Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci.,
Vol. 8, Nos 3 & 4, August 1985, pp. 739–745. © Printed in India.

Mapping the structure of macromolecular assemblies by combining chemical modification and separation methods

C. R. CANTOR, R. BENEZRA, C-F. HUI, C. L. SMITH and J. WELSH
Department of Human Genetics and Development, Columbia University College of Physicians and Surgeons, 701 West 168 Street, New York, New York 10032, USA

Abstract. Several examples will be described in which powerful separation methods are combined with relatively simple chemical modification techniques to provide structural information on complex macromolecular assemblies. Ribosomal RNA structure has been examined by crosslinking, separating individual crosslinked species by gel electrophoresis, and enzymatic methods for determination of crosslink positions in the nucleotide sequence. Chromatin structure has been examined by footprinting the location of individual nucleosomes by a combination of chemical nicking and DNA separations. Virus structure can be examined by using breakable crosslinkers analyzed with diagonal gel electrophoresis. Ultimately such methods may allow structural information to be obtained on systems even as complex as whole chromosomes.

Keywords. Chromatin; virus structure; ribosomal RNA; crosslinking; conformation; electrophoresis.

Introduction

The structures of complex biological assemblies such as ribosomes, viruses, and chromosomes defy analysis by most high resolution analytical methods. Usually the principle of divide and conquer must be employed to deal with such structures. One takes advantage of the biochemist's ability to fractionate the structure into individual protein and nucleic acid constituents. The primary structure and approximate secondary structure of the nucleic acids can then be determined in a relatively straightforward manner. In this communication we will demonstrate several approaches to determining elements of the higher order structure of macromolecular assemblies. All of these approaches take advantage of the increasing power of electrophoretic separation methods to fractionate nucleic acids, and the relative ease of obtaining cloned DNA probes as reagents for analysis or separations.

Ribosomal RNA structure

Crosslinking agents such as psoralens will react with folded RNAs or the RNAs within intact ribosomes. Identification of the crosslinked residues provides direct evidence that they are located in close proximity within the RNA or the ribosome. In most past work, the position of crosslinks has been determined by denaturing the crosslinked RNA and examining it by electron microscopy (Wollenzien *et al.,* 1984). The crosslink produces a looped RNA chain configuration that is easily detected and quantitated in

the microscope. The position of the crosslink can be determined to within around 20 nucleotides. Analysis of a number of different crosslinks at this resolution is sufficient for providing a rough picture of the overall chain configuration of the RNA. However, this resolution is not sufficient to test detailed models of the secondary or tertiary structure.

We have developed a relatively simple method for the analysis of RNA crosslinks by using modern ladder sequencing techniques (Hui and Cantor, 1984). The major strategy of the technique is illustrated in figure 1. An RNA molecule containing a single crosslink is prepared by very mild treatment with psoralen. The RNA is hybridized to a cloned DNA fragment that overlaps one site of the crosslink. A single stranded bubble must form in the DNA near the site of the RNA crosslink because the bases involved in the RNA crosslink cannot possibly be free to base pair to the DNA. Single strand specific nucleases are then used to cleave the DNA strand at the site of the bubble. The length of the resulting DNA fragment indicates the position of the crosslink. This method can be used to find the approximate crosslink position by starting with a relatively long DNA fragment. Once the approximate position is found, the exact position can be determined by using a shorter DNA probe and analyzing the length of the resulting fragments by DNA sequencing gel electrophoresis with a resolution of a single nucleotide. Repeating the process with a second DNA probe can reveal the location of the second crosslinked residue.

