the Protamine-DNA Complex", which is on this web site.

That the protamine type of structure will be found in histones is an essential assumption, but one of which I am absolutely certain. In the 53 years between the publication of the

Watson-Crick structure and the publication of my protamine-DNA complex in 2006, the molecular biological establishment was unable to come up with a single proposal for protamine structure – not one! They drew a complete blank! And why? Because they couldn't think outside the box; the box being the Watson-Crick double-helix, which simply doesn't work with protamine.

SLIDE 71

The militant helicist, of whom I have known many; men who have appointed themselves to defend the Watson-Crick structure against all challenges; might try to persuade you that either protamine has no structure at all, or that the chemical is simply not important enough to bother with.

I would opine quite to the contrary. Protamine, I would suggest, is not only very important, but perhaps the *most* fundamental of all nuclear proteins, being little more than a string of positively-charged arginine residues, which must somehow bind to DNA, which is a long string of negatively-charged phosphate groups. Since *all* DNA-binding nuclear proteins are distinguished by a preponderance of the basic amino acids lysine and arginine, I would predict that protamine will prove to be the *prototype* for the understanding of both DNA *and* protein structure in the living cell nucleus.

Let me add also that in the 7 years since the protamine-DNA structure was published, no one has come up with a better idea, and no one has found any fault with the published model.

SLIDE 72

I therefore propose that the principles of protamine structure be applied to the N-termini of the histone octamer.

A corollary of this is that we shall be presuming a *non*-helical DNA structure for the histone-DNA complex, or nucleosome, as we did for protamine. There's no point in arguing this here, as there are already over 4 hours of PowerPoint slides on this web site, which show in extensive detail why it is a virtual certainty that DNA, in living cells, cannot possibly have a helical twist.

Another assumption we shall be making is that the published crystal structure of the histone octamer core is correct. I wish I could feel as confident of that as I do about the previous two assumptions. The current histone octamer structure was deduced by Roger Kornberg and his associates, who employed the great tools of data analysis and logic, but it was later confirmed by x-ray crystallography, which is a less-great tool. It should hardly be necessary to point this out in the year 2013, but objects that are hydrated do *not* necessarily have the same structures as objects that are *de*-hydrated:

SLIDE 73
(No audio)

SLIDE 74

I only show this because I perceive that there is an ever-increasing unwillingness on the part of x-ray crystallographers to admit that a crystal structure may not be relevant to real life. This was

alarmingly demonstrated in 2006, when, only a few months after the publication of the protamine-DNA complex, the Rutgers Protein Data Bank stopped accepting theoretical models, insisting on x-ray crystallographic structures only. To this they have magnanimously added NMR-determined structures, but, that notwithstanding, it seems to me that they have, in effect, declared war on the human thought process. You see, people, and other cellular life forms, are not crystals, and the only tool capable of penetrating into the living cell, where neither crystallography nor NMR can go, is logic.

SLIDE 75

Be that as it may, we are absolutely compelled, for lack of any alternative whatsoever, to accept the published crystal structure of the histone octamer as being essentially correct. Having been literally living with that structure, night and day, for more than a year now, I have come to greatly respect and admire it, but I harbor a lingering fear that it may be a laboratory artifact. But what options do I have? There are no other published histone structures to choose from! This is it!

A less frightening, but still disconcerting possibility is that the published structure is correct in its basic details, but slightly different than it would be in the 100% humidity environment of the cell nucleus. I would actually presume that to be true as a matter of course, but we have no way of predicting what effect total hydration would have on the structure, therefore I have, by default, accepted the published structure in all its details, and moved nothing except the N-termini, which do not crystallize anyway, and which are depicted, in the current published structure, as projecting outward in totally random fashion in 8 different directions.

It is for the purpose of imposing order on these N-termini that I have undertaken these studies.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 76
Title slide
(No audio)

Limitations/problems with the present modeling endeavor

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 77

Now here are the limitations of the structural studies I shall report shortly:

First of all, I report on 4 histone subunits, H3, H4, H2A and H2B. But there aren't only 4, there are 5: In certain settings we see H1 (or its cousin, H5, pictured here). The trouble is, H1 is not invariably present, and, worse still, I have no clue as to what to do with it – no clue at all. Therefore, and since there are regions of the chromosome where there is no H1, but only H3, H4, H2A and H2B, I have elected to proceed without any consideration of the role of H1 whatsoever – I have just ignored it. Whether this is a valid approach to the problem of histone structure or not, only time will tell.

SLIDE 78

Here's another problem:

The first 10 amino acids of histone subunit H2B contain 4 proline residues, an extraordinary accumulation in such a short space. The sequence is shown here, both in isolation, and in its natural setting at the N-terminus of subunit H2B.

PE**P**AKSA**P**A**P**

In addition to this sequence being an unlikely candidate for binding to DNA, it is furthermore a sequence not likely to "cooperate", so-to-speak, in the sort of ordered structure I wish to propose. Moreover, the length of its DNA-binding N-terminus , because of this sequence, winds up being about 10 amino acid residues *longer* than the DNA-binding portions of the N-termini of the other 3 histone subunits, which increases my suspicion that these first 10 amino acids have some specialized role that I cannot even begin to imagine.

Nevertheless, since there's no way to even guess as to what else to do with them, I have reluctantly treated them the same as the other ~200 N-terminal amino acid residues of the histone octamer.

SLIDE 79

A third problem with our model is that it makes no attempt to deal with the unstructured C-termini of the histone subunits. The reason is that the C-terminal lengths and primary amino acid sequences show no consistency from subunit to subunit, which makes it difficult to envision a single, unified purpose for them all.

Subunits H3 and H4 have almost no C-termini at all. Let's isolate the H3 subunits. Here now are the H3 [a] and [e] chains. Note that these each have essentially *no* C-terminus at all; in fact there are only 4 residues distal to the final α-helix; a sequence so short that I've had to insert a blue arrow to mark their location.

SLIDE 80

Here's the octamer again. This time, let's isolate H4. The ribbons for the C-terminal strands have been highlighted in yellow. You can't tell which end of either yellow curve is the terminal residue, so I'll add some arrows to mark them.

At 9 amino acid residues in length, the H4 C-terminus is longer than that of H3, but not very much longer. Moreover, it contains only a single basic residue; not very promising with respect to DNA binding.

SLIDE 81

Once again we return to our octamer. Let's look at H2A. In marked contrast to H3 and H4, the H2A chains have rather *long* C-termini, which I've again highlighted in yellow. These C-termini are *so* long that they impart to H2A the distinction of being the only histone subunit which extends the full diameter of the octamer, and whose terminal strands protrude from both

sides thereof. Moreover, the H2A C-terminal amino acid sequence is not only long, but also strongly basic.

While this might be construed to suggest a DNA-binding potential, I must note also that there are a surprising number of acidic residues as well. I have no idea what to do with such a sequence, so I'm just ignoring it for now.

SLIDE 82

Finally, let's isolate H2B. First, let's look only at the blue-colored [d] chain. The C-terminus is mostly occupied by a long α-helix, highlighted in yellow. In part I of this slide series, we named this "Helix IV". Now we'll add the pink-colored [h] chain, also with Helix IV highlighted in yellow.

Within this relatively long C-terminal α-helix are a few lysine residues, but unless we arbitrarily assume that Helix IV is an artifact, and unless we unwind it, we will not be able to make any use of these basic residues to bind DNA. Besides, since an analysis of its amino acid sequence reveals that Helix IV is maintained by powerful hydrophobic bonds to neighboring structures, I have no *intention* of unwinding it. Moreover, the C-terminus of the H2B subunit also has some acidic residues which only raise questions, to which no answers can be forthcoming at the present time.

Therefore, in the absence of data to steer us in any useful direction with respect to the modeling of this, or any of the other 3 histone subunit C-termini, I have simply left them as is.

SLIDE 83
Title slide
(No audio)

Review of protamine structure

SLIDE 84

The single overwhelmingly most important precedent for the new histone structure presented here, was the structure of the protamine-DNA complex in sperm cells. There is, on this web site, an entire slide presentation just on this subject, entitled "The probable structure of the protamine-DNA complex". If you're seriously interested in nucleoprotein structure, you owe it to yourself to watch that entire presentation.

I know that many of you will not do this, therefore I shall present a brief review of protamine-DNA structure here, because without at least a basic understanding of it, you will not be able to comprehend the logic behind the new histone structure.

We start with amino acid sequence. Human protamine has 2 chains, called P1 and P2, with lengths of 51 and 57 amino acid residues, respectively. Because the cysteine residues line up very precisely, whether it be human or any other protamine, it's impossible for me to doubt that P1 and P2 are a heterodimer, linked together by disulfide bonds.

You may have noticed that the strands are portrayed here as being parallel, which is a slightly-less-favorable orientation for a β-sheet than anti-parallel, but in the anti-parallel orientation the alignment of cysteine residues was not as good. Or so I thought. An ingenious scientist from the Pasteur Institute figured out how to turn this structure into an anti-parallel one, and I'll show that to you a little later.

SLIDE 85

Let's remove the residue numbers, and focus on the amino acid sequence.

SLIDE 86

The single most important amino acid in protamine is arginine; the longest and most basic of all the amino acids. The arginine structure is shown at the top of the slide.

With the arginine residues highlighted in red, you can see at a glance that they account for over half the residues of the protamine dimer.

SLIDE 87

There's not much lysine in protamine. P1 has none, and P2 has only 2 residues. There are, however, 10 histidine residues, mostly in P2. The lysine and histidine residues are highlighted in blue.

SLIDE 88

The basic residues are now all colored red. You can see at a glance that this is an extremely basic protein.

SLIDE 89

Virtually all of the remaining residues have hydrophilic side chains, capable of hydrogen-bonding with DNA phosphate groups. These residues are Tyr, Ser, Gln, Thr, and Cys.

SLIDE 90

With all these residues colored red, you can see at a glance that protamine is an exceedingly hydrophilic protein. Of its 108 amino acid residues, there are a total of only 8 which are hydrophobic. ***There is thus no possibility whatsoever that this protein could have a classic "globular" conformation,*** with a hydrophobic interior. There's simply nothing hydrophobic about it. Nor could a globular interior be maintained by salt bridges, because there's not a single acidic residue in either strand.

SLIDE 91

There's also no real possibility of an α-helical structure for protamine. I suppose it's possible to make a case for a poly-arginine α-helix in the presence of some sort of external,

negatively-charged support matrix, but Watson-Crick DNA cannot be that matrix, because its helical winding sense and charge distribution are completely wrong.

One might argue that the sperm nucleus *has no structure at all*, and that the protamine and DNA are just there, bumbling about mindlessly, but I would opine that these chemicals will not remain without structure if a highly-ordered and thermodynamically-favorable structure is available, and if there's no bar to it forming. And there is, in fact, just such a structure: the β-sheet.

SLIDE 92

In order to see how DNA can form a logical structure with a protamine β-sheet, we need to consider two key dimensions in DNA. We start by noting that, although the base spacing in helical DNA is 3.4 Å...

SLIDE 93

...the actual distance between residues, when measured along the sugar-phosphate backbone, is 7 Å. This is the first of our key dimensions in DNA. We shall therefore seek, and shall find this spacing, in protamine.

SLIDE 94

We shall also seek, and find, the DNA 20 Å cross-duplex phosphate-to-phosphate distance in protamine. This is our 2nd key DNA dimension.

SLIDE 95

Finding the first of our two "key" DNA dimensions in protamine, namely the 7 Å spacing between residues, is easy. The 7 Å spacing is found when protamine is configured as a fully-extended, flat β-strand, with $\phi=\psi=180^{o}$.

SLIDE 96

Finding the second "key" DNA dimension in protamine, namely the 20 Å phosphate-to-phosphate cross-duplex distance, is a bit more subtle. At the spacing mandated by a standard disulfide bond, shown here in yellow, the distance between the guanidinium groups of a pair of fully-extended arginine side chains in the protamine dimer is...

SLIDE 97

... 14 Å.

SLIDE 98

Here's an axial view of the same structure. 14 Å is considerably less than the 20 Å DNA phosphate-to-phosphate cross-duplex distance, but...

SLIDE 99

...when we leave about 3 Å for a salt bridge, the fit is quite good.

SLIDE 100

Here is a type of unit cell of protamine-DNA structure, in axial view. P1 and P2 are linked by disulfide bonds, and there is a DNA tetraplex on either side. Actually, only DNA duplexes are seen here, not tetraplexes, because the tetraplexes require adjacent unit cells. These turn the duplexes into tetraplexes by intercalating their base pairs between the base pairs shown here. This will be demonstrated in a movie, a few slides from now.

The rotational state of the protamine dimer that is shown in this slide, which is taken from my 2006 publication, could be changed. A rotation of the protein by 90º, would alter the structure slightly, because the protamine dimer, as depicted here, is not exactly square. But with minor changes in bond and dihedral angles, the 90º rotation could be accomplished with almost no significant change to the structure at all.

I mention this 90º rotational change in the protein only because we shall be carefully considering the rotational states of both the DNA *and* protein in our new histone structure, to be described after the completion of this protamine review.

SLIDE 101

Here's a longitudinal view of our protamine structure. This drawing schematically depicts the essence of the protamine-DNA complex. The DNA base pairs are excluded for the moment, so that we can focus on charge interactions. Note the perfect alignment between the positive charges on arginine guanidinium groups, and the negative charges on DNA phosphates.

The only thing wrong with this structure is that the 7 Å spacing in the totally-flat protamine β-sheet is a bit large, and requires that the DNA base pairs be a bit too far apart. So we have to adjust dihedral angles in both the protein and DNA.

SLIDE 102

By altering the phi and psi angles, we can contract the structure.

SLIDE 103

At psi and phi angles of approximately +/– 130.5º respectively, we get an R-group spacing of 6.8 Å, which, as we shall see, is perfect for DNA.

SLIDE 104

These angles place this β-sheet in the most-favorable region of the Ramachandran Plot, as indicated by the blue star.

SLIDE 105

To get the DNA base-pair spacing down to its preferred distance of 3.4 Å, we need to intercalate the DNA duplexes with neighboring ones, as shown in this partially-literal, partially-schematic movie.

SLIDE 106

Here's the entire structure; the complete P1-P2 dimer, binding approximately 59 base pairs of DNA. Now let's watch a *non*-schematic, atomically-accurate movie showing intercalation of adjacent protamine-DNA complexes.

SLIDE 107

This shows how the structure propagates along the x-axis, by mutual intercalation of base pairs.

What about the propagation in the z-axis direction, which, in this view, is either toward or away from your eyes? To see this, we will have to tilt the complex forward.

SLIDE 108
(Start of movie, no audio)

SLIDE 109
(Movie, no audio)

SLIDE 110

Here's the axial, or top view of the structure. Let's enlarge this:

SLIDE 111

Here's a representative cross-section through the column of atoms, at a level where there are 4 consecutive Arg residues. Let's add some labels.

