DNA ELECTRON MICROSCOPY

Author: **Christine Brack**
Basel Institute for Immunology
Basel, Switzerland

Referee: Jack D. Griffith
Cancer Research Center
University of North Carolina
Chapel Hill, North Carolina

I. INTRODUCTION

In recent years, DNA electron microscopy has become a tool of increasing interest in the fields of molecular genetics and molecular and cell biology. The rapid advance of in vitro recombination and DNA cloning technology has asked for new methods to characterize cloned DNA molecules. A variety of techniques are now available to obtain information about the structural and functional organization of genetic material. The size and structure of a nucleic acid molecule (molecular weight determinations, discrimination between double- and single-stranded regions, circularity, replicative structures, secondary structure in single strands, etc.) can be directly determined by viewing the isolated molecules. A general idea of the base composition of a DNA molecule can be obtained by partial denaturation mapping. Various hybridization methods allow detection and analysis of sequence complementarity between two DNA molecules (heteroduplex, D-loop mapping) or between DNA and RNA molecules (R-loop mapping, R-hybrids). Visualization of transcription complexes and transcription R-loop mapping permits mapping of promoter sites and analysis of in vivo and in vitro transcription. Special methods have been developed for the study of protein DNA interaction (direct visualization of protein-DNA complexes or indirect tagging methods).

The reproducibility and the resolution of these methods make it possible to map specific DNA regions with a precision of about 50 to 100 nucleotides. Therefore, electron microscopic mapping techniques have become a prerequisite for determining DNA regions of interest for subsequent nucleotide sequencing.

The present article is not intended to give an exhaustive literature review on the subject. Recent review papers should be consulted for additional complementary information on quantitative aspects,[1] technical questions,[2] or on the preparation of RNA molecules.[3] The main purpose of this review is to illustrate the contribution of DNA electron microscopy in the recent developments in molecular biology. Therefore, many examples have been chosen from recent work on cloned eukaryotic DNA.

II. METHODS FOR PREPARING NUCLEIC ACIDS FOR ELECTRON MICROSCOPY

A. Surface Spreading Techniques

In the late 1950s, Kleinschmidt and Zahn were studying mixed surface films of DNA and proteins. Examination of these films with the electron microscope revealed the DNA as a strikingly clear, extended thread outlined smoothly in the two-dimensional protein film.[4] The value of this method which transforms the random coils of DNA chains in solution into nicely arranged filaments measurable in length was immediately

evident. Since 1959 this technique has been widely used to study nucleic acids, and many modifications have been developed to improve contrast, resolution, spreading of single-stranded nucleic acids, and so on. The principles of the method have, however, remained unchanged.

In the BAC spreading method[5] and the Anthrabis method,[6,7] the cytochrome is replaced by the detergent benzyldimethylalkyl-ammonium chloride or Anthrabis, respectively, which form very thin surface films. The diameter of the nucleic acids is not enlarged as much as it is by cytochrome spreading; therefore, these methods are particularly useful for visualizing protein molecules bound to nucleic acids. A combination of cytochrome and detergent spreading has been described by Zollinger et al.[8] The procedure employs a mixture of cytochrome and sarkosyl and gives a resolution sufficient for the visualization of RNA polymerase bound to DNA.

In this section we shall describe briefly the different spreading methods. Some points relevant to particular types of molecules (heteroduplex, R-loops, protein-DNA complexes) will be further discussed in later sections. For further information, we would like to refer the reader to recent review papers describing in great detail the methodology, the quality, and the sources of material and technical equipment.[1,2,9,10]

1. Spreading in a Protein Monolayer

The basic protein film technique[4,11] is based on the following principle. Nucleic acids are mixed with a slightly basic protein. Several proteins have been used, but cytochrome c proved to give the best results. The DNA-cytochrome c solution is let to flow down a glass ramp onto a hypophase solution. During the spreading out at the air-water interphase, the cytochrome c denatures and forms a monolayer of denatured protein. Because denatured proteins bind nonspecifically and very cooperatively to nucleic acids, the cytochrome wraps around the DNA molecules and, spreading out between them, disentangles the filaments. As a result, individual molecules are embedded and spread out in a monolayer of denatured cytochrome c. A variety of conditions including ionic strength, specific cations or anions, pH, detergents, presence of other proteins, may lead to aggregation of nucleic acids, poor extension, and tangling.[1,12,13]

In practice, the spreading experiment is carried out in the following way: the spreading solution (hyperphase) containing the buffered DNA solution at 0.1 to 1 $\mu g/ml$ and cytochrome c (0.1 mg/ml) is allowed to flow down a clean glass ramp and spread at the surface of a hypophase solution. The hypophase can consist of either distilled water, an ammonium acetate solution, or a buffered formamide solution.[9,11,14] Many variables in the composition of the hypophase and spreading solution can influence the spreading quality, the amount of cytochrome collapsed around the nucleic acid molecules, the granularity of the background, the length of the molecules, etc. Some of these are (1) salt concentration, (2) type of salt or buffer used, (3) formamide concentration.[1,9,13] The protein films formed at the water-air interphase is picked up on a support film (either carbon or collodion film mounted on an electron microscope copper grid) by touching the surface with the grid. The samples are stained either with uranyl acetate ($5 \times 10^{-5} M$ in 90% ethanol)[9,15] or phosphotungstic acid,[16] dehydrated by rinsing in 90% ethanol and isopentane (methyl butane) and dried in air. For many purposes, staining alone gives enough contrast for visualization of the molecules.[1,9,16,17] Additional contrast can be obtained by rotary shadowing the grids with heavy metal. Visualization of nucleic acids with the dark field methods enhances the contrast considerably; excellent results are obtained by staining or even without any additional contrasting.[17-21]

An important improvement to the spreading method was the introduction of the use of formamide to remove the random base interactions in single-stranded DNA.[22] While

with the aqueous method single-stranded DNA appears as collapsed bushes, with the formamide method both double- and single-stranded molecules are well extended. Differences in the thickness and the appearance — smooth or kinky — allow distinction between duplex and single-strand, respectively. In addition, formamide seems to raise the contrast of the molecules.[9] In another method described for the electron microscopy of single-stranded DNA, the molecules are unfolded by reaction with formaldehyde.[13] The formamide method, however, is nowadays the most widely used technique for spreading nucleic acids — in particular, for mounting hybrid molecules.