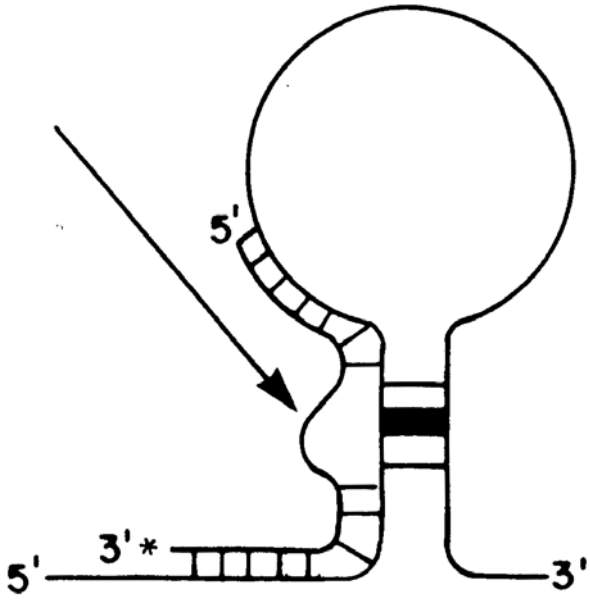

Figure 1. Scheme for determining the location of crosslinks in an RNA after hybridizing the RNA to a cloned DNA fragment. Arrow shows single stranded DNA bubble that can be cleaved specifically by mung bean nuclease.

In order for the method in figure 1 to function, one must have efficient methods of fractionating crosslinked RNA species. Otherwise, if the original crosslinked RNA sample contains a mixture of different products, one can learn the sites of crosslink

locations, but there will be no way to tell which crosslinked residues are paired with which. In the past, the method of choice has been to fractionate crosslinked RNA species by electrophoresis in the presence of extremely strong denaturants. Pure formamide has been used (Wollenzien and Cantor, 1982) as well as various formamide-urea mixtures (Quarless, S. and Cantor, C. R., unpublished results). These are effective but they are technically demanding and are easily compromised by impurities in the reagents. To try to obtain higher resolution separations of crosslinked RNAs, we have recently begun to fractionate structures such as that shown in figure 1, directly by electrophoresis under conditions that preserve the RNA-DNA hybrid. The presence of long stretches of double helix stiffens the crosslinked RNA loops. Preliminary results are extremely encouraging and suggest that almost all major 16S rRNA species known to be crosslinked in inactivated 30S ribosomes can be fractionated electrophoretically in very dense agarose gels.

Chromatin

One of the major questions of interest in chromatin is to determine the position of nucleosomes on individual genes and see whether this position alters as a function of gene activity. A technique that allows nucleosome location to be determined with relative ease is footprinting. The ideal footprinting reagent can react anywhere along a naked DNA duplex and either cut directly, or produce a lesion that can be converted, specifically, to a cut by subsequent chemical or enzymatic manipulation. Proteins bound to the DNA, such as the histones in nucleosomes, are expected to suppress the reactivity of the bound region by simple reduced steric accessibility. The resulting pattern of detection is a footprint of the location of the bound proteins.

Our objective has been to look at the nucleosome positions on a single copy eukaryotic gene. In order to do this, one must visualize the results of the footprint indirectly, as shown in figure 2. A reagent is first allowed to cut the DNA of intact chromatin. The reaction is carried out under extremely mild conditions so that most chains are not cut at all, and the remainder have just a single cut somewhere in the region of interest. Then proteins are removed, and a total cleavage is made near the region of interest by a properly chosen restriction nuclease. The resulting mixture of heterogeneous DNA chains is separated by length, electrophoretically. It is extraordinarily complex since it represents various partial fragments of the entire genome. A single radiolabelled DNA probe is used to examine the mixture of fragments. This is chosen to lie adjacent to the total cleavage site near the gene of interest as shown in figure 2. The result is that only those fragments adjacent to this site can hybridize to the probe. Thus, in a radioautogram one observes just the lengths of DNA corresponding to the cutting sites in the original chromatin structure. The detection technique is called indirect end labelling.

The major problem with most past studies of chromatin footprinting has been the sequence specificity of the reagents used for the initial cutting. These reagents must be used under very mild conditions. Thus, any kinetic preferences for particular DNA sequence will be highlighted. Even a naked DNA molecule under these conditions will show a complex pattern of reactivity (Dingwall *et al.,* 1981; Horz and Alterberger, 1981). The pattern of a nucleoprotein will be a superposition of the naked DNA pattern

