Dahlem Workshop on Organization and Expression of Chromosomes

Closing Remarks - F. Crick

I am not going to make a summary -- who would be so rash -- but I did think it would be sort of fun if I made some personal remarks... by which I mean not remarks about persons, but remarks which are personal to me, as it were, and for which nobody else has to be responsible. I'll try and not go fast so that people can interrupt. If I say something terrible, hold up your hand.

First of all I would like to give you my own personal reactions to one or two general areas. Of course they are a little extreme. When I make some critical comment, it doesn't mean to say that people aren't trying hard, but that though a few interesting results have come out, we are little clearer than we were a year before.

Let's start off with pure genetic methods, emphasizing the word "pure", and my assessment of that at the moment is that firstly, they are slow, and secondly, they are mainly ambiguous. This is just to say that when you want to make a really hard interpretation, you need some biochemistry.

Now hnRNA, as my notes say, is still a mess, and I think that's a fair summary. RNA polymerase -- little progress, but at least they are facing up to the problem, and that's something. Transcriptional complexes -- still a mess, nothing definite, all contradictory, getting nowhere at the moment, but maybe it will come.

Pure hybridization -- I think that's been useful, but my belief is that it is getting well past the point of diminishing returns, and I wouldn't be surprised if you feel the same. I don't mean hybridization as a tool in individual cases -- there it is very useful. But I think to explore the genome, we are getting towards the limit there.

Non-histone proteins: all one can say is that serious study is only beginning. There is a realization that they fall into categories, and the methods have proved decisive over the past few years, of course, and we may hope to see advances but it's still very much in the polymerase stage.

Those are my general comments so far, all right? Then I thought, well what are some of the areas of rapid progress, if any? Now by rapid progress I mean two things: that the people working in the field can't do good experiments fast enough. In most areas you don't quite know what to do next, experiments are not very good, you are stumbling along. But an area of rapid progress is usually one where you just can't make up your mind to do this, or that, or the other, because they are all good experiments and they will all give you good results.

The other thing is that an area of rapid progress is one in which you get firm conclusions, which are often unexpected. As opposed to people who do sloppy experiments just to get the results they think they are going to get. Now, are there any areas of that sort? Of course I'm prejudiced, but my belief is that there are two such areas. They may be limited areas, but they are making rapid progress. One is the characterization of the nucleosomes, and the other I think I could describe as cloning.

In those two fields, as far as I know, it is not possible to get through the experiments quick enough, and the results when you get them are reproducible in other labs, or so good that people don't query them. Why are these fields so active? Because in both cases they are studying, one in the sense of primary sequence and the other of tertiary structure, well-defined entities, and these are being studied by more methods than one. The tools are good.

If you have a method which is very laborious and the assay is very slow and so on, it isn't repeated at other labs, you can't get though things -- this isn't true in the two

fields I mentioned. So let's take those two areas and see where we have got to and what we have to do next.

Going back first to the nucleosome, let me say straightaway that there are many unanswered questions, but there are some things which, though you have to be careful about the wording, are definite. It does look as if the 140 base pair core particle is an interesting, well-defined entity. (I'll weaken that in a minute.) It appears to be universal; nobody has found an organism so far which doesn't have it. It contains most of the DNA, at least probably 50%; this is not a trivial matter and we won't talk about the other things it doesn't contain. And it has got a sufficiently regular structure that it can be crystallized. The first people who produced crystals were the Russian group; they have published photos of the crystals; they are small; they are by refrangence and so on (we have recently done this). I would just like to show this one slide just to show you the degree of regularity.

The point I want to make is, that five years ago that would have been inconceivable. If somebody in histones could say we have a thing which is 50% in chromatin and it is so regular that we can make a structure which is as regular as you see here, it would have been unbelievable. So we can clearly see, whatever your reservations about nucleosomes, that a major advance has been made. And let me say that the primary credit really has to go to Huisch and Begoin for opening up the field. We are not here to allocate credit, but I must mention it because they were two young relatively obscure workers in Australia, way out of all the big labs, and without that discovery of theirs we could still be blundering along. Once they'd opened it, then it was easy -- all we had to do was follow.

I will now put in a personal note, because I had to bully Aaron Cook to get this picture out of him, and I mustn't interpret it because it is so new (the spacing is about

110 Å and it is hexagonal, as you saw) that I am not allowed to say what the interpretation of it is. I could give you a learned lecture on all the artifacts and how it might be this, that and the other, but instead I will give you my guess as to what is going to be the answer to that particle.

I think it is going to be approximately spherical; it will probably have two turns of DNA in spite of the counterevidence from the amount of superhelicity; it is going to be a fairly regular fold, I think; and Struther Arness has convinced me by arguing the opposite that it is going to have kinks in it, because having listened to his arguments I can't think of another way to explain the data. Now we won't discuss when it's going to be sold and whether it can be done and so on -- there are many questions about that, and many about the uniformity of the particle, the effect of modifications, etc.