The best description of the method was written by Davis, Simon, and Davidson,[9] and important points are discussed in the review of Ferguson and Davis.[1] Usually, the spreading solution contains 40 to 50% formamide and should be well buffered (0.1 *M* Tris HCl, 0.01 *M* Na_3EDTA, pH 8.5) because formamide solutions tend to become acidic very rapidly. Up to 70% formamide solution can be used for the spreading of molecules or for demonstrating weak homologies in hybrid molecules.[23-25] The hypophase is 30% less in formamide and has one tenth of the buffer concentration compared to the spreading solution and is made up immediately before use. Since formamide lowers the melting temperature (Tm) of DNA,[26] exact conditions of temperature, formamide concentration, and ionic strength must be considered to obtain spreading conditions such that double-stranded DNA is stable, while random base interactions in single-stranded DNA are melted out. Davis and Hyman[24] introduced a procedure, called isodenaturing mounting, in which the denaturing conditions of the spreading solution are identical to those of the hypophase. The difference in formamide concentration between spreading solution and hypophase are calculated to compensate for the tenfold differences in ionic strength. Thus, the isodenaturing conditions assure that the molecules are mounted for electron microscopy under true equilibrium conditions and not trapped in some metastable state. This is particularly important when the degree of sequence homology between two related DNAs in a heteroduplex molecule should be calculated.[1,24]

The protein monolayer spreading, in particular the formamide technique, can be easily applied to any type of nucleic acid molecule: single- and double-stranded DNAs, RNA, hybrid molecules, and, in certain cases, also DNA-protein complexes (see later section).

2. Diffusion and Spontaneous Adsorption Methods

A variation of the basic protein technique is the diffusion method.[11,27] A monolayer of cytochrome c is spread on the hypophase solution containing the DNA. By diffusion the molecules come into contact with the protein film and adsorb irreversibly to it. Because the molecules are essentially spread by the diffusion movement, they are less subjected to shearing forces that occur during spreading out of the mixed protein film. The diffusion method is therefore recommended for the spreading of very long DNA molecules. Also, because the number of the molecules adsorbed to the film increases with time, DNA samples with very low concentrations can be processed. It should be taken into account that the diffusion time depends both on the concentration and the size of the molecules (see below).

A valuable modification of the diffusion method was described by Lang and Mitani.[28] It is the microversion of their spontaneous adsorption method, also referred to as the "droplet method", which was developed to minimize the amount of material used in the diffusion method. A critical amount of cytochrome c (1.3 $\mu g/ml$) is simply added to the DNA solution. Small droplets (20 to 40 μl) of the solution are placed on a hydrophobic surface (Teflon® or Parafilm®) and covered for protection. Monolayer formation and adsorption of the macromolecules occurs spontaneously. The

number of molecules adsorbed to the film is a function of concentration and time. A critical time of 4 min is needed for the formation of the cytochrome surface film. After that time, and for at least 5 hr, DNA molecules continue to adsorb in proportion to [(time)$^{0.75}$]. In the longer adsorption times, more cytochrome c diffuses to the surface and denatures; this thick protein film decreases the contrast of the embedded DNA molecules. The important factor for obtaining optimal contrast is the relative combination of cytochrome c concentration and diffusion time. When higher cytochrome c concentrations are used (60 $\mu g/m\ell$), a film is formed after 1 min.[19,4] This method permits working with DNA concentrations as small as 2×10^{-5} μg per droplet. In the original method, Lang and Mitani suggested adding 0.07 *M* formaldehyde into the DNA-cytochrome droplet.[28] This was thought to influence the denaturation of the cytochrome at the air-water interphase and to stabilize the formation of a monolayer film.

Delain and Brack studied the influence of different environmental conditions on surface film formation and molecule extension in the droplet.[17] Instead of mixing formaldehyde into the spreading solution, they added a drop of formaldehyde or other solutions (gluteraldehyde, osmium tetroxide, ethanol, acetone, etc.) beside the droplet. The drops were covered so that during the surface film formation the sample was in a saturated atmosphere of the respective chemical. Besides formaldehyde, glutaraldehyde, osmium tetroxide and other fixatives resulted in a remarkable extension of the molecules when compared to a sample spread without any formaldehyde. On the other hand, alcohol, acetone, or ammonia vapor inhibited the film formation.

The simplicity of the method and the small amount of material needed have made the "droplet method" of particular use in many fields. It allows rapid and simultaneous processing of many samples (e.g., kinetic studies of enzyme reactions, checking of gradient fractions) with a minimum of manipulations, and, combined with dark field visualization, also in a minimum of time. When formamide is added, it is possible to spread even single-stranded DNA and hybrid molecules with the same ease.[17,29] So far, it is the method of choice for gentle spreading of high molecular weight DNA associations; e.g., the trypanosome kinetoplast DNA networks (Figure 1).[21,30,31]

Care should be taken with quantitative interpretation of DNA samples spread with the droplet technique and the diffusion method. In a mixed population of molecules with different molecular weights, small molecules diffuse faster and therefore appear in higher concentration on the grid than long molecules.[11] Also, the droplet method is more sensitive to variations in salt concentration and impurities in the DNA sample. In our own experience it is, for example, impossible to spread DNA-protein complexes. The molecules are condensed into rod- or doughnut-shaped particles, probably because the specific proteins bound (e.g., polymerases, restriction endonucleases, antibodies) serve as nucleation points for unspecific interaction with cytochrome molecules during the diffusion period.

3. Spreading in a Detergent Film

A considerable improvement in the resolution of spread nucleic acid molecules was obtained in the BAC method described by Vollenweider and co-workers.[5] Instead of using cytochrome c for surface spreading, the cationic detergent benzyl dimethyl-alkyl-ammoniumchloride is added. Both the spreading on a hypophase and the diffusion droplet method can be used with the same ease as the protein monolayer techniques. In the spreading version, the BAC stock solutions (2mg BAC in 1 $m\ell$ formamide) is diluted into the buffered DNA solution to a final concentration of 2.5×10^{-3}%. In the droplet version, a final BAC concentration of 2×10^{-4} is used. Formamide can be added for the spreading of RNA, single-stranded DNA, and hybrid molecules. The