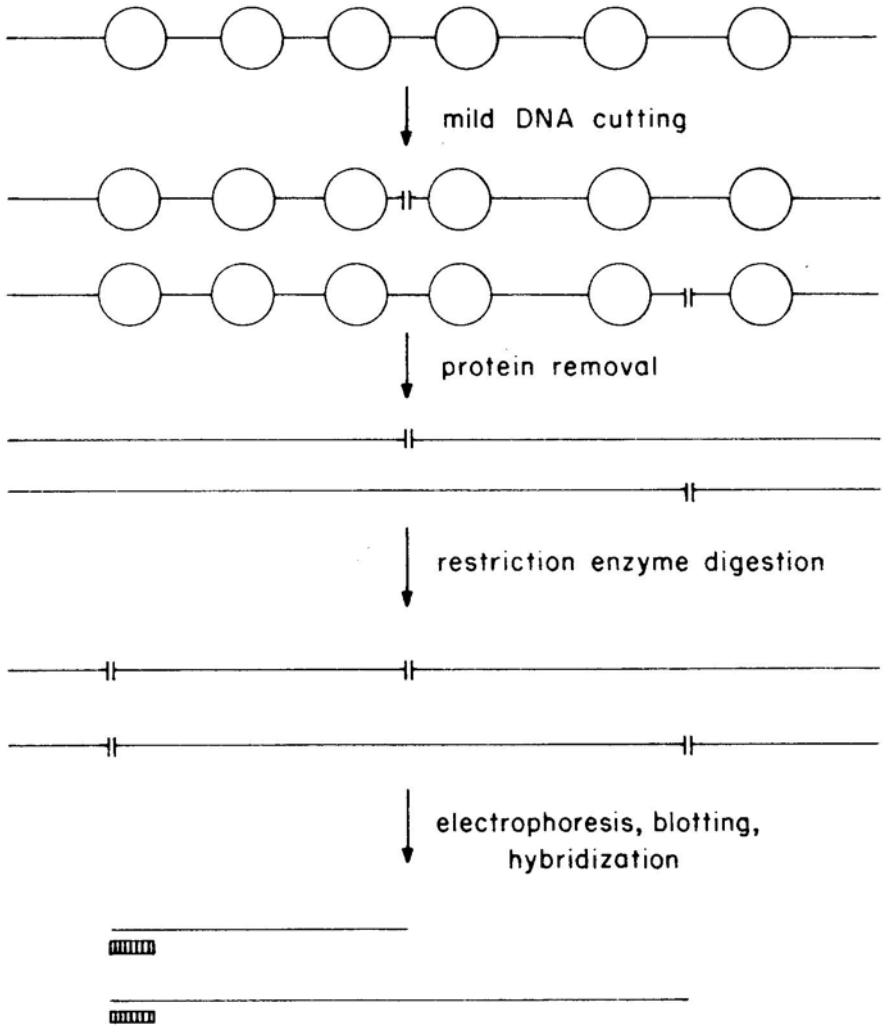

Figure 2. Procedure for determining the location of nucleosomes by footprinting and then using indirect end labelling to visualize a particular set of DNA fragments.

and the protection due to bound proteins. It is often impossible to deconvolute the two patterns. Fortunately one reagent has been found that shows little or any DNA cutting specificity even under the mildest of conditions. This reagent is the iron (II) complex of bismethidium-propyl-EDTA (Cartwright *et al.,* 1983). We have used this reagent to examine the pattern of nucleosomes on the mouse β-globin gene (Benezra, R., Cantor, C. R. and Axel, R., unpublished results). The results indicate that on the 3′ half of the gene, nucleosomes are present in the same precise positions in different cell types, including cells where the β-globin gene is expressed or can be induced, and cells where it cannot be expressed. However, the pattern of nucleosomes on the 5′ half of the gene is altered in complex ways by gene expression.

Viruses

We have been interested in developing methods that provide information on the configuration of the DNA chain in various forms of chromatin and in condensed chromosomes. As a model for these studies we have asked if it is possible to detect how DNA is folded inside the heads of viruses. In principle both DNA-DNA and DNA protein crosslinking agents can provide information on intraviral DNA packaging. However, most past methods of analysis had to rely on electronmicroscopic analysis of the crosslinked products, a very tedious process (Haas *et al.,* 1981, Schwartz *et al*., 1983; Widom and Baldwin, 1983). We have recently developed a much simpler method of analysis using gel electrophoresis (Welsh, J. and Cantor, C. R., unpublished results). This method is summarized, very schematically in figure 3.