Is it allosteric? You have heard of course that it can be split into two, but it is obviously going to have a dyad axis, and anything with a dyad axis in sufficiently extreme conditions can be split into two. Equally, all allosteric proteins have monomers, all right? So we could have more subtle variations that merely splitting into two. All we can say is, we are at the beginning of all that, but we have got a well-defined entity.

Against that, the repeat between the centers of the nucleosomes under 200 base pairs is really a mystery. It has been argued that it might be an artifact of the extraction of the nuclei, since it cannot be shown to be there in the intact nuclei. We don't know if getting the nuclei out of the cells is producing the spacing. We don't know whether this spacing is trivial -- if it really matters to the cell what the spacing is. The other extreme view is that it is

very significant. Let me explain the lines along which people are thinking. There is a strong suspicion that the 140 base pair particle is not too accessible to proteins, is not easily attacked by micrococcal nuclease (it is of course snipped a little bit by the DNAase I). So there is a sort of feeling if maybe you are going to recognize a base sequence on it, it is going to be difficult, as opposed to just snipping the backbone. Consequently it has been not unreasonable to think that that's covered up. What isn't clear is whether the space between the 140 base pair particles is covered up, so it is obviously more accessible to enzymes. The nice question therefore is, is it accessible to non-histone proteins -- and then people build elaborate theories to say, since you've got a stretch of 140 which you can't get at and another 160 which you can, depending on how you phase the nucleosomes you can get all sorts of things in regulation. That is the sort of idea which is being discussed, and I won't elaborate on it. You could easily write down the theories yourself, given this starting point. The answer, of course, is that it isn't known. I'll come back to this topic in a moment.

Now we come to cloning. Here what is clear is that I am entirely in agreement with what I believe is the attitude of the Stanford school, and I myself would say that I don't think there is an adequate substitute for Drosophila. There is nothing I am going to say which I don't think they are fully aware of, that is, the need for more clones, the need for biochemistry in translating the messages and so on, the need for some Swiss sequencing (?) as opposed to just cutting up the restriction enzymes -- but the point I would emphasize really is that in order to show the sequence works you have got to have mutation, and whether you can do that by nibbling or dilution will depend on the situation. Sometimes you can and sometimes you can't.

We in Cambridge in thinking about genetics think of it in terms of chess terminology. We think it's always fun to

try genetics in the opening game -- you may get somewhere, you may not. Genetics in the middle game is a very big question mark. But genetics in the end game is almost always essential to prove anything. You've only got to look at the things that Wally Gilbert does, say, to see that you really can't get away without genetics if you want to get a firm proof at the molecular basis.

So it will be a combination of working on precise elements, i.e. the things that you clone, by a lot of different methods, the biochemical ones and the genetic ones, which is going to get us somewhere -- although notice that it is all being done on primary structure so far. That isn't the thing we have to worry about. So I haven't any serious doubt that that is going to be the major way into the problem. We have all felt this way at Cambridge for some time, and I know they have at Stanford, and I suspect many of you feel the same.

It doesn't mean to say, of course, that people shouldn't try and develop genetic systems and then try and get a clone off them, as in say alcohol dehydrogenase. It would be very important to do that.

This leads us to another topic, which is, how important is the tertiary structure. This is very, very difficult indeed, and it isn't clear how we are going to make progress more than the tiny bit that is being made. As many of you know, what we believe is that the nucleosome thread, the chain of nuclosomes, folds up into what we call a solenoid or what could be called a supercoil, and you see these rather poorly in EM pictures (which I won't show), and it is compatible with the X-ray diffraction in fibers. But I must point out that both of those could be artifactual; the exact pitch and diameter of the helix could have been destroyed by any of these methods. So not only have we got to characterize them as we get them out of the cell, but we have got additionally

to characterize them within the cell. So we have quite a long way to go on that.