SLIDE 112
(No audio)

SLIDE 113

What we're going to show next is an extension of the structure through mutual intercalation of neighboring base pairs, exactly as we saw a few slides up, only this time from the axial perspective. The protamine-DNA complex will increase in the directions shown.

SLIDES114-117
(No audio; rapid sequence of slides)

SLIDES 118-121
(Audio on Slide 117 only)

The association of adjacent protamine-DNA complexes, through mutual intercalation of their DNA base pairs, establishes a pattern of alternating columns of DNA and protein.

SLIDE 122

This pattern of alternating columns of DNA and protein establishes a higher-order 'row' of DNA & protein columns which continues for the ‘length’ of the cell nucleus. But what about the rows of DNA and protein columns which presumably lie to the right and left of those shown here? How will they fit into the structure?

As I pointed out in 2006, in the accompanying slide show, "The probable structure of the protamine-DNA complex", when I began looking at protamine, I had no pre-existing plan to account for the growth of the structure in the directions which, in the present slide, would be properly-described as 'to the right' or 'to the left'.

As it turns out I didn't need a plan. In what I can only account a stroke of divine good fortune, it turns out that when adjacent rows of DNA and protein columns are approximated…

SLIDE 123

…and moved by a distance corresponding to one-half of the unit cell...

SLIDE 124

...a square-array pattern of salt bridges appears, which is so wonderfully-favorable that it is impossible to believe, by any stretch of the imagination, that it is a mere modeling coincidence, with no correlation in real life. Rather, it must be that this is, in fact, the actual structure of DNA and protamine in the sperm nucleus.

SLIDE 125

Here's a slightly-displaced view of the same structure.

The purple box shows one way to define the repeating unit of the protamine-DNA complex. We could also move the box up or down, but this position will suffice for now. If we call this the 'unit cell', then the position of the column on the right may be said to be offset by one-half unit cell. This gives rise to the extraordinarily fortuitous array of salt bridges that are marked by the small gray squares,...

SLIDE 126

...now highlighted in purple. Let’s look more closely:

SLIDE 127

What we see, everywhere that DNA and protein meet, is a perfect square-array of charge interactions. I had no idea I would find this when I began working on protamine-DNA structure in 2005. I don't wish to belabor the point excessively, but I must reiterate that it is inconceivable to me that this is a mere coincidence.

SLIDE 128

If you're like me, then when you look at the square array, you'll be impressed by the obvious charge attractions, highlighted here, but...

SLIDE 129

...you may wonder about the charge *repulsions* at the diagonals. But the diagonal of a square is 1.41x the sides in length, and, since the strength of an electrostatic attraction is related to the *square* of the distance, we see that the attractive forces in the sides of this charge-array are fully *twice* as strong as the repulsive forces at the diagonals, wherefore we may rest assured that this is a *very* stable structure.

SLIDE 130

'Helicists', as I have come to call them, never tire of teaching people that DNA absolutely *must* be twisted, supertwisted, and subjected to even higher-order contortions, if it is to fit into the nuclear space. Really? Let's see what percentage of the available space, in the sperm nucleus, is consumed by our non-helical protamine-DNA complex. We see here that the size of the sperm nucleus is about 40-50 μm^3. Will our structure fit?

SLIDE 131

Here's a calculation I went through line-by-line in the accompanying slide presentation on protamine-DNA structure, so there's no need to do that again here. The numbers are all on this slide, and what they show is that, if this unit cell of protamine-DNA structure, consisting of a single instance of the P1-P2 dimer plus its approximately 59 base-pairs of associated DNA, is extrapolated to the haploid amount of DNA found in a germ cell, it consumes only about 10% of the available space!

This suggests that in a diploid somatic cell, with twice the DNA, a structure like this will take up only about 20% of the available space. We shall, in fact, be doing a comparable calculation for the histone-DNA complex shortly.

I think you can see that the allegation, that DNA must be twisted in order to fit into the available space, is not very well grounded in anything you would want to call 'science'.

SLIDE 132
Title slide
(No audio)

Jean-Luc Jestin (Pasteur Institute)
Anti-parallel protamine structure

SLIDE 133

In November 2012, six years after the publication of my manuscript on protamine-DNA structure, I received correspondence from a brilliant chemist named Jean-Luc Jestin, of the Pasteur Institute. In my model, the P1 and P2 strands were parallel, and he had figured out how to make them antiparallel.

The reason my strands were parallel was that I was trapped in a manner of concrete thinking, with respect to the relative locations of P1 and P2. I had automatically assumed that each molecule of P1 had to be uniquely associated with a single molecule of P2, as depicted both here and in the protamine slide presentation on this web site. The alignment of disulfide bonds was definitely inferior in the antiparallel configuration, which I therefore rejected.

The current slide shows my original model. Note that the strands are parallel, and that the 4th disulfide bond on the right is of poor quality. Justin suggested rotating the P2 strand at Arg42, which is colored orange in this slide. Let's center the picture about the orange point of rotation.

SLIDE 134

Now let's rotate by 180^{o}.

SLIDE 135

OK, we've now inverted P2. The labels are a bit hard to read upside-down, so let's re-label the drawing.

SLIDE 136

There, that's better. Now we have a pair of anti-parallel β-strands. Let's bring in the next pair of protamine monomers.

SLIDE 137
(No audio)

SLIDE 138

Now we have 4 perfect disulfide bonds, and an anti-parallel structure. Since the antiparallel β-sheet has better hydrogen bond alignment than the parallel type, and since this model has onere perfectly-aligned disulfide bond than my original did, we might consider it to be a superior model.

The only problem is that, once one begins "thinking outside the box", then other structures begin to suggest themselves. For example, both these models could be construed to have one small drawback, namely that there's a small, 6-residue-lengthed gap between adjacent P1 molecules. If we wanted to, and if we were prepared to pay the modeling price, we could get rid of that space, as shown on the next slide.

SLIDE 139

Here’s my original parallel structure. Let’s remove the annotation...

SLIDE 140
(No audio)

SLIDE 141

…ad move it over a bit.

SLIDE 142

Now we can bring in a second protamine P1-P2 dimer, this one totally inverted, and move it close to the first one. Note that the inversion of the P1-P2 dimer totally eliminates the gap. The next P1-P2 dimer will have the original orientation...and so on.

The modeling price we pay for this is that every other protamine P1-P2 dimer must have its orientation reversed. Is that too hard for the cell to do? I doubt it, but who can say which of the various alternative structures is correct?

The point is this: as I said in my original publication, this protamine-DNA model is not a single, unique structure, but rather a *type* of structure, with numerous possible variations. The only invariable aspects of it are that the DNA must be a non-helical tetraplex, and the Arg/Lys residues of the P1-P2 dimer must align with the DNA phosphate groups. These are the attributes we shall seek in histones.

SLIDE 143
Title slide
(No audio)

New histone structure (based on protamine model)

SLIDE 144

If we are going to take the principles we learned from study of the protamine-DNA complex, and extend those principles to histones, where a similar protein-DNA arrangement ought to be possible in the 8 lysine- and arginine-rich N-termini of the histone octamer, then we're going to have to find the critically-important structures and spacings we found in the protamine-DNA complex. For DNA, we shall stick with the structure that surely pertains to sperm cell nuclei:

- **Wu "straight-ladder" DNA duplexes, with 6.8 Å base-pair spacing,** and...
- **Intercalation of adjacent DNA duplexes, to restore 3.4 Å spacing.**

Within the histone octamer, the critically-important spacing we shall seek is that which flanks each DNA tetraplex in the protamine structure. This spacing is around 15 Å, as shown in the next slide:

SLIDE 145

Let's review the spacings in protamine. In this slide we see an axial view of the protamine-DNA complex in sperm cells, showing a small portion of the sperm nucleus containing 4 protamine P1-P2 dimers and 4 DNA tetraplexes. Notice that the spacing between protamine dimers, necessary to accommodate the DNA, is in the range of 15 Å. That is the critical spacing between protein components that we shall be seeking in the histone octamer, and we shall, in fact, find it. But before we do, there are some preliminary points which must be made, concerning certain spacings and configurations which are not going to be exactly the same in histone as they were in protamine.

SLIDE 146

First of all, in the protamine-DNA complex there was an approximately 14 Å spacing, shown here, between oppositely-pointing arginine guanidinium groups in any given cross-sectional plane, this distance representing the sum of the lengths of the two arginine side-chains, plus the length of the disulfide bond separating the two arginine residues.

In histone, however, about half the basic residues are lysine, which has a slightly shorter R-group than arginine. Furthermore, there are no disulfide bonds in histone. This means that a structure comparable to the one shown in this slide, when found in histone, will, on the average, have a somewhat smaller spacing between the extremities of the basic residue side-chains.

Fortunately, this smaller spacing turns out to have little effect on the ability of DNA and protein to form salt bridges, because the N-terminal β-strands of the protein can readily move closer to the DNA, if necessary, to form 3Å salt bridges between the lysine ε-amino groups and the DNA phosphates. Arginine residues in the same β-strands, however, cannot then assume the fully-extended conformation shown in this slide, but rather they must curl up a bit, to avoid having their guanidinium groups being *too* close to DNA.

Neither of these things causes any serious problems for the model builder.

SLIDE 147

The reason there are no disulfide bonds in histone is that there are very few cysteine residues, and, of the few which may be found, there is no obvious alignment such as we see in the protamines of most species.

Even without disulfide bonding, however, there's no reason to doubt that adjacent N-terminal β-strands of one histone octamer can align, by hydrogen bonding alone, with the N-terminal β-strands of adjacent histone octamers, to form small β-sheets. Here's a preview of an axial view of one such structure, where a pair of β-strand peptide backbones from two adjacent histone octamers are thusly juxtaposed. Such a structure would be comparable to an analogous sort of structure which almost surely exists in salmon protamine, which has no cysteine, but which, in every other respect, appears essentially the same as other protamines which have been studied.

In salmon protamine, I have previously suggested, as a theoretical possibility, that the best fit with DNA might be obtained by rotating the P1-P2 dimers by 90º relative to the orientation I have proposed for cysteine-containing protamines. In histones, we shall have to re-visit that question, because it is clear that both rotational options are possible. Allow me to illustrate.

SLIDE 148

Here yet again is the old protamine-DNA axial view. Let's put back the labels showing the 20 Å spacing between DNA phosphate groups, which, in the presence of a 14 Å spacing between arginine guanidinium groups, makes for a good 3 Å salt bridge. For completeness, I've also added the 3 Å spacing between protamine chains P1 and P2, which, in human protamine, is enforced by disulfide bridges. In histone, any such spacing will have to be maintained by hydrogen bonding of the β-sheet backbones only, because of the lack of cysteine.

But the purpose of this slide is to show something entirely different. Please note that the P1-P2 dimer, in axial view, is approximately square shaped. Therefore the dimer will have approximately the same relationship with DNA phosphates even if it is rotated 90º.

SLIDE 149

Note that after a 90º rotation, the 3 Å spacing between the negative phosphate groups and the positive guanidinium groups is essentially preserved, simply requiring minor adjustments of the protein side chains. In histones, therefore, we are going to have to decide between these two rotational states.

SLIDE 150

OK, let's restore the original protein rotation. So now we know that if we are going to apply the principles of protamine-DNA structure to the histone octamer, we'll have to adjudicate between two competing protein rotational states. But this axial view also reveals that the same question will arise with respect to the DNA. Please note that the DNA, can, and *does* have two different orientations in its charged-based interaction with the protein.

Here's the first, where the DNA phosphate groups are a little farther apart than the basic amino acid residue side-chain termini; and here's the second, where the DNA phosphate groups are a little closer together than the basic residue side-chain termini. As was the case with the protein, these DNA positions are also related to one another by a 90º rotation.

If there are 2 possible protein orientations, and 2 possible DNA orientations, then we are going to have to adjudicate between 2x2, or *4* different DNA-protein structures, and decide which one, or which *ones,* are correct for histones.

For the model to be presented here, we shall surely choose the rotational options that appear to make the most sense, but there can be no absolute answer to this question without further data. Therefore, from this point forth, nothing shall be 'carved in stone'.

SLIDE 151
Title slide
(No audio)

Determination of the rotational state of protein and DNA

SLIDE 152

Let us put the protamine structure in the setting of the histone octamer, so that we can review the possible rotational states for the DNA and protein chains in the most general sense.

We're going to play a little modeling game now. We're going to pretend that the β-strands of the protamine P1 and P2 chains are N-termini of histone core proteins, and try to figure out what the rotational states of the protamine chains might be, and how they might connect to the histone core.

We have three bodies of data to guide us: (1) The protein and DNA relationships established in the protamine-DNA complex, (2) the spatial relationships and other characteristics of the Luger *et al* crystal structure for the histone octamer core, and (3) the conditions imposed by the requirement that whatever structure we come up with, it be at least possible to incorporate it into some sort of plausible 30 nm fiber, the next step up in chromatin structure, for which there are no currently-agreed-upon models.

So, how might the histone core structure connect to these protamine chains?

SLIDE 153

Well, each of the 8 histone octamer subunits ends in an N-terminal strand, all of which, in the current model of histone structure, are portrayed as being disordered. Let us presume that 2 of those N-termini are configured as β-strands, and that they give rise to two of the protamine-like chains, as shown here, very, very schematically, each arrow representing a short polypeptide link between the body of one of the subunits of the histone octamer core, and that subunit's N-terminal strand, which we are going to represent, for the moment, as if it was identical in structure to a strand from the protamine β-sheet shown.

This, however, is problematical. If the N-termini of the two histone subunits are to eventually be involved in β-sheet structures, as suggested by the second set of arrows coming in now, with backbone-to-backbone hydrogen bonding, *where* will those other 2 strands come from? There's no geometrically possible way for the other β-strands to come from the single octamer shown; therefore they must come from a 2nd octamer,...

SLIDE 154

...giving rise to this sort of structure, where each octamer provides half of a β-sheet, and where the resulting β-sheets provide a scaffolding for DNA. Since such a structure would be copiously-provided with a huge number of well-positioned salt bridges and backbone hydrogen bonds, one could argue that it is energetically-favorable, and I would so argue.

But I doubt this structure exists, because it has what I would regard as a fatal flaw.

To see the flaw, we need to focus on the DNA. Let's remove the upper DNA tetraplex in this slide, which is not involved in the current consideration, and focus on the lower DNA tetraplex.

SLIDE 155

Assuming, as I must, that the DNA structure is the same as that which must, logically, exist in the presence of protamine, we are then dealing with a tetraplex, consisting of two mutually-intercalated duplexes.

Let's define the DNA base pairs, so that we can more thoughtfully consider the merits of this structure. Suppose that it's A-T on the top, G-C on the bottom. Note that the left-hand member of each bp "belongs", so-to-speak, to the octamer on the left-hand side of the slide, by virtue of its salt bridges to the adjacent Arg or Lys R-chains, now highlighted by yellow stars. But the right-hand member of each base pair "belongs", so-to-speak, to the octamer on the right. Well, whatever possibilities there are for such structures in special situations of some sort, this cannot be a structure for long-term storage of DNA in the resting state, and here's why.