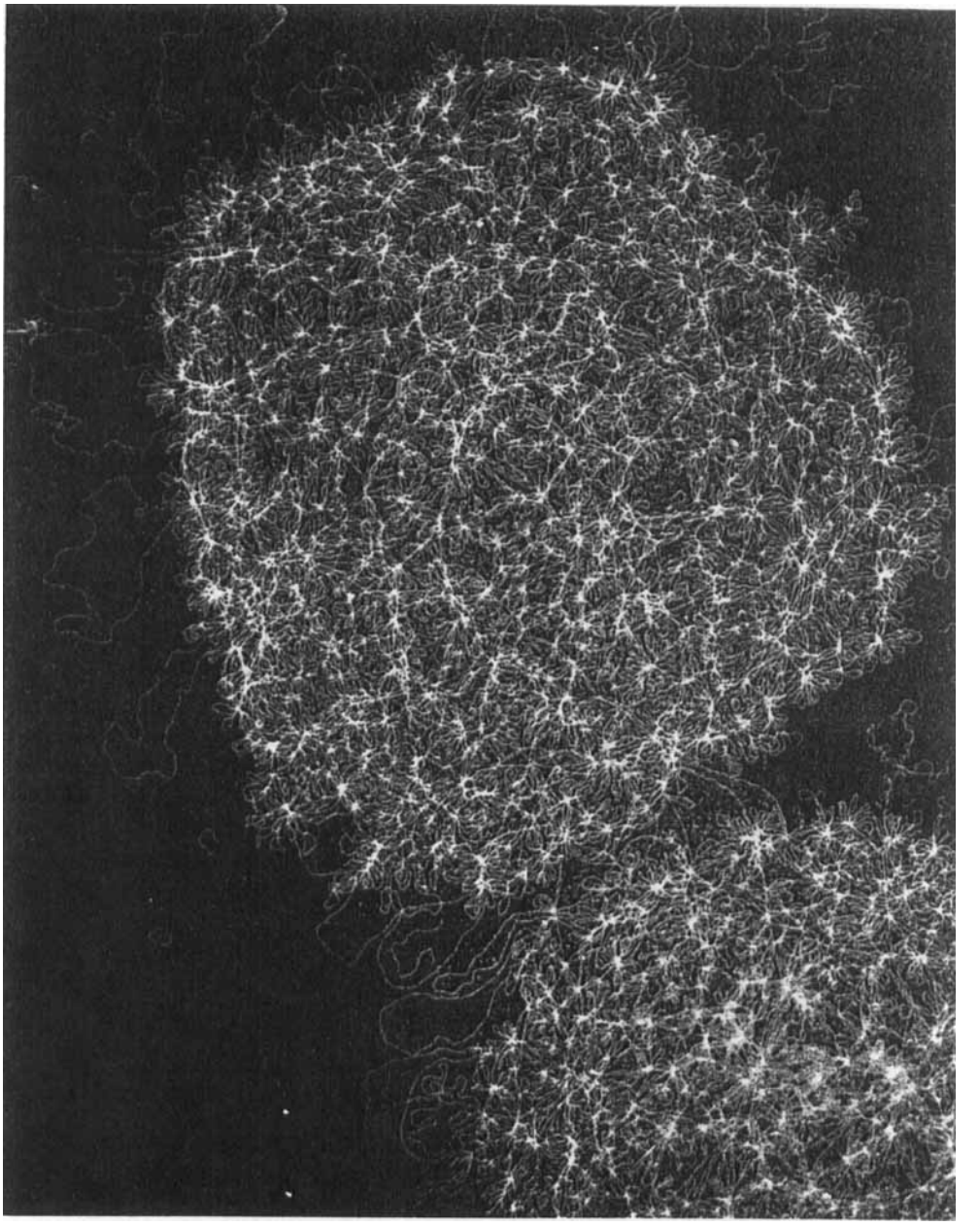

FIGURE 1. Network of *Trypanosoma brucei* kinetoplast DNA spread with the microdroplet method.[28] This large DNA association consists of thousands of minicircular DNA molecules linked together in a network, plus additional maxicircles at the edge of the network. The sample was picked up on a thin carbon film, stained with uranyl acetate (2% aqueous solution), and photographed in dark field.

mixture is spread either over a glass ramp as in the Kleinschmidt method, or by depositing a small amount (2 μl) of solution on the hypophase, which in most cases is distilled water.[32] Because contrast enhancement with staining alone is not sufficient, the molecules are usually visualized after shadowing with heavy metal.[5,32]

The advantage of the BAC method lies in the fact that because of its low molecular weight, a much thinner surface film is formed, one which hardly increases the diameter of nucleic acids. Whereas in cytochrome spreadings double-stranded DNA has a width

of approximately 10 to 15 nm due to the protein coat, in the BAC spreadings the width of a duplex DNA is about 5 to 6 nm. This increases the resolution of small details along the molecule, like denatured regions or secondary structure loops.[10,32-34] Furthermore, it allows visualization of individual protein molecules bound to nucleic acids.[5,35,36] This aspect will be discussed in a later section.

In spite of its advantages, which promised the BAC method to become a routine method equivalent to the basic protein film spreading, many laboratories still encounter difficulties in reproducing satisfactory results.[10,37,38] One source of difficulty is the water of the hypophase; it should be freshly distilled and rapidly cooled before use.[195] The ionic strength of spreading and hypophase solution influence both appearance and contour length of the molecules.[5,10] To obtain high-resolution preparations, samples should be picked up on carbon support films. The use of carbon films introduces additional problems such as variations in hydrophilicity and therefore in adsorption properties. These can be overcome by first floating the carbon films either on distilled water, on an ethidium bromide solution (100 $\mu g/m\ell$) followed by distilled water,[34] or on a 10^{-3}% solution of Alcian blue.[196] These treatments change the surface properties of the carbon film, probably rendering it positively charged.

Recently, another spreading method has been described that uses Anthrabis for surface film formation in the droplet method.[6,7,10] Anthrabis is 1,4-bis-[3-(benzyldimethyl-ammonio)-propylamino]9,10 anthraquinone dichloride and is mixed with the spreading solution (containing 0.1 $\mu g/m\ell$ DNA, any buffer pH 6 to 11, and 30% formamide) to a final concentration of 0.01%. The samples are prepared as in the droplet method,[28] picked up on carbon films, and rotary shadowed. Apparently there were no problems concerning the adsorption properties of the carbon films. The authors claim that both single- and double-stranded DNA, as well as the protein-nucleic acid complexes, can be spread with results superior to those obtained with the BAC method.[10] So far, the method has been used successfully in only a few laboratories, and the Anthrabis can be obtained only from the laboratory of J. O. Thomas, Department of Biochemistry, New York University Medical Center, New York (it is not commercially available).

Thach and Thach have described a spreading technique in which nucleic acid samples containing 0.01% of the detergent Brij 58 are drawn into a thin film by means of a wire loop.[39] The method has been used to study viral RNA, but has not found any wider applications.