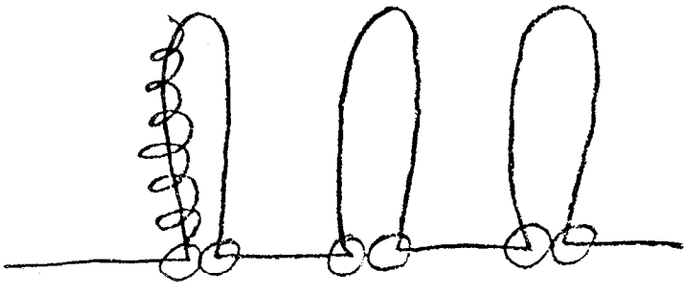
Now, when you build models like that, the thing you are tempted to do is think, "What comes up must go down", so what happens? The Russian group for years, Vashevski and Georgiev in particular, have been drawing models that look like this, with it coming down on the outside. I used to tell Georgiev, put it down on the inside -- symmetry would dictate that that is a more balanced type of structure. The very obvious question is, is there something on the inside? If I may have the next slide ... This is the present Russian suggestion along those lines, that there is some sort of chain of nucleosomes going up like that, and a chain of DNA, the so-called naked DNA, the DNA you don't have with the H1, going down the middle. Now whether the structure is really like this, or is more collapsed, or collapses when we see it, we don't know. But it turns out to be an extremely important question (you will find the same sort of thing is being suggested by Aaron Cook in a paper which is in press) to know whether that so-called naked DNA is an artifact or not.

We are going to have a lot of trouble getting at this answer, not the least of which is that we believe that even the mild extraction procedures we have at the moment may damage the structures. We have to go to even milder ones. After all, these things are quite big -- the things we have are 20,000 base pairs long, and they would be 2,000 Å long typically. We are getting up to the size where shearing due simply to osmotic breaking of the nucleus can do damage, let alone what happens in compaction. So essentially you have to do things like open the nuclei very gently, get them out, characterize them in solution without drying them by scattering methods, and make sure you don't have a lot of RNA synthesis going on at the same time.

The next thing I want to touch on is, are there any new developments. What I said about nucleosomes and cloning, of course, is not especially new, though you don't have to be very perceptive to see that that's where the action is at the moment. The question is, where is it going to be next year in relation to these two? There is one topic which was just touched on, and this is loops or domains.

Two groups of workers have published things on this, and Wersome has given colloquia on it, in which it is claimed (and I think rightly) that there are domains of supercoiling in the Drosophila cell line he is using as there are in E. coli. This leads to the obvious hypothesis, which is supported, I understand, by his own published data but certainly not established, that in Drosophila there might be one loop per band. Now we are talking about not how it is folded up into a solenoid ... (I'll do it on the board).

Here is the DNA; it's going up like that; it's got no protein on it at the moment; it's continuous; and it's held together by some complex of protein or protein-RNA, so that this bit of DNA is near this bit of DNA.



The simplest version of the hypothesis says that this coils up with histones and all the rest of it to form one band (and this is an interband), and this coils up to form the next band.

This is not a new idea (I am told that it is in the latest version of Jim's textbook, and you will certainly find it

in an article by Saucer and somebody in the Cold Spring Harbor thing). So it's not new, but my belief is that it is likely to be right, and that moreover this is one of the lines that has to be pursued. You have to ask yourself whether you can isolate these things, for instance, and can you characterize what is holding things here. They do have the advantage that they are very stable under salt -- he uses 1 Molar NaCl and thidium bromide to push the histones off. That is comforting, because if it were very labile we would have difficulties finding it.

Consequently, my own feeling is that the next breakthrough is going to be this question of these loops on this scale -- that sort of scale would be 20 to 30,000 base pairs in Drosophila.

This leads to an obvious hypothesis that a solenoid essentially is equivalent to a loop. When you fold this thing up, what you are going to get is a "spring" of nucleosomes going round like that, and then <u>possibly</u> it goes back with DNA in the middle, and these two bits of DNA join, and then you have a bit here, possibly with some nucleosomes in more open structures than SV 40, and then another solenoid like that. Many variants of this are possible -- I've just given it to you as an example.

Do remember that when we do a light projection with micrococcal you would say that will cut here, and what you will get out is those things -- I mean, they've already been seen in our electron micrographs! Though of course we haven't proved mutual (inaudible) of a thing like that, it's a nice experiment. If you then do a tRNA removal, what you should see is a bit of naked DNA, and then you should see a length like this, probably with some nucleosomes, and the ratio of this to that total length should always be the same. If that model is correct. So we could do some very definitive experiments which, if they work, would become convincing.

We aren't in the horrible situation where we can't even think how to begin; with a little luck one can see how to do it.

So if you ask me what is likely to be the next hot subject, I think it is going to be the loops. And obviously we would like to know what is holding the loops here, and we have a nice lot of elaborate models in Cambridge for this. And what process is involved, is it always the same process? Can you eventually recognize the sequence? If this model is true, then (inaudible)..... , but that means that on your plasmid you should be able to say where this point is, and this is going to be a matter of extreme importance.

Now things are starting to get wild. So far what I have said has been fairly sober; I would be surprised if I am grossly wrong. The question now is, can we take it one stage further? Of course there are a lot of technical questions about this structure, and we would like to know where the H1 is. But what we really want to know is, is this structure an inert one?