Based upon what we already know about the dynamic aspects of histone function, it is a certainty that in real life, histone octamers are in no way static, lifeless crystalline objects. Considering all that goes on in a cell, it must be the case that histone octamers are extremely dynamic objects, undergoing conformational changes on a regular basis. But if either of *these* octamers so much as hiccups, the base pairs will be torn apart!

SLIDE 156

This model cannot be said to be without merit, because, whatever its weaknesses, it does one thing well, namely it provides an exposed surface for reading of the base sequence by polymerases. Even there, however, there are at least two serious problems. The first one is this: Although the correct, hydrogen-bonding faces of the bases are properly exposed by the separation shown here, there are *two* DNA duplexes here, with mutually-intercalated bases. Although that is not visually evident in this axial view, you can quickly bring it back to mind by returning to Slides 104-106. With respect to the subject matter of the current slide, this intercalation has the result that every-other-base in the polymerase's reading frame is from the

wrong DNA chain. That means that the polymerases would have to be capable of reading every-other-base, which strikes me as being hugely unlikely.

My second objection is that this opening up of the reading frame for polymerases requires the *joint activity* of these two adjacent histone octamers, which means that they are, so-to-speak, a "team", and must always stay together, and must always either remain frozen in place together, or else move together in perfect concert. Since each octamer has other histone neighbors, with whom such relationships would presumably also exist, the implication is that the entire cell nucleus' multitude of histone octamers are all marching in lock-step. Although this cannot be logically ruled out, it nevertheless strikes me as sufficiently *unlikely* that I have opted to not waste any more time on this model.

SLIDE 157

Might rotating the DNA give us a better structure?

SLIDE 158

Let's find out. We'll need to move the protein a bit, then we can rotate the DNA 90º. Now let's move our polypeptide connection to the octamer core.

SLIDE 159

Now we have a new structure with complete base pairs uniquely-associated with either the octamer on the left, or the octamer on the right. Is this a viable model?

I would opine not. Although we've removed the problem of base-pairs being violently torn asunder, we've replaced it with a new but similar problem: If either histone octamer so much as hiccups, base stacking will be disrupted.

SLIDE 160

Although, once again, we cannot see it in this axial view, the separation of the tetraplex into two duplexes leaves both columns of DNA with 6.8 Å base spacing, which is not a stable configuration. One might conjecture that this sort of separation would be non-problematical, because it would usually occur only when polymerases or other enzymes passed briefly by. There are several problems, however. First of all, such separation once again requires that two adjacent octamers function continuously as a single unit, either both remaining frozen, or both moving in perfect concert, which strikes me as unlikely.

SLIDE 161

Secondly, as we can see after we bring in these little pink lines to highlight the traditional hydrogen bonds of the A-T and G-C base pairs,the polymerases would be reading the base sequence from the non-hydrogen-bonding side of the DNA. This is thought to be possible, but has not, to the best of my knowledge, been proven.

And finally, there is the matter of devising a model for a 30-nm fiber.

SLIDE 162

This slide doesn't look anything like what we have been studying above, but it is, in fact, a direct manifestation of it. What we have here is a skeleton model of a single histone octamer, containing only Helix I. Although it may be difficult to appreciate in this view, the hypothetical structures of the subunit N-termini and their associated DNA, at the 4 corners of this very octamer, are essentially identical to the structure we have been looking at in the slides immediately above, differing only in minor details concerning the mode of connection between the subunit N-termini and the octamer core.

Don't worry if you can't see it, because we'll be looking at this sort of structure much more closely later. The important thing to note now is that this, like *all* protamine-based histone models, places a DNA tetraplex at each of the 4 corners of the histone octamer. If, in the slides above, I have been explaining myself adequately (which is probably a "big if"!), then you'll understand that in this particular model, we have attributed each instance of mutual intercalation of base pairs to *two* adjacent histone octamers, and since the fundamental motif shown here is only one of those two, we see therefore an octamer core associated with 4 DNA *duplexes,* not tetraplexes, at the corners, each duplex impossibly stacked at 6.8 Å. These only become tetraplexes when other octamers are brought in and intercalated,...

SLIDE 163

...which we show here. Watch the lower RH corner of the first octamer, as we add a second. Although the graphics of this slide are a bit crude, you'll hopefully appreciate that the association of these two octamers comes about through mutual intercalation of base pairs at their corners. If we now bring in a 3rd and 4th octamer, then these further mutual intercalations totally solve the base-stacking problem in the middle of the structure. Moreover, look at this: Voila! We have created, without even attempting to have done so, a beautiful-looking model for a 30-nm fiber!

But look at the periphery! As has been the case all along with this structure, the periphery consists of badly stacked, that is, 6.8 Å-spaced DNA duplexes, and the number of them has only *increased.* We can, of course, bring in yet *more* similarly-structured octamers,...

SLIDE 164

...and in principle we could fill the *entire* cell nucleus with these things. But there will *always* be badly-stacked bases at the periphery, and the number just keeps growing. Moreover, since there is general agreement that the next higher step in chromatin organization is the 30-nm fiber, we would like there to be some logical reason to stop at that stage, but there is none. At whatever stage we stop, there are still multitudes of non-intercalated 6.8 Å-spaced DNA duplexes crying out for partners with which to intercalate; a process which will continue until the entire cell nucleus is filled with this unsatisfactory structure.

For this reason, and for the reasons given previously, I reject this structure, and any other structure in which the DNA tetraplex is shared between histone octamers.

SLIDE 165
Title slide
(No audio)

Likely model for N-terminal structure

SLIDE 166

OK, we've looked at some models that do *not* work; now let's look at the one that *does* work.

After spending much time with the models we looked at previously, and several other similar models I will not trouble you with, I have concluded that the entirety of each DNA tetraplex must be under the jurisdiction of a single histone octamer. There are still, however, multiple ways to structure such models. The form I consider most likely starts, as they all do, from this same view of the protamine-DNA complex we have looked at several times before.

First we must rotate the protein 90º, because, as we have already seen in the slides above, if the protein is oriented as shown here, then the two members of each associated base pair will wind up belonging to different histone octamers.

Next we remove the extraneous DNA and protein,...

SLIDE 167

...then we center the structure, remove the β-strands which do NOT bind directly to this DNA...(these β-strands, by the way, will be restored by bringing in adjacent histone octamers, as we shall see momentarily)...finally, we bring back the body of the current histone octamer, and connect the N-terminal β-strands to the octamer core, from which they must, after all, emerge.

This is going to be the fundamental building block of our new histone structure, the exposition of which is the entire purpose of the current slide presentation. The DNA base-pairing and intercalation are now exclusively associated with a single histone octamer, so our DNA tetraplexes can no longer be disrupted by histone movements.

SLIDE 168

The only problem is that we now have a pair of β-strands whose peptide backbones lack β-strand *partners* with whose backbones they can associate, by hydrogen bonding, to form β-*sheets*. As we shall see, however, the β-strand partners that we seek, can, and will come from neighboring octamers. But before showing that, let's finalize our DNA rotation.

The DNA orientation in the current slide mandates an approximately 15 Å spacing between the protein strands.

There is, in fact, a critically-important 15 Å spacing in the histone octamer. 15 Å, as we shall see shortly, is the distance between the N-terminal ends of adjacent Helix I α-helices, and the same spacing is found in each quadrant of the octamer.

Could it possibly be mere coincidence that the protein spacing we require, in order to apply protamine structure to the histone octamer, not only ***is*** found in the octamer, but, moreover, is found at all 4 quadrants, and at the N-terminal ends of the octamer subunits, where the majority of the basic residues are known to lie?

I would opine that it's no coincidence at all, but rather an important clue beckoning us to a new, and perhaps *improved* histone structure. But, before moving on to discuss this new structure, I shall ask one last question: We stated earlier that, if one wishes to apply protamine structure to the histone octamer, there are two possible protein rotations, and two possible DNA rotations. I have ruled out, to my own satisfaction at least, the alternative protein structure to the one shown here, but is it possible that a better histone structure could be had by rotating the DNA 90º from its current position? Let's take a fast look at this:

SLIDE 169
(No audio)

SLIDE 170

Here is the rotated structure. From most points of view, it's a perfectly reasonable model, and there's not much to distinguish it from the model we were just looking at a moment ago. But look at the spacing: By rotating the DNA, we have increased the distance between the protein strands to about 20 Å. This spacing is not auspicious, because there is no naturally-occurring 20 Å spacing in the histone octamer, with which to accommodate the DNA in this position. I'm not suggesting that a 20 Å spacing is impossible, but it seems foolish to assume, as a first approximation, an orientation which creates a problem, when there's a better-looking one that fits the natural scheme of things.

Therefore, pending evidence against it, I'm going to assume, for the duration of this slide presentation, that the previous DNA orientation, with the 15 Å protein strand spacing, is the most likely one.

SLIDE 171

Our introduction to the new histone structure is almost complete. Our only remaining task is to complete the two β-sheets that are started by the two β-strands shown. As I said above, this will be done by bringing in adjacent histone octamers. Here's the first, and here's the second.

As we shall see later, when we begin to look at higher-order chromatin structure, starting with the 30-nm fiber, the orientations and alignments of protein and DNA strands, at the N-termini of the histone octamer, are probably not finalized, and cannot be at this point in history. I suspect that's because we're dealing with a core structure that was determined from x-ray crystallography, and I fear that the fully-hydrated form in the cell nucleus might have a somewhat different structure.

SLIDE 172

In closing this section, I should mention that I have no trepidation about invoking adjacent histone octamers to complete this structure, as is done here. I was not happy doing that

with the several models which I proposed earlier, then rejected, because in those earlier models, the integrity of the DNA base-pairing and stacking were threatened.

But, as shown in the model depicted here, the hydrogen bonding of adjacent β-strands, to form a β-sheet, is non-informational. I would think, therefore, that breaking and re-establishing those hydrogen bonds would not likely cause trouble for our model, since I don't imagine it would have any serious disruptive effect on the read-write processes involved in the never-ending propagation and transcription of our genetic information.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 173
Title slide
(No audio)

Details of the new histone octamer structure, and the derivation thereof

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 174
Title slide
(No audio)

I. Histone equatorial plane
1. Tetramer 2-fold axis

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 175

Here's another look at the histone structure I shall propose. We've seen this same picture before, so we can't really call it a "preview" anymore, but I still haven't explained all the critical steps in developing this model.

The model consists of 4 DNA tetraplexes, each one uniquely associated with this single octamer, and no other, so that neither the base pairing nor the intercalation of base pairs can be disrupted if the octamer moves. As I showed you above, the N-terminal β-strands can be hydrogen-bonded to the β-strands of adjacent histone octamers.

How was this structure deduced? There were two key discoveries that led to it. The first was the finding of the necessary 15 Å spacing in the protein, to accommodate the DNA tetraplexes. This was found to be the N-terminal spacing between the α-helices known as "Helix I", in each of the 4 quadrants of the octamer. One of these 4 spacings is now indicated in the upper right-hand corner. We'll look more closely at these spacings later.

The second key discovery was the realization that there exists, in the published crystal structure of the histone octamer, a thing I call the "histone equatorial plane". This was a plane of approximate symmetry which enabled me to define a likely direction for DNA chains to travel, which otherwise would have been a complete mystery.

Before looking at the "histone equatorial plane", I wish to review the 2-fold axis of symmetry that underlies it. For these purposes, I'm going to step back and go through all the

steps from the un-ribboned atomic view to the ribbon view of the truncated octamer core, so that we can be sure we are firmly oriented in our thinking, with respect to the view at which we are looking now.

SLIDE 176

Here's the complete nucleosome.

SLIDE 177

Here's the ribbon representation of it.

SLIDE 178

Here's the complete octamer without the DNA,...

SLIDE 179

...and here's the ribbon representation of that. Let's enlarge and center it,...

SLIDE 180

...and remove all the varying N-terminal and C-terminal ends, leaving only the octamer core, which is very regular.

SLIDE 181

Now, as I've said, in several places in this web site, the octamer consists of two tetramers, both of which contain one instance of each of the four subunit types, that is, H3, H4, H2A and H2B. These tetramers join together to form the octamer, according to this "inverted V" relationship between the C-termini of their respective H3 subunits, as discussed in Part I of this slide series.

Let's now re-visit the two tetramers, which I have referred to as "logical" tetramers, and begin to focus on the geometric relationship between them.

SLIDE 182

This is the a-b-c-d tetramer, that is H3[a], H4[b], H2A[c] and H2B[d].

SLIDE 183

This is the e-f-g-h tetramer. If you have a good eye for geometry, you may have already noticed the approximate 2-fold axis of symmetry involving this pair of tetramers. We discussed this axis in a general way in Part I of this series. Now we'll look at it more closely, because it's important to what follows:

SLIDE 184

Here's the 2-fold axis. Understanding this axis is key to understanding the equatorial plane we shall define. We'll have to shrink the octamer a bit to see this.

SLIDE 185

Now let us separate the two tetramers, so that we can look at them simultaneously, and judge their similarity, or lack if such be the case. OK, now we can see both H3-H4-H2A-H2B tetramers clearly; the a-b-c-d chains are on the right, the e-f-g-h chains on the left. Now let's rotate the e-f-g-h tetramer 180º.

SLIDE 186

In this position, I think you can see that the structures of the two tetramers are quite similar. As we're going to see, they each lie approximately in a plane, but the planes are not the same, and, although the planes are approximately parallel, they are not exactly so. Therefore the structures will not be perfectly superimposible, although perhaps we can come close, if we try. Let's try.

SLIDE 187

When I performed this exercise in 3D, using Schrödinger virtual modeling software, I was surprised at how good the fit was. In PowerPoint we can only move in 2 dimensions, yet the fit is still pretty good.

We have thus determined that these two tetramers have a definite spatial relationship to one another, although we cannot yet take advantage of that, to define a direction for DNA chains to travel. We will be able to do that, however, after the next step. It's now time to look at the Histone Equatorial Plane.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 188
Title slide
(No audio)

I. Histone equatorial plane
2. The plane itself

The two "logical" tetramers occupy nearly-parallel planes, which together define a definite orientation for the histone octamer, with likely implications for the directions of DNA duplexes which might bind to the N-termini thereof.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 189

The chance observation of the existence of the histone octamer "equatorial plane" was one of the two key observations which made possible a solution to the mystery of how DNA might bind to the basic-residue-rich N-termini of the eight octamer subunits (the other key observation being the 15 Å spacing between Helix I helices) .

The term "equatorial" is, of course, arbitrary, in that the octamer is not an opinionated human being with a firm conviction as to which part of itself is "up", and which part "down". Nevertheless, I am now persuaded that the view we have been starting most of our discussions from, which I have referred to as the "root position", which is the position depicted in the Luger *et al* crystal structure file, is well-thought-of as an axial view looking down from the equivalent of the north pole of a globe; looking from north to south.