B. Adsorption Methods

The adsorption methods were developed primarily for high-resolution visualization of nucleic acids. Initially, the effort of many laboratories was directed towards developing electron microscope methods for nucleotide sequencing. The detection of individual, selectively stained nucleotides would require a spreading technique which does not involve any carrier protein.[40,41] The availability of excellent chemical methods for DNA sequencing makes it unlikely, nowadays, that electron microscopy will contribute to this field. Therefore, the adsorption methods are mainly used to study protein-nucleic acid interaction.

The essence of these methods is to treat a supporting surface such that it will adsorb nucleic acids. When no protein film or other monolayer surrounds the molecules, it should be possible to see the "naked" DNA molecules and small proteins bound to them. Because no spreading forces act to disentangle long molecules, most of these methods are limited to low molecular weight nucleic acids. Furthermore, the demand for high resolution requires the use of thin carbon films as support films in many adsorption methods.

1. Adsorption to Charged Carbon Films

The method developed by Dubochet et al.[19] involves positively charging carbon films by submitting them to glow discharge in pentylamine vapor. The carbon coated grids are put in a small vacuum chamber which is evacuated to $\sim 10^{-2}$ Torr. A small amount of pentylamine vapor is introduced through a needle valve so that an equilibrium is reached at $\sim 2 \times 10^{-1}$ Torr, with continuous pumping. Glow discharge — that gives a characteristic bluish-purple color — is maintained for 15 to 20 sec. It is thought to deposit positive charges on the surface of the carbon film. The grids become highly hydrophobic and adsorb nucleic acids and negatively charged proteins with high efficiency, probably through electrostatic and hydrophobic interaction. Alkylamines of different chain length, as well as other amines, give similar results.[197] The films slowly lose their positive charge and should be used within a few hours after the glow discharge; the best results are obtained after 10 to 30 min.

Samples are prepared by placing a small droplet of DNA solution onto the charged grid. Adsorption of nucleic acids is proportional to time, and it is therefore possible to use DNA solutions of low concentrations (0.1 to 0.5 $\mu g/ml$) by increasing the time of adsorption. As standard conditions, we use 5 μl droplets of a DNA solution containing 1 $\mu g/ml$ DNA; after 1-min adsorption, a reasonable distribution of molecules is observed. The adsorption efficiency under these conditions is about 10%.[19] Any buffer solution, including formamide solutions, can be used, provided their ionic strength is not too high (<0.1 *M*). After adsorption, the grids are rinsed by floating on distilled water to wash off the buffer and salt solutions, stained briefly with aqueous uranyl acetate (2%)[19,42] or uranyl formate (0.5%) made up shortly before use,[43] and dried by blotting on filter paper. Molecules are visualized directly either by bright field or dark field,[19,43] or after shadowing with platinum or tungsten.[45-48] In stained preparations, the width of double-stranded DNA is about 1.5 to 2.5 nm (Figure 2), which indicates that the uranyl salts mainly interact with the phosphate backbone of the DNA duplex. After shadowing with platinum, the width of duplex DNA is 4 to 5 nm (Figures 3, 23). The very smooth and clean background on thin carbon films gives a high resolution, and DNA molecules as short as 30 base pairs (bp) can be measured,[43] which would be completely lost in a cytochrome monolayer. Length measurements on double-stranded DNA adsorbed to charged films give a higher reproducibility (± 1% SD) than cytochrome spread molecules (± 1 to 4% SD). This is understandable because little spreading forces act on the molecule. As a result, the absolute length of duplex DNA may be up to 5 to 10% shorter than in monolayer spreadings. (These values vary, depending on the ionic strength and the buffer used.)

The main advantage of this method lies in its application to high resolution studies of nucleic acid-protein complexes, and will be discussed in Section X. Some limitations have to be mentioned: molecules larger than $\sim 30 \times 10^6$ mol wt often appear too tangled to be measured. Therefore, the adsorption method can be routinely used only for low molecular weight nucleic acids. Spreading of single-stranded DNA or RNA requires relatively high formamide concentrations (70 to 80%), which reduce the adsorption capacity of the charged films. Even then, single-stranded molecules are not well extended and difficult to measure (e.g., Figure 4). Attempts to visualize very short R-loops or hybrid molecules have failed so far.

Griffith described a method for direct adsorption of nucleic acids to carbon films activated by glow discharge in an air atmosphere of reduced pressure (2×10^{-1} Torr).[49] Samples were prepared essentially in the same way as described for the pentylamine charged grids and were shadowed with tungsten. The method was successfully applied to various investigations on protein-DNA interaction and chromatin structure.[49-51] In a recent review, Griffith and Christiansen discuss certain technical points of different