Bis-psoralens have been synthesized that, in the presence of light, will react with two near-by double stranded DNA regions as might exist in any compact DNA-containing structure. A typical compound consists of two aminomethyltrioxsalens joined by a long water soluble chain containing a disulphide group. The reagent is diffused into the virus in the dark. Then irradiation with 360 nm light forms DNA-DNA crosslinks. The virus is disrupted and protein removed enzymatically. The free crosslinked DNA can then be cleaved into convenient fragments by restriction nucleases. Some of these fragments will be crosslinked together by the bis-psoralen. Since that crosslink is breakable, the identity of the crosslinked fragment pairs can be easily studied by diagonal gel electrophoresis. A fairly standard agarose gel procedure is used to separate the uncrosslinked restriction fragments. Under these conditions, the crosslinked pairs migrate roughly as the sum of the molecular weights of the two constituent pieces. Then the crosslinks are broken directly in the gel by irradiation with 254 nm light. This photo-detaches psoralens from the DNA and also appears to cause considerable cleavage of the disulphide bond, possibly, by psoralen photosensitized reactions. Then electrophoresis is run perpendicular to the original migration conditions. Now all pairs of fragments originally crosslinked run as free DNA molecules. Each fragment will resolve into a pair of bands that can be identified by their electrophoretic mobilities. This technique has been applied to the analysis of bacteriophage λ which consists of a naked DNA molecule, approximately 48 kb in length, coiled inside a protein capsid. The results suggest that the structure of the DNA inside the virus is either extremely complex, or more likely it is a mixture of many different structures or even a random structure. No evidence was found for any region of the DNA to crosslink preferentially to any other particular region. Similar experiments have been carried out on SV40 virus. This consists of a minichromosome containing 5243 bp of DNA organized into about 24 nucleosomes and contained within a protein capsid. Efficient bis-DNA crosslinking is obtained with SV40 showing that the bis-psoralens can reach between two adjacent DNA strands in packaged chromatin. The pattern of crosslinking shows some evidence for preferential contact between particular DNA regions but there is not yet enough data to allow specific structural interpretations.

Discussion

In this communication three approaches to study the structure of complex nucleo-protein assemblies have been discussed. In each case the methods are relatively simple