A plane through the "equator" of the octamer, according to this point of view, will divide the structure into a "northern hemisphere" consisting almost exclusively of the 4 subunits of the a-b-c-d tetramer, and a "southern" hemisphere consisting mainly of its e-f-g-h "cousin". This has important implications for the orientation of DNA chains which bind to the N-termini.

SLIDE 190

This view, from which we have started several discussions, shall therefore henceforth be regarded as an axial view, looking down upon the "north pole" of the octamer. This direction of viewing, which is normal to the plane of the computer monitor, corresponds to the z-axis of various virtual molecular modeling software programs I've had the good fortune to be able to use. Somewhere down this z-axis is the equatorial plane, dividing the octamer into two hemispheres, the "northern" one occupied mainly by subunits a, b, c and d, the "southern" one by subunits e, f, g and h.

To see this equatorial plane, we shall once again start by stripping everything away from the octamer core except the most N-terminal of the core α-helices, namely Helix I, and β-Strands 1 & 2.

SLIDE 191

Here's the stripped-down octamer core, showing only Helix I and β-strands 1 &2. We've seen this view before. But now we're going to do something we *haven't* done before. We're going to display the very first, *i.e.*, the *very most N-terminal amino acid residue* on each Helix I, and highlight that residue in yellow.

SLIDE 192

There they are. These eight residues, standing at the extreme periphery of the octamer core, are exactly and precisely what stand between the ordered core structure, and the eight disordered N-termini, the latter of which we are proposing herein to be the main sites of DNA binding. Let's remove the yellow highlights, and color these eight N-terminal residues by atom.

SLIDE 193

Now, before proceeding, just for review, and to keep our bearings, let's re-label the histone subunits,...

SLIDE 194

...and finally, let's add labels for these important N-terminal amino acid residues.

SLIDE 195

Actually, the chemical natures of these residues, as to whether they are basic, acidic, hydrophobic, etc., are less important than their relative positions. Nevertheless, for completeness, let us review them:

For H3 — and this is for *both* H3 subunits — the N-terminal amino acid of Helix I is Lys-64. Look closely at the white-colored [a] chain, and you should be able to discern the form of the lysine side-chain. Lys-64 of the yellow-colored [e] chain is mostly out-of-sight in this view.

For the H4 subunits, the N-terminal amino acid is Lys-31. Both of these lysine side-chains are clearly visible.

For H2A, the N-terminal amino acid is Gly-28. You can see them, but there's not much to see, since glycine has no R-group, and, finally, for H2B, the N-terminal amino acid is Ala-35. Ala-35 for the pink [h] chain is just about discernible, but Ala-35 for the blue-colored [d] chain is on the *other side*, and is barely visible in this view.

By the way, you may have noticed that the Helix I N-terminal residue numbers for 3-out-of-4 of these subunits are about 30, but for H3, the N-terminal residue number is Lys-64, which is twice as high. I've made mention of this length discrepancy before, but we will soon be in a position to account for it.

For now, let's continue with our study of these Helix I N-terminal residues. We don't need the labels anymore, so let's remove them.

SLIDE 196

OK, we're back to our previous view. Let's focus now on the N-terminal residues of the a-b-c-d tetramer. As I said above, it's not so important what the chemical natures of these residues are, as to whether they are basic, acidic, or hydrophobic, but a very interesting fact emerges when we connect their *alpha carbon atoms.*

SLIDE 197

Although it cannot be rigorously demonstrated from this viewing angle, it is perhaps evident, even in this 2-dimensional drawing, that these 4 atoms lie pretty much in the same plane. I stumbled across this quite by accident, and immediately realized its significance for the N-terminal DNA binding mechanism I shall present shortly. The next thing I did was examine the other tetramer, the e-f-g-h tetramer.

SLIDE 198

Not surprisingly, the alpha carbons atoms of the most extreme N-terminal residues of Helix I of the e-f-g-h tetramer collectively defined a similar-looking plane. Were these planes co-planar with one another? The answer proved to be “Not exactly”. But close. Let’s take a look.

SLIDE 199

Here are the two α-carbon-defined planes, looking very much, from this viewing angle at least, as if they are the same plane. But are they? We'll need to rotate the structure to learn the answer to that question.

SLIDE 200

Let us therefore rotate the octamer 90º. I've installed one of my mini-coordinate-axes, in case there's any ambiguity as to the direction of rotation.

SLIDE 201

Since we rotated to the left, *i.e.*, in the negative y-axis sense, the “north pole” of our octamer, which was facing us directly before, is now on the left, the south pole is on the right. From this perspective, we can see several things. First of all, the surfaces defined by the α-carbon atoms of the a-b-c-d and e-f-g-h tetramers are not perfectly flat after all, but kinked. Yet they are close to being true planes, wherefore they *almost* disappear when viewed from the side like this. Secondly, we see that the two tetramer planes are very close to being parallel to one another. Although the deviation from parallelism cannot be precisely defined, because of the complexity of the shapes involved, I think you can see that it’s only a few degrees.

Can we not therefore define an equatorial plane, dividing the octamer into the tetrameric equivalent of two hemispheres; the a-b-c-d tetramer at the “north”, and the e-f-g-h tetramer at the “south”?

SLIDE 202

Here is just such a plane. To the "north" of this purple plane, *i.e.*, the left in this view, is the a-b-c-d tetramer, and to the "south", at the right, is the e-f-g-h tetramer.

SLIDE 203

OK, let’s rewind back to our root position.

SLIDE 204

The object of this exercise, of defining a “histone equatorial plane”, is to create a logical basis for orientation of the binding of DNA to the 8 subunit N-termini, whose locations are now indicated by yellow stars.

If we simply assume, as common sense in any event would predict, that for each of the 8 histone subunits, there will be a similar joining geometry connecting that subunit's Helix I to its N-terminal strand, then, instead of the 8 N-termini pointing every which way, as they do in the currently-accepted model, they could all be oriented in parallel with each other, each bearing a DNA strand or strands, each of which would therefore also be logically oriented with respect to the nuclear space.

That, in fact, is precisely what we are going to assume. We shall assume furthermore that the logical direction of these N-termini and associated DNA strands will be approximately normal to the plane of the computer screen, since that accords with the orientation of the histone equatorial plane, and thereby avoids collisions of DNA strands, either with other DNA strands, or with the octamer core.

What would be the distances between adjacent histone subunit N-termini, in the positions indicated by the yellow stars? Let's find out.

SLIDE 205
Title slide
(No audio)

II. Discovery of the key 15 Å spacing in the histone octamer

SLIDE 206

I said earlier that the discovery of the histone equatorial plane was one of the *two* key discoveries that made possible the elucidation of the new histone structure we are proposing. The other key discovery was that of the location, or I should say *locations,* of the critically-important 15 Å spacings between protein strands, necessary to accommodate DNA tetraplexes, if such association is to be in accordance with the manner of association of protein and DNA found in the sperm cell. I've previously shown that spacing for one corner of the histone octamer only, and I said we'd look at it more closely later. Let's therefore do that now.

What would be the distances between adjacent histone subunit N-termini, in the positions indicated by the yellow stars? Let's remove the stars,...

SLIDE 207

...and consider these dimensions. Remember that the two approximate planes shown were defined by the α-carbon atoms of the most extreme N-terminal amino acid residues of Helix I. I have measured the distances between those atoms. Starting in the upper right-hand edge of the a-b-c-d tetramer, the distance between the α-carbon atoms at the corners of the plane is very close to 15 Å. Proceeding clockwise, the comparable measurement for the lower right-hand edge of the e-f-g-h tetramer is also 15 Å. They're *all* 15 Å!

SLIDE 208

Returning momentarily to protamine, note again that the distance between the P1 and P2 strands, which bind opposite sides of the DNA tetraplex, is just over 15 Å. I predicted that we would find this spacing in the histone octamer, and indeed we have found it.

Let's review what we've determined concerning the application of this structure model to the nucleosome.

SLIDE 209

This is a slide we have looked at previously. It shows, in a cartoon setting, what I consider to be the best candidate for protamine-like structure in histones, with respect to the rotational states of the protein and DNA.

We've now firmly established that the 15Å spacing does indeed exist in the histone octamer. We also know that the highest concentrations of positively-charged basic amino acid residues are found in the 8 histone subunit N-termini, portrayed here for two of the eight subunits by simple light-brown lines connecting to β-strands "borrowed", so-to-speak, from protamine. Now we're going to have to work out the details of the protein-DNA structure on the atomic level. We'll start by returning to our skeleton view of the octamer core.

SLIDE 210

Here again is the skeleton core structure, showing mainly the Helix I helices, and their 15 Å spacings . Let's add back the rest of the octamer core.

SLIDE 211

OK, we're back to our root position, only a little older and, hopefully, wiser. The 15 Å spacings between adjacent Helix I helices is now evident in the setting of the complete octamer core. It's time to discuss the orientations and directions of the N-terminal β-strands that will bind DNA.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 212
Title slide
(No audio)

<u>"Up" vs "down" orientations of Helix I</u>

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 213

There are eight N-terminal β-strands in the histone octamer. These strands are high in basic amino acid content, wherefore we may safely assume that whatever the orientation of these strands, the DNA will follow them. We now know that there is a geometric reference plane,

which I have dubbed the "Histone Equatorial Plane", with respect to which we can define directions of protein strands. Does nature provide us with any *other* clues concerning these directions?

The answer is "Yes". It turns out that the Helix I helices all point either "up", meaning toward the so-called "north pole" of the octamer core, or "down", that is, toward the "south pole". Furthermore, a regular alternation of the "up" and "down" directions is seen, as you travel around the octamer core in root position.

It's convenient, in considering these directions, to think of the octamer as being a square, with a *pair* of Helix I α-helices in each of its 4 corners. Let's look first at the lower left-hand corner of the square, where these orientations are best displayed in this viewing angle.

I think it's pretty clear that the "natural" direction of the green subunit, H2A, is up, that is, pointing, albeit obliquely, away from the equatorial plane, towards your eyes. Conversely, H2B, the blue subunit, is pointing in the approximately-opposite direction.

Now please look at H3 and H4, the white and red subunits at the right upper quadrant. H4, the red subunit, is pointing up, although it's difficult to see at this viewing angle, but H3, the white subunit, is clearly pointing down. We can better-appreciate these directionalities if we remove the other 4 subunits from the picture.

SLIDE 214

OK, now we're looking at the isolated a-b-c-d-tetramer. Recall that these 4 subunits form what we called a "logical tetramer", and that, like all the tetrameric substructures in the histone octamer, it has a 2-fold axis of symmetry. You can see that the directions of the Helix I helices in the left lower quadrant, and those in the right upper quadrant, are indeed related by this 2-fold axis. That which is "UP" on one side is "DOWN" on the other, and vice-versa.

SLIDE 215

Let's look at the up/down orientations of the remaining two quadrants, which are easily determined by simply noting the 2-fold axis between the left and right sides of the complete octamer core, as indicated by the blue axis symbol running down the middle of the slide.

As implied by the axis, the orientations of the pair of Helix I α-helices in the right lower quadrant, now labeled, are opposite those of the H2A and H2B subunits immediately to the left, in the left lower quadrant. Likewise, in the left *upper* quadrant, we see that the yellow subunit, H3[e], unfortunately largely hidden, and the brown H4[f] subunit, have orientations opposite those of the cognate subunits in the adjacent right *upper* quadrant.

Thus, when we consider all the "UP" and "DOWN" directions simultaneously, we learn that they alternate with perfect regularity, as we go around the octamer core.

Let's look at this from a different perspective.

SLIDE 216

Let's return to our "root position", and include everything, specifically the long N-terminal β-strands. We'll rotate it +90º about the y-axis, to take another look at this "up/down" business.

SLIDE 217
(Movie - no audio)

SLIDE 218

Now let's do a small positive rotation about the x-axis, to get a better viewing angle for the two octamer corners we studied in the slides above.

SLIDE 219
(Movie - no audio)

SLIDE 220

Now we can clearly see the path of the red and white H4 and H3 subunits on top of the slide, and the green and blue H2A and H2B subunits on the bottom. See whether you agree with my modeling of the N-terminal β-strands, with respect to the question of whether or not they follow the natural inclinations of the Helix I helices from which they emerge:

SLIDE 221

In order to grasp the meanings of the terms "up" and "down" in this view, we need to insert labels for the so-called "north" and "south" poles of the octamer. Please note that the North Pole, which was on the left in previous slides, is now on the right. I chose this position only because it better displayed the up-down orientations of the strands we'll be looking at.

Starting with H2B, the blue subunit, note that the orientation of Helix I is decidedly "south", or down. Therefore, I have modeled the N-terminus as a linear strand, perpendicular to the "histone equatorial plane" and running *"down"* toward the "south pole".

The green H2A subunit is largely obscured by H2B, but if you look closely, you'll see it in the background, and you'll see that it's pointing decidedly *"up"*, or toward the "north".

The *"up"* orientation of the red H4 subunit is very clear, and, last but not least, we see the *"downward"* orientation of the white H3. H3 differs from the others in that there are extra helices between Helix I and the N-terminal β-strand, but I think the general direction of the N-terminus is well-established by the orientation of Helix I, which is unequivocally "down".

This completes our review of the strand directionality of the [a], [b], [c] and [d] subunits, which comprise the so-called "logical tetramer #1", or the "northern hemisphere" of the octamer. The situation with the "southern pole" subunits, that is, the [e], [f], [g] and [h] chains, is similar, except reversed in direction, because of the octamer 2-fold axis.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 222
Title slide
(No audio)

IV. Summary review of histone structure

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 223

And here is the nucleosome structure which results from application of the principles we have been discussing. This, in fact, is the structure I wish to propose for the nucleosome in its everyday state. I'm about as sure of it as I am of the Kornberg-Luger-*et al* crystal structure for the histone octamer core. If that core structure is correct, then I would assert that the structure shown here, or one very similar to it, is very likely also correct.

If, of course, it turns out that the currently-accepted histone octamer structure is an artifact of crystallization, then I have just wasted an entire year of my life, which is how long it took me to complete this model, *plus* an additional half-year to create this slide presentation.

And here's a summary review of the structure. In this view, which I have variously referred to as "frontal view" or "root position" view, the octamer, with its eight N-termini modeled as β-strands, has a decidedly square appearance. Each corner of the square has a protein-DNA complex which mimics that previously shown to be the likely structure of the complex found in sperm cells, between DNA and protamine.

The DNA is a tetraplex, consisting of two DNA duplexes, each one modeled after the "straight ladder" model of Tai Te Wu, with 6.8 Å base-pair spacing. The two duplexes are then mutually intercalated, which restores the base-spacing to 3.4 Å. Their positions with respect to the octamer core are determined by the positions and orientations of the octamer subunit N-termini, which are modeled as β-strands, pointing alternatively either "up" or "down", as these terms are herein defined.