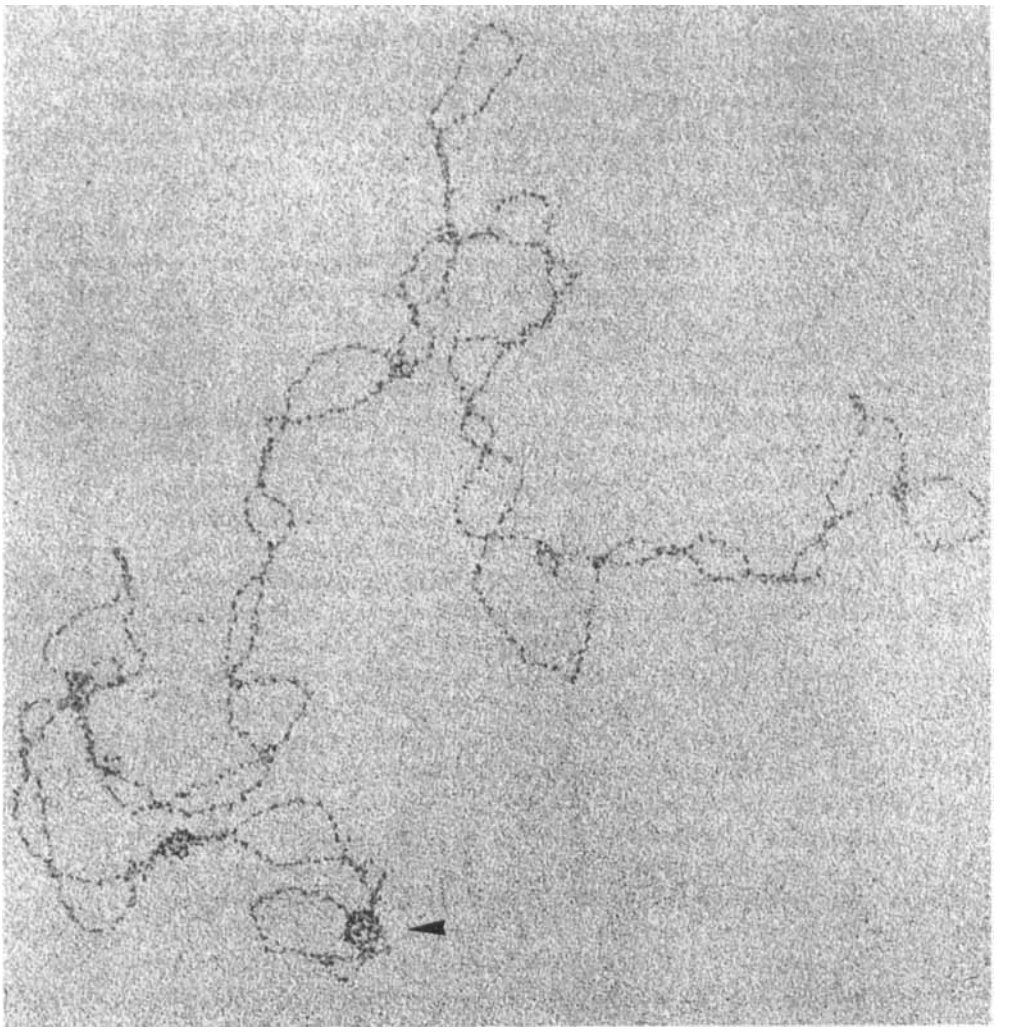

FIGURE 2. Supercoiled PM2 DNA molecule adsorbed onto pentylamine glow-discharged carbon film, stained with aqueous 0.5% uranyl formate solution, photographed in bright field. One molecule of the restriction endonuclease *EcoK* bound to the DNA appears negatively stained (→).

adsorption methods, in particular the importance of fixation for preparing DNA-protein complexes.[37] Glow-discharged carbon films are used also in the Miller technique[52] for spreading of transcriptionally active genes. This method will be discussed in Section IX.

2. Adsorption to Pretreated Carbon Films

Other approaches have been proposed to render the surface of carbon films suitable for direct adsorption of nucleic acids. A film of cytochrome c has been applied to the carbon and after drying, these precoated carbon films were used for subsequent adsorption of DNA.[53,54] Williams has obtained good results by treating glow discharged carbon films with a polylysine solution.[55] These films adsorb single- and double-stranded nucleic acids, protein-DNA complexes, and proteins from various solutions, provided they contain a minimum of 3 m*M* Mg^{++}.

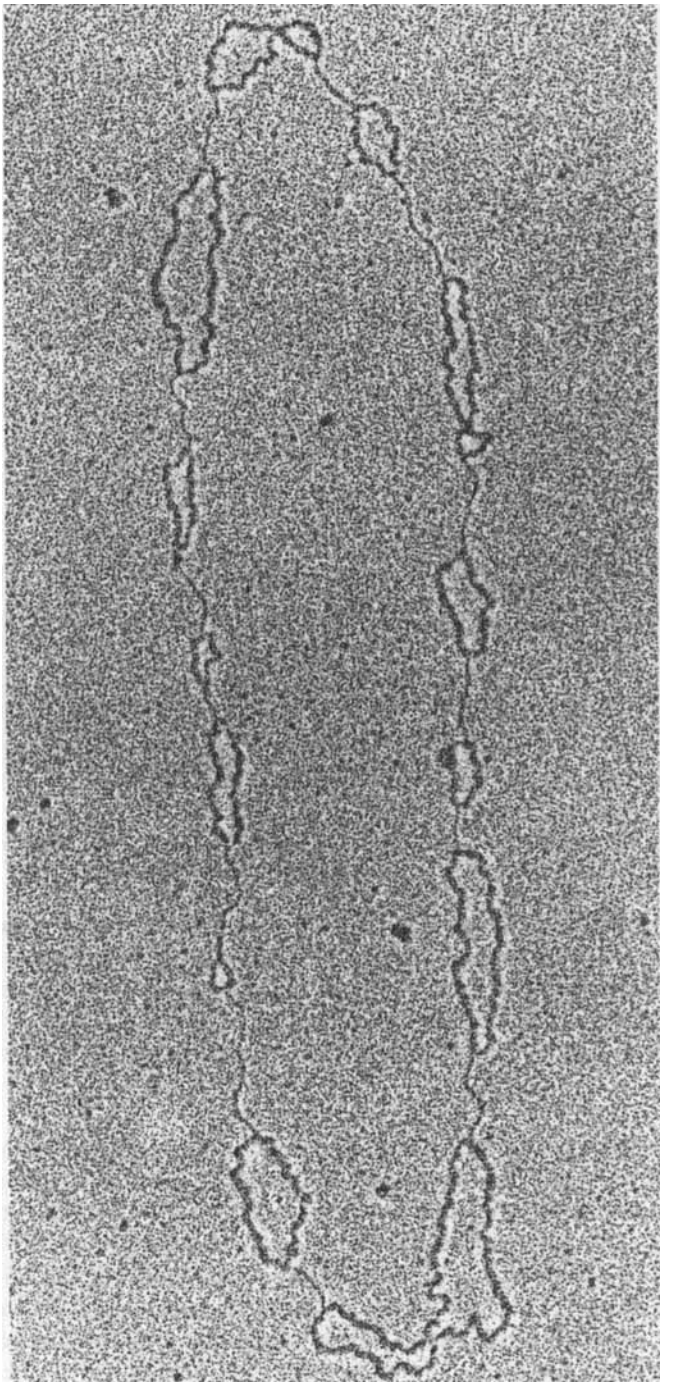

FIGURE 3. PM2 DNA molecule partially denatured with the T4 gene-32 protein. The complex was adsorbed onto pentylamine glow-discharged carbon film, stained with uranyl acetate, and rotary shadowed with platinum. The protein-coated single-stranded regions appear thicker than the double-stranded DNA.[45]