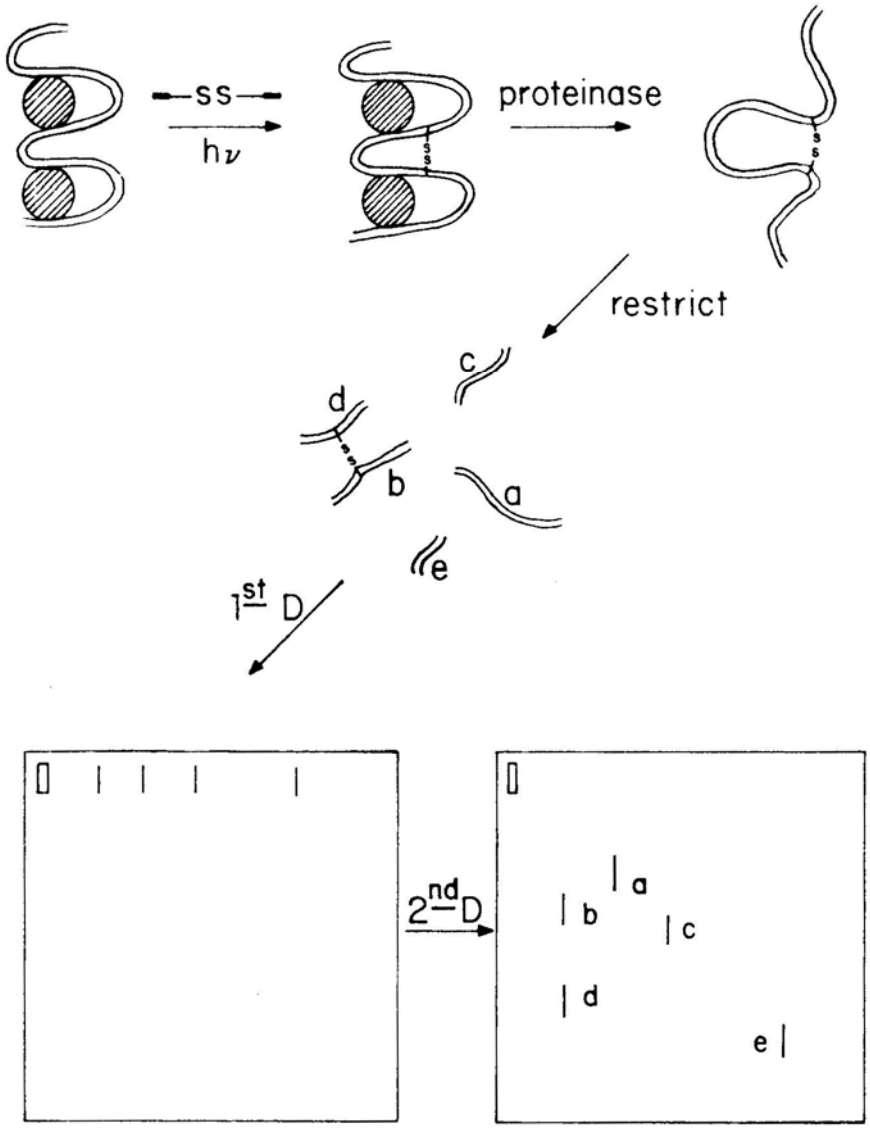

Figure 3. Analysis of the arrangement of DNA packaging in a complex nucleoprotein like a virus or chromosome, by using a breakable bis psoralen. After fragmentation of the DNA by restriction nuclease cleavage the resulting pattern of crosslinked fragments is analyzed by diagonal agarose gel electrophoresis.

and general because they rely on the enormous power of current technology for handling DNA molecules. Restriction nucleases allow one to cleave DNAs, specifically, more or less at will. Recombinant DNA methods allow the isolation of DNA reagents specific for any desired sequence. Electrophoretic methods can fractionate DNA molecules up to 500 bp with single nucleotide resolution and up to 20 kb with high enough resolution

so that one can always proceed with an analysis in stages, first low resolution detection of large fragments, then high resolution detection of small fragments.

We have shown how fractionation and enzymatic or chemical modification can be combined to analyze the structures of RNA in ribosomes, and DNA in chromatin and viruses. In the near future it should be possible to extend these techniques to examine the structures of whole prokaryotic and small eukaryotic chromosomes. The new technique of pulsed field gradient gel electrophoresis allows the fractionation of DNA molecules as large 4000 kb. This size range includes the DNA molecules from intact chromosomes of many simple organisms including yeast (Schwartz and Cantor, 1984), *E. coli.,* and many protozoan parasites such as trypanosomes (Van der Ploeg *et al.,* 1984a), leishmania (Van der Ploeg *et al.,* 1984b), and Plasmodium. By combining this separation method with footprinting and crosslinking, one should be able to begin to map the higher order structure of the chromosomes of these species.

Acknowledgement

This work was supported by grants from the United States Public Health Service and the National Science Foundation. We are grateful to the many collaborators who participated in aspects of the work described including Richard Axel, Piet Borst, David Schwartz, Lex Van der Ploeg, and Paul Wollenzien.

References

Cartwright, I. C., Hertzberg, R. P., Dervan, P. B. and Elgin, S. C. R. (1983) *Proc. Natl. Acad. Sci. USA,* **80,** 3213.

Dingwall, C., Lomonossoff, G. P. and Laskey, R. A. (1981) *Nucleic Acids Res*., **9,** 2659.

Haas, R., Murphy, R. F. and Cantor, C. R. (1981) *J. Mol. Biol*., **159,** 71.

Horz, W., and Altenberger, W. (1981) *Nucleic Acids Res*., **9,** 2643.

Hui, C. F. and Cantor, C. R. (1984) *Proc. Natl. Acad. Sci. USA*, (in press).

Schwartz, D., Saffran, W., Welsh, J., Haas, R. Goldenberg, M. and Cantor, C. R. (1983) *Cold Spring Harbor Symp. Quant. Biol*., **47,** 189.

Schwartz, D. C. and Cantor, C. R. (1984) *Cell*, **37,** 67.

Van der Ploeg, L. T. H., Schwartz, D., Cantor, C. R. and Borst, P. (1984a) *Cell*, **37,** 77.

Van der Ploeg, L. T. H., Cornelissen, A., Barry, J. D. and Borst, P. (1984b) *EMBO J*., (in press).

Widom, J. and Baldwin, R. L. (1983) *J. Mol. Biol*., **171,** 419.

Wollenzien, P. L., Murphy, R. F., Cantor, C. R., Expert-Bezancon, A., and Hayes, D. (1984) *J. Mol. Biol.* (in press).

Wollenzien, P. L. and Cantor, C. R. (1982) *J. Mol. Biol*., **159,** 151.