As in protamine, the protein strands are 15 Å apart, a spacing just large enough to accommodate the DNA, and allow for the formation of 3 Å salt bridges between the positively charged R groups of lysine and arginine, and the negatively-charged phosphate groups of DNA.

At every level, the DNA strands are only paired with protein on one face of the tetraplex; the protein for the opposite face is provided by an adjacent histone octamer, the one either immediately preceding or following the current one along the DNA tetraplex chain.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 224
Title slide
(No audio)

<u>Stoichiometry</u>

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 225

What is the DNA base pair count in this structure? Shown here are two representative corners of the four corners of the histone octamer, only the square shape is gone, because we've rotated from the root position. For reference purposes, so that we may remain fully-cognizant of exactly what it is we're looking at, I'm bringing in a cameo of the root-position octamer on the left. The larger linear structures, shown at the top and bottom of the screen, have been extracted

from the complete octamer and rotated by +90º about the y-axis, to facilitate the performance of a base-pair count.

Each of the DNA complexes shown at the top and bottom of the slide is a tetraplex, consisting of *two* intercalated duplexes. In addition to the +90º y-rotation, we've additionally rotated each of these tetraplexes by an arbitrary amount about their respective *x-axes* in the current view, to optimize each view specifically for the counting of base-pairs.

Let's start with the upper complex, which is a rotated view of what would be the white and red H3-H4 corner of the root position octamer square. If you take the trouble to count the rungs in either of the two DNA ladders showing, you'll see that they contain about 32 bp per DNA duplex. I say "about" 32 bp because there's an ambiguity concerning exactly which base pairs constitute the beginnings and endings of these regions of histone-bound DNA. Moreover, there's no way to know exactly *how* closely packed adjacent octamers will be. The 32 bp count pre-supposes that the next octamer will be positioned so that it almost touches this one, which, as we shall see later, may not necessarily be the case.

If you count the rungs in either of the two DNA ladders in the DNA complex at the *bottom* of the screen, representing the green and blue H2A-H2B corner of the histone octamer, you get either 30 or 31 bp, slightly less than the count we just did in the upper picture of the slide. The difference is that little α-helix which is in the top picture, but not the bottom, which adds a bit of extra length to the DNA-protein complex.

If we take the average number of bp's per DNA ladder to be 31, then it might be thought that each tetraplex corner of the histone octamer square, containing two such duplexes, therefore should be considered to be uniquely associated with ***2*** x 31, or ***62*** bp of DNA. But that's not the case, because in this model, only half the DNA "belongs", so-to-speak, to the octamer shown.

Let's illustrate this for the upper protein-DNA complex, by bringing in the next two octamers that bind this tetraplex of DNA chains. Moving in from the right now is the next octamer over in that direction, and you'll hopefully appreciate that a portion of the upper duplex of DNA belongs to *it.* Now the next octamer in the *other* direction comes in from the left, and you'll see that a portion of the *lower* duplex of DNA belongs to *it.*

OK, let's remove these additional octamers. So, in our new histone-DNA model, we now see, that although there are approximately 62 bp of DNA located in each of the four corners of the histone octamer, only half of them, or 31 bp, are uniquely associated with that octamer, the other half being uniquely associated with one or the other of the neighboring octamers.

SLIDE 226

Here's the numerology again. In our new histone-DNA structure, there are 4 tetraplexes, that is, 8 intercalated *duplexes* of DNA associated with each histone octamer core, distributed as two duplexes in each of the octamer's 4 corners. Each of those 8 duplexes has an average of 31 bp of DNA, so the total bp count for the entire histone octamer is 8x that number, or 248. But since, as we just saw above, only *half* that DNA is uniquely associated with any given histone octamer, the other half being associated with the two adjacent histone octamers, the actual base-pair count, that is, the count of bp's uniquely associated, by salt bridges, with a single octamer core, is half that, or 124.

Relative to the current superhelical ramp model of DNA binding, 124 represents a decrement of about 22 base pairs, or about 28% *fewer* base pairs than are found in the current model, which would be a cause of great consternation to me, except for the fact that the current

twisted model is based upon an *in vitro* crystal structure made from purified, reconstituted histone subunits with added DNA. I reiterate that I greatly respect the superhelical ramp model, and I cannot doubt that anything so *non-random* with respect to the placement and number of basic amino acid residues, has some sort of function in life, but I doubt that that function is the regular storage of DNA, in a form amenable to the everyday working processes of the cell.

I should add that the DNA:protein ratio of our *new* structure could be *increased* by increasing the spacing between adjacent octamers, but I have no basis for assigning any one spacing over any other, wherefore I have pushed adjacent octamers as close together as possible. The subject of octamer spacing in the new model is discussed further below.

SLIDE 227

Here's another chart showing the lengths of the N-terminal β-strands in our new histone model, for each of the four unique histone subunit types. The N-terminus is defined as all the amino acid residues proximal to Helix I, except for subunit H3, in which the N-terminus is defined as everything proximal to the structure we've been calling "Helix Sub-Zero". Also shown in this table is the count of the basic residues Lys and Arg for each of the N-termini, but we're not concerned with that right now. What we're interested in at the moment is the penultimate number in the first column, the sum total of the amino acid residues in the N-termini of all four unique histone subunit types combined, that is, H3+H4+H2A+H2B.

And that number is: **101**. But that is not the total number of N-terminal amino acid residues per octamer, because the octamer contains *two* of each subunit type.

So the total number of amino acid residues in the N-termini of all 8 histone subunits combined is twice as high, namely **202**.

SLIDE 228

Since our histone-DNA structure is based upon the structure of the protamine-DNA complex, I thought it would be a good idea to compare the protein:DNA residue count ratios for the DNA-binding regions of each protein. For protamine, the DNA-binding region is...well, it's the whole thing. For histone, in the present model at least, it is the total count of all the amino acid residues in the N-terminal β-strands. As we just saw in the previous slide, that number, now reproduced in the top of this table, is 202. Also as shown in the slides above, the total number of DNA bps uniquely associated with these same eight N-termini is 124. So the ratio of protein:DNA residue counts is 202/124, or **1.6**.

The comparable number for protamine is **1.8**. The slightly higher number in protamine, if held to be significant at all, would be quite consistent with the presumed lack of metabolic activity in a sperm cell. This is suggested by the fact that most of the basic residues in protamine are arginine, which accounts for more than half the amino acid residues found, and which has the longest and most positively-charged of the 20 amino acid side-chains. I would imagine that the DNA in a sperm cell nucleus would be relatively immobilized by all this arginine.

In contrast, the *histone* N-termini are only about 1/3 basic, and over half of the basic residues are the shorter and less-powerfully charged lysine. This would allow more freedom-of-motion for DNA in the eukaryotic cell nucleus, a place where enzymes are interacting with DNA continuously.

SLIDE 229

Here's a table showing the molecular weight totals of protein and DNA in our histone model, on the left, and comparable data for protamine on the right. There are a lot of numbers here, but the only ones I want to mention now are the molecular weight totals for histone and DNA, which are about 110,000 and 76,000 respectively. This gives a protein:DNA weight ratio of 1.44, which is a bit high. Published figures on the histone:DNA weight ratio vary quite a bit, but they tend to be around 1.

I could say "1.44 is about 1", but it really isn't. I'm not in a position to defend the 1.44 ratio, because, first of all, I have no way of knowing what the spacing of octamers will be in the living cell nucleus. As I've already said previously, for lack of a logically-compelling reason to do otherwise, I have pushed the octamer units as close together as possible, so that the N-terminal β-strands of adjacent octamers are almost touching. But it may not be that way.

Most other histone models, which, of necessity, are based on studies of disrupted chromatin, assume long stretches of "linker" DNA between octamer cores, and I have made no attempt to incorporate that concept; not that I don't believe in it, but simply because I have no logical basis for defining either a length for "linker" DNA, or a basis for assigning some lengths of DNA to linker status, and other lengths to the status of being bound to the octamer's N-terminal β-strands by salt bridges.

Moreover, I have made no attempt to define a role for the histone subunits H1 or H5, whose inclusion could alter the picture considerably. So I post here the protein:DNA ratio of 1.44 as an interim number, to be re-adjusted as further relevant data become available.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 230
Title slide
(No audio)

Volume of the structure

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 231

The picture at the top is an Accelrys Discovery Studio cross section of our structure. In the area calculation shown in the slide, I'm treating the cross section as if it was a perfect square with a side whose length is the average of the 4 sides of this irregular quadrangle.

Keep in mind that Accelrys portrays axial views in perspective mode, so that parallel columns appear to diverge as they approach your eyes. In other words, the 4 columns of protein-DNA, that mark the periphery of the structure, are completely parallel, and the cross-sectional area determination shown would be about the same at any level which might be chosen for measurement.

I suppose I don't have to point out that the 3rd decimal places in the measurements shown are *not* significant! In fact, *none* of the decimal places are significant; even the whole numbers are merely estimates. I am simply reporting the distances exactly as measured by the Accelrys

Measurement Tool. Therefore, I would suggest that the cross-sectional area shown would be better thought of as being "around 7000 $Å^2$".

The length measurement, indicated by the pink arrow in the middle of the slide, is simply the distance between comparable atoms of adjacent histone octamers. As I have emphasized above, I don't know how close together adjacent octamers should be placed. In the current PDB virtual structure file, as posted on this web site, and as portrayed in this slide, they are squeezed together so that the end of one N-terminus *almost* abuts on the beginning of the next one. Also as stated above, had I added "linker" DNA to the model, or tried to incorporate histone subunits H1 or H5 into the structure, the size of the unit cell would have *increased.* But I currently have no logical basis for doing either, so I'm leaving things as portrayed here, pending further information by which we may refine the structure.

The length of the current model is 120 Å. This gives a nucleosome volume of about $8x10^5$, or 800,000 $Å^3$, which I urge you to *not* take overly seriously, but to think of as representing some number in the vicinity of about 1 million cubic Å, comparable to a cube with an edge length of 100 Å.

SLIDE 232

Here's another slide with a lot of numbers. We could go through the detailed math, but why don't we just continue rounding off? The volume of our unit cell, shown here as being around 800,000 $Å^3$, can be rounded off in our minds to around *1 million* cubic angstroms, which is equal to *one millionth* of a cubic micron (*i.e.*, 10^6 $Å^3$ = 10^{-6} μ^3).

Taking the standard figure of 6.4 billion bp (6.4×10^9 bp) for the human diploid genome, and using "124", the number derived in the slides above, as the number of DNA bp's uniquely associated with each histone octamer in our new nucleosome model, we come to the next important number, namely, the {number of nucleosomes per human genome}, according to our model.

That number, which is simply the total genome bp count divided by 124, is here given as 5.16×10^7, which we might as well round off to "about 50 million". So what's the total volume of 50 million nucleosomes, each 1 millionth of a cubic micrometer in volume — you don't need a slide rule for this! — we're talking about a total nucleoprotein volume of something like 50 cubic microns (50 μm^3). If you actually go through all the math, you get the more exact figure of about 41 cubic microns, which is just about the same as the total nucleoprotein volume predicted by the current Kornberg-Luger *et al* crystal model, the latter of which is based upon DNA binding to the superhelical ramp.

We can now ask: What percentage of nuclear volume is consumed by the histone-DNA complex? Well, that depends upon the size of the nucleus. Published values, however, differ wildly, and, depending upon which one you choose, the percentage of nuclear volume occupied by histone and DNA is anywhere from 2-41%.

At the bottom of the slide is the comparable statistic for protamine and DNA in the sperm cell, from my 2006 publication on that subject. I showed at that time that the protamine-DNA complex occupied a mere 10% of the available space in the sperm head. Even taking into consideration the fact that the sperm nucleus is *haploid,* containing only half the full diploid complement of DNA found in somatic cells, it is entirely evident that even the full diploid amount of 6.4×10^9 bp, if hypothetically complexed with the highly-basic protamine, would require no more than 20% of the volume available in a sperm head.

Interestingly, that number, 20%, is about the midpoint of the range of values given one line above, for histone and DNA in somatic cells. We can therefore confidently conclude that our non-helical DNA model will fit very nicely into the available nuclear space.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 233
Title slide
(No audio)

<u>What madness causes people to believe that a helical twist is *necessary*, if DNA is to fit into the available nuclear space?</u>

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 234

This brings us to a topic I have been wanting to discuss for many years. Molecular biologists are positively obsessed with the absurd notion that DNA must be secondarily twisted, then tertiarily supertwisted, then quaternarily twisted yet again, if it is to fit into the tiny nuclear space. I proved, in 1976, that the known topological properties of circular DNA, not only do *not* require the presumption of helicity, but rather are explained *better* by a non-helical DNA model. That paper was not, however, accepted for publication until 2002, because the whole world, from the layman to the Nobel laureate, is in love with the double-helix, and even scientists refuse to give it up; even in the face of powerful evidence against it.

The idea that native plasmid DNA is supertwisted, because such supertwisting is somehow necessary to pack the DNA into the available space, has been emblazoned into the minds of several generations of molecular biologists, even though the topology simply doesn't favor that theory. Now we see, in both the protamine-DNA complex in sperm cells, and the histone-DNA complex in somatic cells, that a *non*-helical DNA model occupies only a small fraction of the nuclear space. *Who says* that twisting, supertwisting, and quaternary twisting on top of the supertwisting, are necessary for DNA packing? And *why* do they say it?

If you simply cannot grasp the point from pure molecular biological considerations, please then consider a macroscopic analogy. Libraries have a book-packing task not very dissimilar to the DNA-packing task of the eukaryotic cell nucleus. Libraries need to compact thousands of volumes into the smallest possible space. If this is done in a *rectilinear* fashion, the compaction is logical and effective. This slide illustrates an admittedly melodramatic, but nevertheless I'd say *appropriate* view of this subject. Of course, library books are not stacked on their sides, but, in the upright posture in which they are usually found, the principle is the same.

Here, however, is another packing mode for books which you'll *never* see in any library. What librarian, desiring to efficiently shelve hundreds or thousands of books, would pile them up in a helical array like this? Why, the very notion is absurd.

Here's another conformation for book storage you're not likely to see soon at any library in the real world. It looks a bit like the histone octamer, but without the rectilinear N-termini we have added.

The point I'm trying to make is that DNA, like books, is best stored in the Wu straight-ladder conformation, which facilitates both efficient DNA packing, and the everyday

biochemical tasks of life, such as reading of the base sequence by enzymes, and strand separation during replication.

Ok, I've gotten that off my chest. It's now time to proceed to our last subject, namely higher-order chromatin structure. This starts with the 30 nm fiber.

||

SLIDE 235

Title: **The 30 nm fiber**

The remainder of this slide presentation is a bit problematical. Before embarking on a discussion of the 30 nm fiber, and on the development of a higher-order structure for entire chromosomes, I must reveal at the outset that this was not part of my original plan. I never intended to look at any higher-order structure than the nucleosome itself.