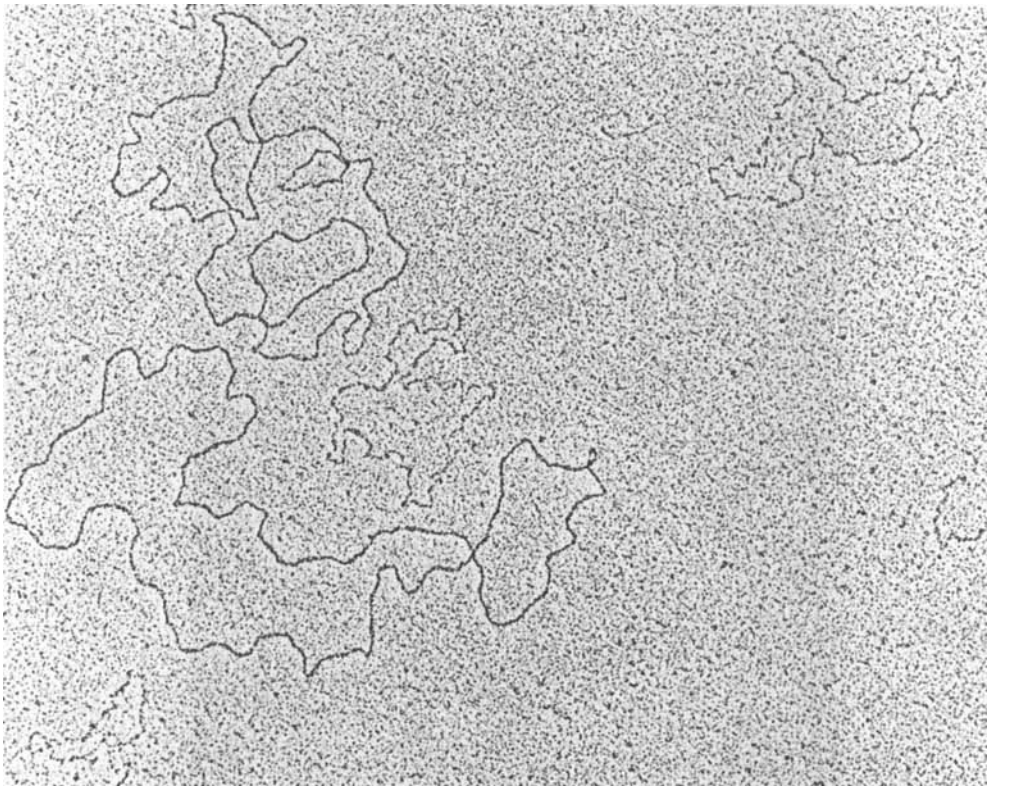

FIGURE 4. Double-stranded PM2 DNA and single-stranded fd DNA adsorbed onto pentylamine glow-discharged carbon films in the presence of 80% formamide; rotary shadowed with platinum.

To render the surface of carbon films positively charged, Koller et al. have pretreated them with BAC solutions.[40] The best results were obtained when the carbon films were floated overnight on a 10^{-5}% BAC solution, dried, and then used within 1 hr. It is also possible to pretreat carbon films with alkyl amine salts, which confers surface properties similar to those obtained after glow discharge in alkyl amines.[197]

3. Adsorption to Mica

Another supporting material that has been used for adsorbing nucleic acid molecules under various conditions is freshly cleaved mica sheets. After adsorption, the molecules are visualized in the electron microscope as metal shadowed carbon replicas. Mica sheets treated with aluminium salts were used to adsorb double-stranded DNA.[56]

Koller et al. have observed that DNA can be adsorbed directly on mica (and also on carbon grids) from the surface of DNA solutions containing ethidium bromide, actinomycine, or propidium diiodide.[57] These intercalating dyes are used to increase the rigidity of the DNA molecules and to mediate their adsorption to mica (or carbon surface). In this method, ethidium bromide dissolved in redistilled water is added to the DNA solution immediately before use, to give a final concentration of 50 to 250 μg/ml. Droplets of 0.1 to 0.15 ml of this solution are placed on a clean Teflon® or Parafilm® surface and kept for 10 to 15 min at room temperature. A 2 × 3-cm piece of freshly cleaved mica is then touched to the surface of the DNA droplet. It is very important to remove the excess DNA solution and dye by washing the mica sheets in redistilled water (several hours or overnight). After dehydration in ethanol, the mica sheets are dried and shadowed with platinum. A carbon layer is then evaporated onto

sheets are dried and shadowed with platinum. A carbon layer is then evaporated onto the mica in order to produce a replica which can be floated off onto distilled water. At DNA concentrations of 0.1 to 0.3 μg/m*l*, the molecules appear well separated and smoothly spread out. Although ethidium bromide mediated adsorption is possible on carbon film, better results were obtained with freshly cleaved mica.[57].

When the solution contains 10 m*M* Mg^{++}, in addition to 100 m*M* ethidium bromide, good adsorption of DNA molecules to mica is obtained even from samples containing high salt concentrations (up to 0.5 *M* NaCl).[196] This spreading method can then be routinely used to check fractions directly from a CsCl gradient or from enzyme reaction mixtures. The ethidium bromide adsorption method has been widely used for mounting protein-DNA complexes; e.g., for the visualization of RNA polymerase binding sites,[57] for gene-32 protein coated hybrid molecules,[58] and for chromatin.[59,60]

Adsorption to mica can be also mediated by divalent cations in the absence of ethidium bromide. Portmann and Koller[61] described the use of magnesium, manganese, calcium, and cobalt salts for the spreading of DNA molecules. DNA-polymerase complexes were adsorbed from a droplet containing between 8 and 10 m*M* of these divalent cations as described above. Partially denatured DNA was spread from a 6 m*M* acetate solution containing 30% formamide; single- and double-stranded regions were well distinguished.[61]

One many easily get confused when confronted with this variety of preparation methods. The advantages of certain methods and applications to specific problems will be illustrated in the following sections. It should be kept in mind that the best approach is usually to select one particular method and adapt it to one's own problems. Most of the current methods are flexible enough to be applied to a variety of conditions. When high resolution is not required, the cytochrome monolayer spreading gives the most reproducible results. For visualization of smaller details and proteins bound to nucleic acids, either the BAC spreading method or adsorption to activated carbon films can be used.[19,37,49] Some authors state that the BAC method has several advantages over the adsorption to charged carbon films: double- and single-stranded DNA and RNA can be spread equally well; no additional technical equipment is required. In our own hands, the adsorption to pentylamine glow-discharged carbon films has given best results. Once the system had been set up and experimentally optimized, reproducible results could be obtained by many different users.

As a recommendation for beginners in the electron microscope techniques, it should be emphasized that the "Kleinschmidt spreading," i.e., the cytochrome c monolayer technique, is still the easiest and least problematic method. The direct mounting techniques, however, may need more experience both in the technical part and in the interpretation.