You can easily make an argument in higher organisms, you've got to cover up most of the DNA because you don't want to have a lot of non-specific binding and you want to leave open some of the control sites -- I mean that is a pretty straightforward argument following from the sort of thing you want to get in the lab. Now, is this such that in fact that there are no proteins sitting on it in that form, or possibly they might sit at the end here, but never mind that... Maybe all this body of stuff is so nicely tightly wrapped up here in possibly a regular structure... Now I don't know whether it is, and you may be skeptical about it and that chromosomes are built at that level in a regular structure, but how many of you would have thought that nucleoseomes existed five years ago? So be careful, all right? There may be a regular structure there.

So then you would say that this is clearly where there are going to be genes in coded sequences, as we know from the histones. And we also know, I think, that it is plausible that once you've got going, you have multiple promotor sites, when you've opened things up, with your puffs -- in the lampbrush case you can actually see it. But how do you get things going? It may be that there are three states -- the inactive one, or let's call it the preactive one is the first.

To make this preactive you have to start either here or here, somewhere outside this inert structure, and transcribe it a bit. The mere act of transcribing will certainly loosen the structure, whatever the transcription complex is. Therefore in order to get it such that this can begin to work, maybe you have to transcribe all that, and maybe that's what the hnRNA basically is. Of course it may be used for other things afterwards. The hnRNA may be sort of an unwrapping device, basically. Never mind whether some of it has messenger on the end and all the rest of it.

Once you've got to the preactive state, which means that you are transcribing with the messenger RNA, then maybe it becomes just like E. coli -- you are exposing the promotors here, you may be using this one faster than the rest of it. So maybe the secret of this thing is that it has been wrapped up in an inert package; you've got to loosen it, and that explains the hn RNA; and after that it's E. coli with a few variations.

Now, we are in hot trouble about establishing all this, and this is where the plasmid business is going to be difficult, because if we are going to test it we want this to go along the lines of previous things and we want to establish what the structures are (they don't have to be exactly as I've drawn them; they can be more complicated). Obviously if we want to get the primary sequences in the pure form, what we

want is this in the pure form.

There are two ways -- one is reconstruction and the other is isolation. Both are going to be difficult. Reconstruction is going to be quite tricky -- you have to ask how you fold all this up, and there are probably lots of minor proteins that help, etc. It's not going to be an easy job to make sure that if you have a plasmid DNA bit that you can reconstruct that inert structure, if it exists. But equally, it isn't very easy to see how you're going to isolate it. Maybe you'll try to see if you can get it with the tails off, and get this bit here... So if tertiary structure matters, we are going to have problems later on, because sooner or later the only way to get it is to get a well-defined gene with a tertiary structure. That's a long way away, but that is what you will have to try for in the end.

Now, I don't want to pretend that all genes are going to be the same, or the same in all species, because this is clearly not true. But we do have a right to assume that there is some general principle of organization which is similar in many different species -- that is what we are trying to find. The best way to find it is to get specific examples and work it out with them beyond doubt, and we can worry about generalizations later. So I'm not worried by people who say, well what about notch and this that and the other, since that's for the future. For the moment let's have one or two well-defined genes characterized. We know all the sequences in the primary sense; we're trying to get the tertiary structure; we can worry about the generalizations later.

I think that is most of what I wanted to say. It seems to me we have things in a fairly interesting state. At least some branches of the field are moving forward, and I don't think there is any bit of the field which isn't moving at all. Of course what we hope is that the bits that are advancing rapidly will feed into the ones which have got stuck

a bit, and produce something in the way of... I won't say "models", but at least plausible theories. My experience in such matters has always been that unless you can see through to the right idea, you can't see which experiments are wrong. Once you suspect an experiment is wrong, then you can really go and prove it's wrong, rather than just smearing somebody else's work, say. Then you make progress. The hope is that one part of the subject will feed into another to such an extent that you can say, well now we see that it was bad luck with such and such and so on.

Take the genetic code -- once we had the genetic code, we could see that some of the data was wrong. It was repeated, and it turned out to be clearly wrong. So that's the hope how things will go. On the other hand, we have to realize that it's a very complicated subject.

I have said nothing about the biological side of the subject, because first of all I read again the biological end of the report, and they cheated by going back and putting in all the biochemistry! They haven't adjusted themselves to all the problems of how you make tissues and all that sort of thing. The signals within cells, etc. I'd left that out because it didn't seem to me to be the essential gravity of the meeting, but we have to realize that even if we'd solved all of this for Drosophila, we'd have many, many problems left in development. I could give examples in many insects where you get the strangest things happening that have very little to do with this. So you have problems left at the developmental level.

On the other hand, I must say the technical breakthroughs have really been decisive. I think that without genetic engineering, where should we be? Without the new methods of characterizing protein fractions and the gels -- hopeless. We have been very lucky in having these technical advances, and it would be nice if we had a few more.