When I took a cursory look at these higher-order structures, however, I was rather surprised to find that our new histone model lent itself so readily to the development of models for higher-order chromatin structure, that I could hardly resist the temptation to look further.

In the case of the 30-nm fiber, I set out expecting to find very little, because the traditional histone structure found in all textbooks has not led to any compelling theory of 30-nm fiber structure at all, but only vague models, with no experimental support for any of them.

But with the new histone structure presented in this slide series, the tables were turned. Not only did it readily suggest a 30-nm structure, but *several* such structures, so that the problem changed entirely, from one of coming up with any plausible model at all, to one of grappling with a plethora of models, all plausible, all competing for our attention, and between which we must now choose!

It was the same with chromosomes. I had no plan, in the beginning of these studies, to look at chromosome structure at all. But once I had settled on what I regarded as a good model for a 30-nm fiber, I found that it almost effortlessly extended itself, according to electron-microscopically observed 300-nm and 700-nm folding patterns, into an object with the known dimensions of a chromosome.

I said at the outset that this section was "problematical", yet all I've told you thus far is how *well* the investigations went. "So what's the problem?", you might be wondering. The problem is that my temporary license to use Schrodinger virtual modeling software expired in June of 2013. I spent the subsequent 6 months putting together these slide presentations, and it was during that period that I discovered the potential of the new histone structure to generate higher-order chromatin forms. That actually introduced *two* problems: The first was that all my research on higher-order chromatin structure had to be done using graphic models, not virtual molecular models, because I no longer had access to virtual modeling software.

The second problem was that, in the course of investigating higher-order chromatin structure, I encountered some relatively minor, but still alarming *imperfections* in my current histone model. But I cannot revise it now, because I have no software even remotely capable of doing that. So the revisions will have to be done by someone else with more money than I have, since the cost of virtual molecular modeling software is $100,000 and up; way outside my budget, which is literally zero.

If the subjects of these slide presentations interest you, and if you're also interested in the histories of things generally, you can click the link on this page which reads "Software Nightmares", which will take you to another page which relates the terrible difficulties I

encountered in my efforts to obtain even temporary access to software for the determination of both protamine structure, back in 2005, and histone structure now.

Having now warned you about the difficulties and limitations of the following studies of higher-order chromatin structure, let us nevertheless delve into it.

SLIDE 236

Here's the axial view of our somewhat-asymmetric nucleosome structure; the same view we looked at previously, except now with most of the lengths of the eight N-termini cut off, to minimize the splaying effect of perspective view. Also, we have finally rounded off the dimensions to eliminate the totally-insignificant decimal places.

The structure is a rather irregular quadrangle. The electron microscopist, however, would not be likely to discern this geometry, but only an amorphous cylinder-like structure, whose dimensions would then be measured.

SLIDE 237

Aha! So what we're seeing here is a structure with a diameter of about 100 Å, or 10 nm, which is, in fact, the next level up in chromatin structure, having the so-called "beads on a string" appearance, which has a 10 nm diameter, and is punctuated periodically by [what are presumed to be] histone octamer cores, giving it the appearance after which it is named. Our new structure could certainly give a sort of "beads on a string appearance" also, although there are some spacing issues we would need to address:

SLIDE 238

Here's a typical EM photograph of single strands of nucleoprotein, revealing the "beads on a string" appearance. The spacing of the nucleosome cores is difficult to define precisely; in various published studies, it seems to range broadly from about 150-350 Å. These spacings are larger than the spacing in our model, which is a mere 120 Å. This might, for me, have been another cause for consternation, but if you look at this picture you'll see immediately that there are problems with the preparation of the material.

First of all, it's clear that the DNA has been tremendously degraded; even in this single small electron micrographic field, we see a free end, which merely illustrates what would otherwise be self-evident: that a 6.4×10^{9} bp genome cannot be isolated intact, but will be badly fragmented, regardless of the technique used to extract it.

I'm going to dare to propose that these pictures all show individual strands of duplex DNA, which, once released from the tetraplex nuclear environment, immediately assume the most stable known conformation for the unconstrained DNA duplex, namely the Watson-Crick right-handed helix. Why? That is, if, as I believe, the new histone structure I have proposed is thermodynamically favorable, why should it disappear merely because the cell is disrupted?

The answer I would suggest is that, while the new proposed structure does exist in a free energy valley, it has a high free energy barrier to formation. To speak more simplistically, it is as highly-ordered and *non-random* as a structure could possibly be, requiring, for its stability, that all components be present and properly arrayed. If any part of it is disrupted, the rest

quickly follows, as the individual parts all scramble to assume the conformations they are accustomed to when existing separately. For DNA, this is surely the Watson-Crick double helix. This then wraps itself around the octamer core, exactly as it does when nucleoprotein is reconstituted in the laboratory, giving essentially the structure portrayed in the Luger *et al* crystallographic study.

The spacing between octamer cores will then be more-or-less random, which explains the large discrepancies seen in different electron micrographs. This spacing, however, may perhaps be influenced by the positive charges of the N-terminal strands, which will repel each other, giving a spacing of twice that shown in my new histone model, as illustrated by the following demonstration:

SLIDE 239

This slide starts with our new structure, with octamer cores spaced at 120 Å, as described above. After the electron microscopist extrudes the entire nuclear contents from the cell, which probably causes total disruption of the highly-ordered native structure, then the protein N-termini, their positive charges being no longer neutralized by the negative charges on tetraplex DNA, and they furthermore being unable to establish comparable charge interactions with twisted duplex DNA, will repel each other, giving rise to a spacing in the neighborhood of 240 Å, which is in fact the general magnitude of the spacing seen in the "beads-on-a-string" electron micrographs.

You may be wondering: Do I *really* believe that this is the explanation for what we see in these electron micrographs? The answer is, I don't *really* know. I only know that, from my perspective, this explanation is far less a stretch of the imagination than the alternative belief, which to me is the very height of absurdity, namely that DNA can be twisted, supertwisted, then quaternarily twisted, so that the entire cell nucleus becomes one massive Gordian knot, and yet still replicates not only flawlessly, but with astonishing speed. *That* I find to be utterly beyond belief.

SLIDE 240

If we are going to construct a 30-nm fiber from our new nucleosome model, we shall need to define exactly, or at least approximately, how adjacent 10-nm nucleosome strands will associate with one another to build up larger structures. As I mentioned in the introduction to this section, the model, in its current form, does *not* lead to a single, logically-compelling structure for the 30-nm fiber, but rather a set of competing structures between which we must now adjudicate.

They all, however, employ the same two mechanisms of stabilization which we found previously in the protamine-DNA complex. These are:

(1) The square-array pattern of salt bridges between basic amino acid side-chains and DNA phosphate groups, and...

(2) Classic β-sheet-type hydrogen bonding between the peptide backbones of adjacent β-strands.

The simplest example of a possible 30-nm fiber structure is the one which can be made from 4 nucleosomes:

SLIDE 241

Here's our most simple model, although it's not one I greatly favor. Nevertheless, I'll show it, because it illustrates some general principles. Chief among them is that the diameter is about 30 nm. This can be seen at a glance, since each octamer structure is about 10 nm in diameter, and, in this model, the space in the middle is also about 10 nm. You don't need a ruler; just look at it. The individual octamer units are about 10 nm, and the conglomerate structure, which has a diameter of 2 nucleosomes plus a nucleosome-sized space in the middle, will measure out at just about 30 nm.

SLIDE 242

In this, and in all the other candidates for the 30 nm fiber that I'll show you, there are two types of interactions that can potentially impart stability to the structure; both of these interactions were also found in protamine:

First, there are those rectangular arrays of charge interactions which we've already seen previously, involving the positively-charged Lys and Arg residues in the N-terminal β-strands, and the negatively-charged DNA phosphate groups. Keep in mind that you're looking down axially at a cross section of a very, very long strand, and that these charge interactions will be found at every level.

Secondly, there are peptide backbone hydrogen bonding interactions between adjacent β-strands; exactly the sort of interaction found in β-sheets such as keratin.

These are precisely the interactions we saw by way of preview, in the previous section entitled "Likely model for N-terminal structure".

Let's look a little more closely.

SLIDE 243
(No audio)

SLIDE 244

Please note the square charge array involving two positive charges from lysine or arginine, diagonally-situated from one another, and two negative charges from DNA, also diagonally-situated. Let's bring in a little box to mark these charges. This is precisely the square-shaped charge array we "borrowed", so-to-speak, from the protamine-DNA model, and is one of the key features of that structure. The principle here is quite the same; adjacent columns of protamine β-strands and DNA tetraplexes can align as shown, with the respective positive and negative charges lying at the corners of a square.

Since the diagonals of a square are 1.41 times the lengths of the sides, and since charge interactions vary with the square of the distance, the attractive forces in the sides of these squares are approximately twice the magnitude of the repulsive forces at the diagonals, making this a very stable structure where it can be found.

Although the basic residue content of histone is less than that of protamine, it remains true that in the histone-DNA complex, at the N-terminal regions of each of the 8 histone subunits, the fraction of basic residues is still quite high; about 1/3 of the residues being either Lys or Arg, so, although there will be fewer of these square charge arrays than are found in protamine, they will, wherever they are found, stabilize the structure.

In the protamine P1-P2 heterodimer, because of the high arginine content, a large percentage of arginine residues in P1 were fortuitously adjacent to arginine residues in P2, giving rise to many of these square-shaped charge arrays. In histone, however, the number of protamine-like square charge arrays will be lower, involving only about 1/3 of the basic residues; if you do the math you'll immediately realize that, approximately 2/3 of the time, basic residues in one N-terminal β-strand will ***not*** fortuitously align with any of the basic residues in an adjacent β-strand. In other words, since an average of only about 1/3 of the N-terminal amino acid residues are basic, then the chances that any one of them will align with a basic residue from an adjacent histone N-terminus are 1/3 x 1/3 = 1/9. Therefore, most of the basic residues will not be part of a perfect square array of charge interactions, but rather...

SLIDE 245

...part of a square array containing one positive charge only, the other corner being occupied by a non-basic residue, most likely a hydrophilic residue such as serine or threonine. I can only depict that schematically, because there's still a lysine residue in the background, but what I can do is represent the hydrophilic R-group in the square array by the letter "H" for "hydrophilic". This hydrophilic R-group may associate by hydrogen bonds to either of the adjacent DNA phosphate groups, while the lone positively-charged lysine or arginine side chain, opposite it in the square, can still form a salt bridge by sharing its positive charge with either of ***its*** adjacent phosphate groups.

SLIDE 246

The other species of interaction between adjacent 10 nm nucleosome strands is a classic sort of hydrogen-bonding between the peptide backbones of their respective N-terminal β-strands, shown here in axial cross-section. We cannot see the hydrogen bonds in this view, however. We're going to have to rotate to a longitudinal view.

SLIDE 247

This slide shows a length of nucleoprotein encompassing two adjacent histone octamers and their accompanying DNA. The N-terminal β-strands of the white H3 subunits, chain [a], are on the bottom of the nucleoprotein complex .

SLIDE 248

Now we introduce a 2nd length of nucleoprotein, identical to the first, except for orientation and position. For this one, the N-terminal β-strands of the yellow H3 subunit, chain [e], are pointing up, toward the 1st nucleoprotein complex. What we want to do is bring them

together so that their β-strand peptide backbones can hydrogen-bond to one another. Since H3[a] and H3[e] are antiparallel, the hydrogen bonding should be ideal.

The β-sheets are represented here by simple lines. Just so we can keep firmly in mind what those lines represent, here's a reminder. To bring about this sort of hydrogen-bonded interaction, all we need to do is to approximate the two nucleoprotein complexes:

SLIDE 249

This is approximately the correct ribbon spacing for 3Å hydrogen bonds between the antiparallel peptide backbones. We have thus joined this pair of nucleoprotein complexes together via hydrogen-bonding of their N-terminal β-strands.

In all the higher-order chromatin forms we shall look at, the basis of the interaction between adjacent structures will be either the type of β-sheet interaction shown here, or the salt bridge interaction shown in the previous slides.

SLIDE 250

Because these hydrogen-bonding β-sheet interactions are absolutely critical to everything which will follow, I must, before leaving this subject, admit to a bit of sleight-of-hand I have been compelled to execute. As I mentioned at the beginning of this section, in the creation of higher-order chromatin structures, I had to work with graphic models, because I no longer had proper virtual modeling software at my disposal.

The "Software Nightmare" link is still at the bottom of the slide, if you wish to know how this unfortunate state of affairs came about.

When I arrived at this juncture, I learned that my Helix Sub-Zero was partially blocking the pathway of formation of the β-sheets I showed in the previous slide, requiring that I create a space to accommodate it.

SLIDE 251

With the space, as you can see, the octamers fit together perfectly.

The trouble is, I have included, on this web site, a set of downloadable PDB files for this new histone model, and, as I have pointed out several times previously, in the model, I arbitrarily positioned adjacent histone octamers to be as close together as possible.

SLIDE 252

In other words, if you download the PDB structure files, this is what you'll see: The N-termini of adjacent octamers will be almost touching, and there will be no SPACE such as was seen in the slides above.

SLIDE 253

The irony is that by introducing this space, I've added a very short stretch of "linker DNA", which I said previously my model did *not* have. It's a very small amount of linker DNA, and I also admitted previously that I have no logical way of establishing an exact longitudinal octamer spacing in the first place. In other words, the spacing shown here may, in fact, be

correct. But it's still a slightly different spacing than that shown in the downloadable PDB file. It's just unfortunate that I had to grapple with these issues using graphic software instead of virtual modeling software.

Well, there's nothing I can do about it now, so I'll have to proceed with the 30-nm fiber model, and with the chromosome model that arises from it, in the presence of this slight but annoying ambiguity about the exact spacing of adjacent octamers. Qualitatively, this will have essentially no effect on any of the logical conclusions I shall draw, but quantitatively, with respect to such issues as stoichiometry, a 5-10% new uncertainty has now been introduced. In actuality, even this won't really change things much, because in dealing with higher-order chromatin structure we are, of necessity, invariably dealing with large degrees of approximation anyway.

SLIDE 254

Let us return now to our 4-nucleosome model for the 30-nm fiber. These are the sites of square-array salt bridges, and these are the sites of peptide backbone hydrogen bonding. I don't like the hole in the middle of this structure; I've heard that 'nature abhors a vacuum', and that's how I see the large empty space. I can't imagine that it would be a tunnel for enzymes to slide down; it's much too small for that. So let's look at a few other candidates for the 30-nm fiber.

SLIDE 255

Here's a 30-nm structure which fills the space in the 4-nucleosome model by adding 2 more nucleosome units, one in the center, the other in a pentangle of nucleosomes surrounding it. It's got minor alignment problems which I have not bothered to correct, because I find the pentangular structure somehow unpersuasive; perhaps it's just prejudice against pentangles in general. But next I'll show you a hexangular structure which has much to recommend it.