C. Reagents

1. Cytochrome c

Every type of cytochrome c and every batch gives different results in the spreading technique. Criteria for the quality of a cytochrome c are smoothness of background and aggregate-free spreading of double- and particularly single-stranded DNA. In spreadings with high formamide and high salt concentrations (e.g., R-loop buffers), the DNA can be very thickly covered with cytochrome, and lateral aggregation can occur (Figure 11).[1] Some purification can be achieved by passing the cytochrome solution through a Millipore® filter (0.45 μ) or by chromatography with Sephadex® G100. The concentration of stock solutions is measured spectrophotometrically — the A_{410} of a 1 mg/m*l* solution is about 10^1. This solution is diluted to 1 mg/m*l* with distilled water and can be stored at 4°C for several months.

Considerable improvement in the smoothness of the background, particularly in formamide spreadings, can be achieved by cleaving cytochrome c with CNBr.[62,63] About 20 mg/ml cytochrome c in 1 ml 0.1 *M* HCl is incubated overnight with 30 mg CNBr at room temperature (poisonous, use a hood!) The reaction mixture is passed through a Sephadex® G25 column and peak fractions are checked for their quality of spreading by electron microscopy. Usually lower concentrations of the cleaved cytochrome c can be used than of intact cytochrome c. Some batches may spread double-stranded DNA better than single-stranded DNA.[63] Figure 5 shows two transcription R-loop molecules spread with uncleaved (b) and cleaved (a) cytochrome c respectively.

The smoothness of the cytochrome c film can vary also considerably — from rough to very smooth — with the amount of cytochrome used and the time the film is allowed to remain on the hypophase before being picked up on grids.[194]

2. Formamide

The quality of cytochrome c spreadings with the formamide version varies to a great extent with different batches of formamide. Irregular extension or aggregation of molecules can be observed with certain brands of commercial formamide. In that case, it may be recommended to purify the chemical. Several procedures have been proposed for purification of formamide. Nonvolatile material can be removed by distillation in a rotary evaporator under high vacuum.[63] A 10% solution of formamide should have a conductivity not higher than five- to ten-fold the conductivity of deionized water. Recrystallization[64] can lower the conductivity up to ten-fold in one step. If necessary, the crystallization can be repeated two to three times. Crystals are collected by centrifugation and stored at −70°C. Formamide can also be deionized with ion exchange resins.[65] In the author's laboratory, recrystalized formamide is used for hybridization reactions (R-loop or heteroduplex formation) and for the hyperphase solutions. The hypophase is usually prepared with commercial *pro analysi* grade formamide.

D. Support Films

Cytochrome monolayer spreadings are usually mounted on collodion (parlodion) or formvar support films. They are prepared by spreading a drop of collodion solution (2 to 3% in pentylacetate) on distilled water.[1,2,66] Freshly prepared films (not over 1 to 2 days old) result in the best adsorption properties and contrast.

For adsorption methods carbon films are used. They are more difficult to prepare, but they are more stable, have a lower electron density and a finer background, and contaminate less in the electron beam than parlodion films. Carbon films are prepared by evaporating carbon onto freshly cleaved mica and floating off the film on distilled water.[66] The surface properties of carbon films are variable and can be modified by glow discharge or various other treatments (see previous section). Very thin carbon films mounted on holey films are used for dark field visualization of nucleic acids.[19]

E. Staining and Shadowing

The contrast of nucleic acid molecules spread with or without a protein monolayer is usually enhanced by staining with heavy metal salts and/or low angle rotary shadowing with heavy metal. Most frequently, alcoholic uranyl acetate staining is used for improving the contrast of cytochrome c spread molecules.[9,15] The stock solution of 0.05 *M* uranyl acetate in 0.05 *M* HCl can be stored in the dark for months. It is diluted 1000-fold into 90% ethanol not more than 1 hr prior to use. The contrast is determined by the amount of cytochrome collapsed around the DNA, which depends on many variables.[1] Ionic strength, pH, formamide concentration of the hypophase, and spreading solution are important factors. Aqueous solutions of uranyl acetate have

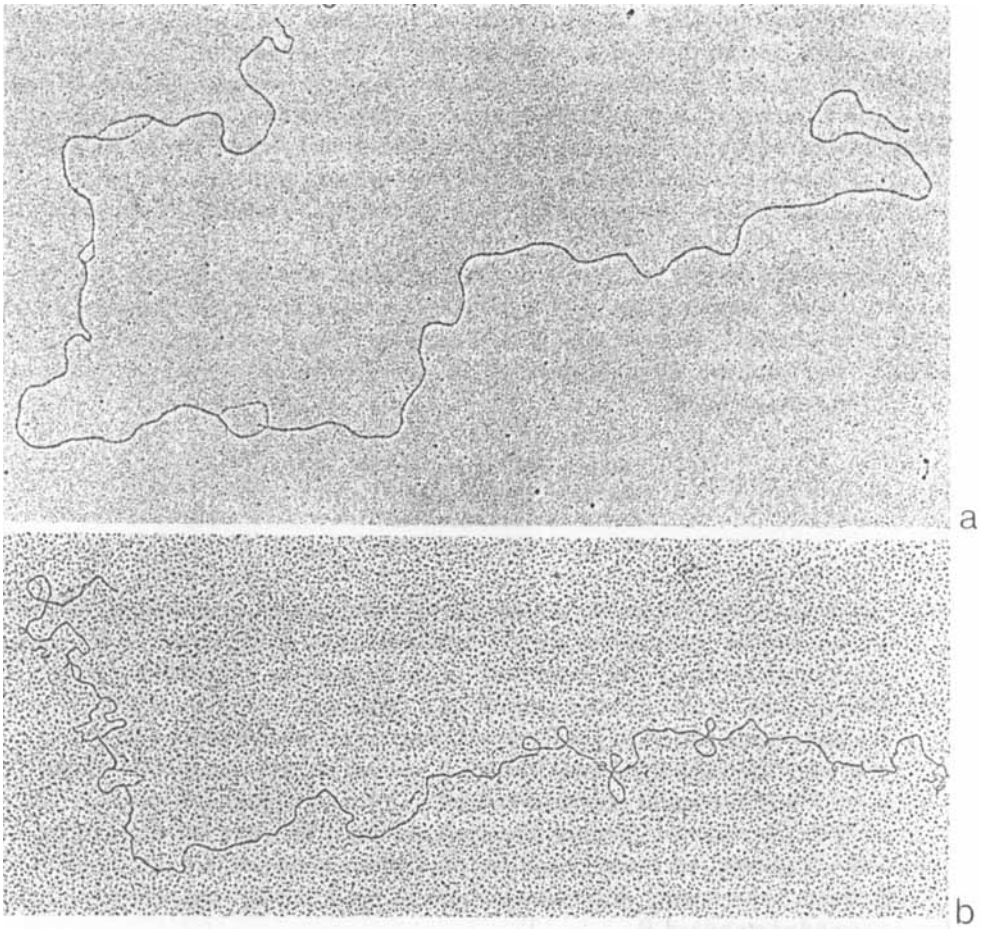

FIGURE 5. Transcription R-loop molecules. λgtWES DNA was transcribed with *E. coli* RNA polymerase, and R-loops were formed as described. The samples were spread with CNBr cleaved cytochrome c (a), or with uncleaved cytochrome c (b). (From Brack, C., *Proc. Natl. Acad. Sci. U.S.A.*, 76, 3164, 1979.)