SLIDE 256

Now here's a structure which has compelling attributes. The cross-section, shown here, is not exactly cylindrical, but it's 30 nm wide, and about 20 nm in height. Like all the subunit monomers, dimers and tetramers which comprise the nucleosome, this structure has a 2-fold axis of symmetry, parallel to the z-axis, that is, normal to the plane of the monitor screen, so that the three lower nucleosomes can be thought of as arising from a 180º rotation of the three upper nucleosomes about the axis shown. A similar structure could be made by performing the 180º rotation about the x-axis instead of the z-axis

Please note that as you move your eye from left to right, there's a slight vertical offset, with each successive pair of vertically-stacked nucleosomes displaced upward a bit *(i.e.*, in the +y direction; toward the top of the screen). This offset allows for both excellent and extensive square-array charge alignment, as described in the previous slides. Additionally, the six pairs of adjacent β-strands along the horizontal midline have excellent hydrogen-bonding potential, notwithstanding the minor modeling problem I alluded to earlier.

Nevertheless, I am not at all suggesting that this should be accepted as ***the*** structure of the 30-nm fiber; but merely that it's about the right size, and that it has attributes which would more than suffice to explain its stability as a unified structure; more so than any of the other structures

I've looked at. We will also see, In Part III of this PowerPoint series, that it readily adapts itself to the 300-nm and 700-nm foldings which have been observed in studies of chromosome structure. Therefore, I shall tentatively presume, for the sake of continuing our investigation of higher-order chromosome structure, that it be accepted as a reasonable working model of the 30-nm fiber, for the time being at least.

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 257
Title slide
(No audio)

VII. *What are* the identities of the DNA strands in the histone octamer?

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

SLIDE 258

Our new histone octamer model has 4 quadrants, each of which contains a pair of mutually-intercalated DNA duplexes. That's 8 DNA duplexes in all! *What are* these DNA strands? Are they simply 8 strands of unrelated DNA being arbitrarily mounted on a single histone octamer? Or are they ultimately connected to each other? Or, is it even remotely plausible that they represent DNA from different chromosomes, particularly from *homologous* chromosomes?

With respect to that last possibility, let me say that I have devoted a great deal of time to the consideration of the hypothesis that the pairs of DNA duplexes, at each of the four corners of our new nucleosome structure, are from homologous chromosomes. Why even bother thinking about such a thing? For one reason only: it would vastly facilitate genetic recombination events, and would, if true, imply that throughout the entirety of genetic history, homologous genes have never actually been very far apart, except during metaphase.

SLIDE 259

While I am almost certain that such a thing is not beyond the realm of topological possibility, I have rejected the idea anyway. In the first place, there seems no doubt that the cell is quite adept at juxtaposing bits of DNA that are far apart in nuclear space, wherefore the cell needs no help from us clumsy molecular modelers to accomplish the task of recombination.

Secondly, although the topological problem of separating the homologous strands may not be impossible to explain, that doesn't mean either that it's easy. In fact, it would be quite a nightmare. Finally, the conceptual *simplification* of genetic recombination, by the permanent juxtaposing of homologous genes, would come at the price of great *complication* of the replication process, since before the chromosomes were even replicated, they'd first have to be separated, *then* replicated, then painstakingly put back together after cell replication was complete. The more one thinks about this scheme, the less attractive it becomes.

After long deliberation, I have concluded that the DNA content of a single nucleosome, in the final analysis, is most likely a single DNA duplex winding back and forth, then continuing

on to the next nucleosome in the 30-nm fiber. Let us now see how this might work, and how it might lead to a plausible model of chromosome structure.

SLIDE 260

In each quadrant of this structure, a DNA double-non-helix extends upwards for some long distance, as represented semi-schematically by this not-so-long segment.

SLIDE 261

When this segment reaches the proper point, it will curve around,...

SLIDE 262

...which I have, perhaps somewhat clumsily, represented by this curved ladder cartoon...

SLIDE 263

...then the strand simply goes back the way it came, giving rise to the tetraplex of mutually-intercalated duplexes. In Part III of this slide series, I will show you how this structure can give rise to an entire chromosome, by being readily extended according to the 300 nm and 700 nm folds which have been observed.

SLIDE 264
Title slide
(No audio)

Primary Structure

1. What lengths should we assign to the β-strands at the octamer subunit N-termini?

SLIDE 265

Structural considerations always begin with *primary* structure. Therefore, if you ever decide to get involved in histone structural studies, you're going to need to have this primary structure page, or one like it, showing the amino acid sequences of the four subunit types H2A, H2B, H3 and H4. There will not be many days in which you will not have to resort to this page at least once, if not several times.

The amino acid sequences shown here are from www.Uniprot.org, the "Universal Protein Resource". To the right of each sequence is the URL from which it was obtained.

H2A: http://www.uniprot.org/uniprot/P06897
H2B: http://www.uniprot.org/uniprot/P02281
H3: http://www.uniprot.org/uniprot/P02302
H4: http://www.uniprot.org/uniprot/P62799

I have my own color code, as you can see. The N-terminal and C-terminal strands, which, in our new proposed histone structure, are ordered as straight-ladder β-strands, are depicted in a dark blue. Alpha-helical sequences are red. This allows me to focus quickly on Helices I, II and III, as well as on the small additional α-helices, when such are present.

The *internal* β-strands connecting the long, hydrophobic Helix II at its N-terminal end to Helix I, and at its C-terminal end to Helix III, are colored a baby blue.

The diagram also shows the starting and/or ending points of the H2A and H2B chains, insofar as they appear in the Luger *et al* crystal structure, all 4 of which chains were degraded in that study. The corresponding starting and ending points are not entered for H3 or H4, probably because I encountered no issues there.

The most frequent issue that *did* arise, regarding primary amino acid sequence, involved subunit H2B, which, in the Luger PDB crystallographic model, is missing 3 residues from the N-terminus. Since the Luger N-termini for both of the H2B chains; chains [d] and [h]; are truncated proximal to lysine-27, I have no way of knowing which three proximal amino acid residues are missing from the PDB structure. There is therefore a 3-amino-acid numbering discrepancy between the Luger H2B sequence and the Uniprot sequence shown here. At times, this discrepancy was maddening in the extreme, wherefore I would say that an amino acid sequence page such as this one, while absolutely necessary, may not, by itself, be sufficient to keep your house in order. But it's a start, at least!

SLIDE 266

Here's a geometrically-accurate graphic representation of the amino acid sequence data in the previous slide, employing the same coloring scheme we used there.

SLIDE 267

In the beginning, it was primary structural considerations that led me to the conclusion that the N-termini of the 8 histone subunits, by virtue of their high basic amino acid content, were the likely sites of DNA binding in the normal resting state.

The term "Normal resting state", in the sense that I'm using it, refers to metabolic inactivity. It may also refer to routine reading by polymerases, which I'd imagine is not seriously disruptive of the resting structure. Of course there's no way to know that with certainty.

In special circumstances, however, I can also readily envision the DNA performing acrobatic feats involving the superhelical ramp, as such is portrayed in this current textbook model of nucleosome structure. I envision this as occurring in a manner such as that presumed to occur with the lac repressor.

But in the normal resting state, I would expect to see this, with the DNA back at the strongly-basic N-termini.

SLIDE 268

Once one dares to breech the integrity of the published histone structure, by remodeling the N-termini into β-strands for the binding of DNA, another question soon arises: *How much* of each subunit should be considered “N-terminal"?

The anatomical regions of each subunit that are part of the so-called “superhelical ramp” also have a high content of positively-charged basic residues. Should we perhaps unwind some or all of those as well, and add them to the N-terminus as part of a larger DNA-binding β-strand?

SLIDE 269

Here's histone subunit H3, chain [e]. If we wanted to increase the length of the N-terminal β-strand, we could start by unwinding the next α-helix in the C-terminal direction, which, for this particular subunit, would be the structure, un-named in the Luger *et al* PDB file, but which we have dubbed "Helix 0". As this somewhat crude movie shows, if we were to unwind Helix 0, we could thereby double the length of the N-terminus, which would double the length of DNA which this subunit could bind.

If we wanted to continue this exercise, we could next unwind Helix I.

SLIDE 270

In fact, if we thought it appropriate, we could unwind *all* the peripheral helices, leaving only Helix II to form a small hydrophobic core for the octamer, as suggested schematically by this drawing. This is the same graphic representation we saw a few slides up, except that, instead of left-aligning the 4 subunits, we have aligned them according to Helix II. Note that the central portions of all 4 subunits align quite well, and that it's only in the periphery that they differ markedly.

The question we're asking is "What would be wrong with a model in which only Helix II remained in the α-helical conformation, and everything else was unwound to the β-strand conformation?

Let’s add some helical twists for Helix II, but let’s leave everything else straight, *i.e.*, in a β-strand configuration. This greatly increases the DNA-binding potential of the histone octamer, to as much as 2-3x what it is in the current, largely-α-helical model. Why should we NOT do this?

SLIDE 271

I spent quite a bit of time investigating this question, and came to the conclusion that almost everything in the octamer core should be left as-is. Here are five pictures illustrating the sorts of reasons why I cannot dismiss the many α-helices that comprise the core of the currently-accepted histone octamer model. We’ve reviewed most of this previously. In short, everywhere you look in this octamer, wherever subunits are contiguous, you see very non-random appearances of bulky hydrophobic residues, giving rise to so-called “hydrophobic bonds”, and, present also but to a lesser extent, a number of ionic bonds, all of which things work collectively to cement the structure together as an octamer of the shape to which we have become accustomed.

Several examples are shown here. In the upper left, we see the hydrophobic bonds between the C-terminal ends of H3[a] and H3[e], considered to be critical to the very existence of the octamer structure itself. Next over to the right we see the same sort of picture at all four of the β-bridges of the same structure. To the right of that is the strongly-hydrophobic alley between Helix 0 of H3, chain [a], and the adjacent C-terminus of H2A, chain [g].

In the second row, to the left, are the hydrophobic bonds between the C-termini of H4 and H2B, and, last but not least, in the lower right-hand corner we see a hypothetical tetramer consisting of a pair of H2A-H2B subunits, described in Part I of this slide series, which, although hypothetical, nevertheless perfectly well illustrates the total picture of hydrophobic and ionic bonding between the subunits of any of the histone tetramers.

Should we regard all this as mere coincidence? I think not. There's simply too much order, too much symmetry, and too many fortuitously-positioned hydrophobic bonds and salt bridges, to dismiss it all as mere coincidence. Therefore I conclude that all the α-helices of the octamer core crystal structure should be respected, and that the modeling of DNA-binding β-strands should be limited to the currently-unordered N-termini of the 8 histone subunits.

SLIDE 272
Title slide
(No audio)

Primary Structure

2. How can we solve the length mistmach between the N-terminal β-strands of H3, and those of the other 3 histone subunits?

SLIDE 273

Here again is our previous graphic representation of the four histone subunits, aligned by means of their Helix II α-helices. In this alignment we can clearly see both the similarities, as well as differences, between the subunits. There are 7 distinct anatomical regions common to all, namely the N-terminus, Helix I, β-Strand 1, Helix II, β-Strand 2, Helix III, and the C-terminus.

The 5 central regions, from Helix I to Helix III, are fairly constant in length from subunit-to-subunit. But the N- and C-termini differ markedly in length. For reasons given previously, we have elected to ignore the C-termini for now. We shall have no such luxury with the N-termini, however. Let us then begin to examine them.

The first thing we notice is that the N-termini of 3-out-of-4 of the subunits; H2A, H2B and H4; are about the same length; 27, 37 & 30 amino acid residues respectively. As I have pointed out previously, if we were to dismiss the first 10 residues of H2B, which have little DNA-binding potential due to their preponderance of proline, then we would have an even greater consistency in N-terminal lengths for these three subunits, as all would then contain approximately 30 residues. But look at the forth subunit: H3 has 63 amino acid residues; *that's more than twice the length of the others!* This length discrepancy is so large as to be positively

disconcerting. No attempt to model the N-termini for DNA binding can succeed unless this terrible size-mismatch is resolved.

SLIDE 274

To begin to solve this size-mismatch problem, we'll need to start with a side-by-side comparison of the 4 histone subunits. I've arbitrarily chosen, for this comparison, the four which comprise the tetramer I previously referred to as the "southern hemisphere" of the histone octamer, namely H3, H4, H2A and H2B, chains [e], [f], [g] and [h] respectively. I've lined them up so that Helix I is approximately co-linear for all four subunits in this view. What I want to show you is that 3-out-of-4 of these subunits have an extra α-helix in their N-termini, proximal to Helix I. I'm not aware of any formal terminology for these extra α-helices, so I shall call them by the terminology I introduced in Part I of this slide series, namely "Helix 0", since they're closer to the N-terminus than Helix I, and ought therefore be numbered with a number less than "1". Well, the only number less than one is zero, wherefore I call them "Helix 0", for lack of a better name.

Helix 0 of subunit H2A, the turquoise subunit on the left of the slide, has a length of only 5 amino acid residues, 2 of which are arginine. Because this helix is too short to have any obvious structural significance, and because the arginines are presumably there to bind DNA, I have arbitrarily, but not quite "capriciously", elected to treat the α-helical structure in this particular Helix 0 as an artifact of purification, wherefore I have opted to unwind this short helical twist in our new nucleosome structure, and re-model it instead as a β-strand; continuous with the rest of the N-terminal β-strand of this subunit.

Although it remains possible that this small α-helix is really *not* an artifact of crystallization, especially in view of the fact that the two arginines are 3 residues apart...(here comes the amino acid sequence now, flying across the screen)...note that those 2 Arg residues, if presumed to be part of an α-helix, are potentially well-aligned with respect to the possibility of charge interactions with an adjacent linear structure like DNA, because their spacing in the primary sequence would put them on the same side of the α-helix.

That's an argument one could assert in favor of leaving these residues in the α-helical conformation, ***but*** if this α-helix is ***not*** unwound, then there will be an additional size-mismatch between the N-termini, because this one, which is *already* the shortest of the four subunit N-termini, will then be even shorter. So my decision, for now at least, is to disallow this small α-helix in our structure, and rather to portray these 5 residues as part of the contiguous β-strand.

H2B, the pink subunit next to H2A, is alone among the histone subunits in having no additional α-helices in the N-terminus proximal to Helix I.

Next is H4, the brown subunit. H4 does have Helix 0, but it's pathetically-small, having a length of only 3 amino acid residues, their sequence being Ile-Gln-Gly. I have dismissed it as an artifact of crystallization, and, as I did in the case of H2A, I have accounted its 3 amino acid residues as being merely an undistinguished 3-amino acid segment of the N-terminal β-strand.

This brings us to H3, the problem subunit with the immensely-long N-terminus. It has Helix 0, but in this instance, I decided to preserve it as an α-helix in the final structure. Why? Several reasons. First of all, it's far too long to just ignore; 12 residues in length, extending from Thr45→Lys56. Secondly, it lies adjacent and parallel to the C-terminus of H2A, as illustrated by the next slide.

SLIDE 275

This slide shows what appears to be a very significant hydrophobic alley between these two subunits. Look at the amino acids in this alley, listed at the bottom of the slide. Could this possibly be a coincidence? As we saw in Part I of this series, this is exactly the sort of hydrophobic subunit-subunit interaction which is responsible for the structure and stability of the histone octamer generally, and I would therefore be very hesitant to dismiss it as mere artifact.

SLIDE 276

Secondly, there's the matter of the excessive length of the H3 N-terminus. Allowing for residues 45-56 to remain in the subunit core as "Helix 0", binding hydrophobically to adjacent H2A, reduces the length of the H3 N-terminus from 63 residues, a length which is quite impossible for a modeler to deal with, to 44 residues. This is *still* significantly longer than the other three N-termini, which have average lengths of only around 30, but at least it's a major step in the right direction. The final step in solving the H3 length problem turned out to be yet another helix, shown in the next slide.

NEW SLIDE 277

I have had the audacity to *add* a new α-helix to our structure, one not found in the published crystal structure of the nucleosome. With respect to terminology, please note again the positions of Helix III, II, and I, as well as the additional helix from the published crystal structure which I have dubbed "Helix 0", and now I've introduced yet a fifth helix, and, as there are no numbers lower than zero, I have, in desperation, dubbed this one "Helix *Sub*-Zero". I know; it sounds like a refrigerator brand. Well, that's what I call it, anyway.

The complete amino acid sequence of the H3 N-terminus is shown at the bottom of the screen, in which Helix Sub-Zero has been underlined, with its basic residues highlighted in green. It extends from Ser28 to Arg42, and connects to Helix 0 (which is indicated by the red letters in the amino acid sequence) by a short 2-3 residue strand. The green letters in the underlined "Helix sub-zero" sequence are Lys & Arg residues which can potentially bind DNA, as we shall see momentarily.

This structure was readily created by the Schrödinger/Maestro Build panel, then adapted to our new histone octamer. By what authority do I add an α-helix where none is currently known to exist? This I do in accordance with three lines of argument: (1) It corrects the length of the H3 N-terminus, without which correction there is no possibility I can see for re-modeling the histone octamer N-terminal sequences, (2) it accords with the natural direction and spacing of adjacent subunit H4; a subject briefly discussed in the previous section entitled " 'Up' vs. 'down' orientations of Helix I", and (3) it has the very fortuitously-spaced salt bridges shown here. Lys-36 and Arg-40 from Helix sub-zero, and Arg-49 from Helix 0, are so well-situated with respect to binding electrostatically to the DNA in this model, that it is hard for me to believe that they are there by mere coincidence.

I am strongly inclined to believe, therefore, that Helix Sub-Zero is real, and not merely a figment of my imagination, or an act of wishful thinking.

SLIDE 278

The yellow H3 subunit shown in the previous slide was H3[e], from what I have been referring to as the "southern hemisphere" of the histone octamer. Here's the comparable Helix-Sub-Zero structure for H3[a], in the "northern hemisphere" of the octamer. I'm showing you this primarily because there's an error in it.

I've already admitted that there are a few relatively minor imperfections in my new histone structure model, and I explained earlier about the software access problems which prevent me from repairing the imperfections now. One of them involves Arg-49, located here. Keep your eyes fixed on the guanidinium group as we add the DNA tetraplex. It looks partly obscured now, does it not?

SLIDE 279

Here's a closeup of the problem. This is a detail from H3[a] Helix 0, at the start of the loop leading to “Helix Sub-Zero”, showing the guanidinium group of Arg-49 being “pierced” by the phosphate group of DNA chain I, thymine-16. I’m fairly certain that this is an arginine rotamer problem, easily corrected, but I no longer have software with which to make the correction. So I guess someone else will have to do it.

I am bringing the error to your attention because I know that among the people who look at this new histone structure, will be experienced virtual modeling scientists, and I don't want them to find the error, and imagine that I missed it.

SLIDE 280
Title slide
(No audio)

Steps in the creation of N-terminal β-strands that bind DNA:

- **Creation of β-strand templates; role of proline.**
- **Creation of DNA tetraplex templates, correction of close contacts in 2006 structure.**
- **Geometries at the 4 corners of the histone octamer.**
- **Selection and editing of lysine and arginine rotamers.**

SLIDE 281

This is the protein motif from which the histone N-terminal β-strands were created. It is borrowed from the protamine-DNA structure file. The slide shows part of Fig. 4 from the 2006 J Theoret Biol article on the subject. The criteria for this motif are described in the accompanying slide show on this web site, entitled "The probable structure of the protamine-DNA complex".

To summarize briefly, the spacing between DNA base pairs is presumed to be the idealized distance of 3.4 Å, wherefore the spacing between amino acid residues on either side of

the β-strand must necessarily be twice that distance, or 6.8 Å. This distance is attained by a β-strand with the approximate angles $\psi = +130.5^{o}$, $\phi = -130.5^{o}$, which also places the structure into the most-favored quadrant of the Ramachandran Plot for β-sheets. You'll notice, however, that the angles shown here are a bit different. That's because setting ψ and ϕ to exactly $\pm 130.5^{o}$ results in a structure which is very slightly curved.

A perfectly straight template was created by cloning a single amino acid residue from an idealized ψ and ϕ $\pm 130.5^{o}$ polypeptide, transposing it exactly 6.8 Å along the y-axis, and re-attaching it to the parent residue. The new ϕ and ψ angles at the junction between the members of the new dipeptide were slightly different than $\pm 130.5^{o}$, wherefore minor adjustments to the structure were made, eventually resulting in this motif. The final dipeptide can now be cloned to any length, without meandering from the longitudinal axis of the structure.

SLIDE 282

This is the 228-residue β-strand template used to make all 8 histone octamer N-termini in our new histone model. The Schrödinger-Maestro "Mutate" panel was used to establish the respective 8 amino acid sequences.

It should perhaps be specifically noted that the "Mutate" module does not introduce a kink to the backbone when proline is inserted in the place of a previously-existing residue. Although it is perhaps unrealistic to expect no alteration at all to the polypeptide backbone in the presence of proline, I do believe that in our new histone model, the backbone orientation will be determined by DNA, and not importantly altered by the occasional proline residue.

What then is the teleological purpose of proline in the histone octamer at all, we might ask? I have always suspected that proline occurs for the specific purpose of slightly *un*-aligning, or straining the backbone, to discourage the formation of biologically-inert crystal-like structures. If protamine is to be taken as a model, then its amino acid sequence, which is 50% arginine, is perhaps the benchmark basicity for DNA-associated protein in a relatively-metabolically-inert storage environment.

With respect to the location of proline in the histone octamer, there seems to me to be a trend for proline to occur in areas with particularly high concentrations of lysine and arginine residues, of which the N-terminus of H2B is a striking example. We've looked at this before; of the first 10 residues, an astonishing and inexplicable 4-out-of-10 are proline; surely not a formula encouraging any obvious protein-DNA structure. Yet the remainder of the N-terminus contains an equally astonishing preponderance of basic residues, numbering over 50%; 14-out-of-27 basic residues to be exact. What is the meaning of this? I don't think anyone can answer that question at the current time.

SLIDE 283

In histone subunits H2A, H2B and H4, proline is found at, or close to the extremities of Helix I or Helix III, where these α-helices connect to the subunit N- and C-termini. In this location, proline may be contributing to the transition from core α-helices to N- or C-terminal β-strands. It *might be significant* that there are small-to-medium-sized additional α-helices in the N- and/or C-termini of these three subunits. It has occurred to me that proline might be there to discourage the extension of the α-helical conformation from the core structure into the N- and/or C-termini, wherefore proline could then be alleged to be an "α-helix breaker".

This is not a totally compelling argument, however, because this pattern is not literally followed in H4, shown here, and even less so in H3, which I haven't even included in this slide.

More details concerning our β-strand protein structure can be found in the accompanying slide show “Probable structure of the protamine-DNA complex”, or, if you possess virtual modeling software, you can download the various protein structure files on this web site, to your own computer.

SLIDE 284

Now to our DNA structure.

This is a slightly-tilted view of the 64-bp tetraplex template used for all 32 DNA strands in our new histone-DNA model. This too was originally derived from my 2006 protamine-DNA structure, wherefore the sugar-phosphate backbone bond lengths, bond angles and dihedral angles may be found in the accompanying slide presentation on that subject, on this web site, to which the interested reader is referred.

Although the sugar-phosphate backbone has been copied exactly from my 2006 protamine-DNA model, the remainder of the structure has been slightly modified. A recent re-examination of my original model revealed that there were two alarmingly close hydrogen contacts per nucleotide residue, both purines and pyrimidines. These close contacts have now been lengthened by use of the Schrödinger Maestro "Minimization" module. Neither the sugar-phosphate backbone nor the bases were included in the minimization, yet even with only a fraction of the atoms per nucleotide residue being included in the minimization, the close contacts were nevertheless improved significantly.

SLIDE 285

For those interested in the details of these close contacts, they are all included in this table. It may be generally stated that in the 2006 protamine-DNA structure file, there were two close hydrogen-to-hydrogen contacts per pyrimidine and purine residue, averaging about 1.6 Å, and that the Schrödinger minimization increased the distances to about 2.0 Å. While that is still a short distance, the preferred hydrogen-to-hydrogen distance being closer to 2.4 Å, it nevertheless represents a significant improvement, since the van der Waals repulsions between the atoms increase exponentially as the distance decreases.

SLIDE 286

Having described the DNA structure, let us look at its mode of electrostatic association to the histone N-termini. To begin with, I wish to say a word about the peculiar geometries at the corners of the octamer, as viewed in this frontal position.

Although it became clear to me, at the outset of these studies, that the DNA belonged *approximately* at the four corners of the octamer, adjacent to the N-terminal ends of Helix I, there has never been any guideline for *precisely* positioning the DNA.

There are two H3-H4 corners, namely the H3-H4 [a] and [b] chains at the upper right, and the H3-H4 [e] and [f] chains at the upper left; likewise, there are two H2A-H2B corners, namely H2A-H2B chains [c] and [d] at the lower left, and H2A-H2B chains [g] and [h] at the lower right.

As discussed previously, in the section entitled " 'Up' vs. 'down' orientations of Helix I", each of the pairs of Helix I N-termini at these corners have what we may call a “natural” direction in the crystal structure, and, having no other recourse, I followed that natural direction as closely as possible. That accounts for the rather incongruous directions of the H3/H4 corners compared to the H2A/H2B corners, where the former seem to be aligned approximately with the x-axis, while the latter lie along lines which seem to emanate radially from the center of the octamer core.

It's possible that in real life, *i.e.*, at 100% humidity in a living cell nucleus, this entire structure would be significantly different from this in its small details, but we won’t know what those differences will be until we develop either *methods* to peer into the living cell nucleus, or advanced deductive methods to know what’s going on there without actually looking.

It's also possible that these corner orientations *are* exactly correct, since they do in fact facilitate the two types of octamer-octamer binding interactions we've described previously, namely the square-shaped array of salt bridges, which are associated with the radially-arranged corners, and the antiparallel β-strand to β-strand backbone hydrogen bonding, which is associated with the corners that are x-axis-aligned.

Once settling on a general structure at the four corners of the octamer, the spacing of the DNA and protein is determined by the establishment of salt bridges between the basic amino acid residue R-groups and the DNA phosphates. Whereas in protamine, most of the basic amino acids were arginine, in histone about half are lysine, so the configurations are a bit different.

In histone, the arginine side-chains cannot be fully-extended, as they are in protamine, because if the histone subunit N-termini are positioned by setting the Arg-to-DNA salt bridges to 3 Å, then lysine residues will be too far away from the DNA, namely 4 Å, due to the relative shortness of the lysine R-group. Optimizing the spacing for lysine works better, since the arginine-DNA salt bridges can still be set to 3 Å, by simply curving the arginine side chains a bit, which is what we have done.

Let's focus our attention on one of the 4 corners, to see exactly how lysine and arginine R-groups anastomose with the negatively-charged phosphate groups on DNA. For these purposes, we’ll look at the lower right-hand corner, namely the H2A[g]-H2B[h] subunits:

SLIDE 287

Here’s a cross-section of the N-terminus and associated DNA from the lower RH corner of the octamer, where H2A and H2B, chains [g] and [h], reside. This is actually a composite graphic construct, because unlike the case with protamine, there are few places in the histone octamer where a single cross-sectional slice will give this many basic residues at the same cross-sectional level.

The top of the slide shows lysines 11 and 12, according to the residue-numbering conventions in the main downloadable PDB structure file on this web site. These two lysine residues have been moved from their original location. Their positions with respect to DNA, however, are entirely correct, because the DNA structure is the same at every level.

The bottom of the drawing, where the basic residues are in fact in their actual locations, shows a cross-section containing two adjacent lysine residues, lysines 27 and 28, which themselves are adjacent to two arginine residues, arginines 29 and 30.

Each rotamer was initially selected from the Schrödinger-Maestro rotamer library, but in most cases, they had to be modified, which means, of course, that the new rotamers had to be

carefully checked for intra-R-group clashes, as well as clashes with surrounding structures. It's always better if you can use a pre-fabricated rotamer, but in this project, that was usually impossible.

The model has been constructed so that each salt bridge is very close to 3 Å, as indicated by the yellow bars.

The difference in the mode of electrostatic bonding between protein and DNA, for arginine and lysine respectively, is best-demonstrated at the lower left-hand corner of the slide, where it may be seen that the lysine R-group does not extend laterally as far as arginine. In the lower *right*-hand corner, you can get some idea of the curvature of the arginine side-chain, which is necessary to facilitate the formation of a 3 Å salt bridge with DNA, in a structure optimized for the shorter lysine R-group.

SLIDE 288

Well, this completes our discussion of our new histone structure, which places the DNA at the N-termini of the 8 histone subunits, which is, after all, where the preponderance of basic, positively-charged amino acid residues are located. If you look carefully at the two instances per revolution where your view is totally axial, you'll see the histone octamer structure as currently portrayed in textbooks. Then, as it rotates beyond the axial view, you'll see what the textbooks have thus far left out.

If you'd like to see how readily this new structure gives rise to a 30-nm fiber which, in a 10-million-fold amplified model, takes 6.4×10^9 bp of DNA from a cable so long it extends almost the length of Manhattan Island,...

SLIDE 289

...then, by merely following the 300-nm and 700-nm folding patterns which have been reported in EM studies of chromosomal DNA, takes the entire structure down to a neatly-folded package, a mere city block in length,...

SLIDE 290

...then please take a look at the next and last installment in this series, entitled "Histone Structure, Part III. A possible new structure for chromosomes".

I hope you enjoyed this presentation. Thanks for watching.

SLIDE 291
(THE END)