been used for staining DNA adsorbed to charged carbon films for dark field visualization,[19,42-44] or for cytochrome spread DNA in dark field.[17,18,20,21] A 2% phosphotungstic acid solution in 50% ethanol gives sufficient contrast for direct observation of cytochrome c spread molecules in bright field.[16,17] Aqueous solution (0.5%) of uranyl formate was used for staining DNA and DNA-protein complexes.[43,44,67] Uranyl formate solutions are not stable and should be prepared and Millipore®-filtered immediately before use.

Additional contrast is obtained by low-angle shadowing with a heavy metal, usually platinum or a metal alloy Pt:Ir, Pt:Pd. Detailed description of the procedure was recently reviewed.[2] Griffith has observed the phenomenon of selective nucleation by evaporating tungsten on samples adsorbed on activated carbon films.[49] When small amounts of tungsten are evaporated onto DNA-protein complexes, selective binding or nucleation occurs primarily on protein, somewhat less on DNA, and much less on the carbon support.

III. LENGTH MEASUREMENTS

In all spreading and adsorption methods described above, variations in the absolute

length of DNA molecules are to be expected. In particular, the physical stretching of the nucleic acids during the surface denaturation of the cytochrome film cannot be reproducibly controlled. In addition, many factors like ionic strength, formamide concentration in hypophase, and spreading solution influence the length of nucleic acid molecules[12] Lengths should, therefore, always be determined relative to standard DNA molecules of known molecular weight mounted on the same grid and never as absolute lengths. Exact calibration of microscope and magnification systems thus become unnecessary.

DNA sequencing of the last few years has provided us with absolute molecular weights of several small circular DNA molecules, which are preferentially used as standard DNAs because they are readily distinguished from a population of unknown molecules. The following list gives some of the DNA molecules whose nucleotide sequence has been determined:

SV40	5226 base pairs[68,69]
pBR322	4361 base pairs[70]
øX174	5386 nucleotides[71]
fd	6408 nucleotides[72]

Although these circular molecules are most frequently used as absolute length standards, they may not be optimal for length determinations of molecules with a wide range of molecular weights. The spreading forces exerted on a closed circle may be considerably different from the ones exerted on a short linear or very long linear molecules.* We have often observed that standard deviations measured on circular molecules were smaller than the ones measured on co-spread linear samples. This is particularly true for DNA molecules prepared by adsorption to charged carbon films. When many experiments involving different techniques are made with one DNA sample, e.g., a cloned DNA fragment, we usually calibrate the fragment relative to a standard DNA. In all subsequent experiments, this cloned fragment itself can be used either as single- or double-stranded standard.[25]

All length measurements are expressed as length ratios between the standard DNA and the sample DNA. They are usually expressed as mean lengths with variability expressed by sample standard deviations.[1,73] More quantitative considerations about error analysis and determination of the number average molecular weights have recently been reviewed.[1] The length unit most commonly used is kb (kilobase);[74] it can be used for double-stranded DNA where it means 1000 base pairs, or for single-stranded DNA and RNA, where it means 1000 nucleotides or bases.

Determination of molecular weights of nucleic acids by electron microscopy is usually more accurate than with other methods and uses less material. Stüber and Bujard have determined the absolute molecular weights of a number of linear and circular phage DNAs that are frequently used as length standards.[73]

It is possible also to determine the DNA concentration by electron microscopy by mounting a sample of defined molecular weight with a distinguishable DNA standard of known molecular weight and known concentration.[1,75] As little as 5 ng of DNA is needed for these determinations.

* In addition, linear molecules may give greater standard deviations because the ends are more difficult to define.

IV. PARTIAL DENATURATION MAPPING

Differences in the stability or reactivity between AT- and GC-base pairs can serve to establish denaturation maps which reflect the base composition in different regions of double-stranded DNA. The physical maps thus obtained have been of great value for structural and functional studies of many bacteriophage genomes. The denaturation mapping method was first introduced by Inman.[14,76,77] Double-stranded DNA is incubated at high pH or high temperature in the presence of formaldehyde. The formaldehyde reacts with the non-hydrogen-bonded bases and prevents their reassociation. When the DNA is prepared for electron microscopy by one of the spreading methods the denatured stretches show up as single-stranded denaturation loops. The position and the extent of the loops can be measured and the resulting denaturation maps can be combined into a histogram which is characteristic for a homogeneous population of DNA molecules. Specific denaturation patterns can only be obtained with DNA molecules that have local differences in A + T content. The extent of the denaturation depends on the pH, the temperature and time of incubation, the formaldehyde concentration, and the ionic conditions. Apparently the formaldehyde destabilizes the duplex regions not denatured by the pH and heat; as a result, the size of specific denaturation loops varies considerably. The uncertainty is usually in the order of several hundred bases so that it is difficult to obtain characteristic patterns for DNA fragments smaller than 2 to 3 kb.

Different methods have been proposed for partial denaturation of DNA molecules besides heating in the presence of formaldehyde ,[76-79] and alkaline denaturation in the presence of formaldehyde.[14,80-82] Wolfson et al.[83] suggested the spreading of DNA molecules and heteroduplex molecules from high formamide concentrations, at which DNA is partially denatured.

Choosing salt conditions under which the difference in melting between AT and GC base pairs is increased, one can obtain partial denaturation at neutral pH and low temperature.[84] Delius and Clements have thus used high concentrations of sodium perchlorate in a formaldehyde-containing buffer. These conditions lower the melting temperature of DNA and increase the difference in melting between AT- and GC-rich regions.[85]

Proteins that have a selective affinity for single-stranded DNA over double-stranded DNA, like the gene-32 protein of phage T4 or the DNA unwinding protein of *E. coli*, have been used for partial denaturation mapping.[45,86,87] Delius et al.[86] have shown that when λDNA is incubated with the gene-32 protein, the A + T rich regions that are preferentially denatured and covered with protein are the same as the ones that appear single-stranded in alkali denatured molecules. Progressive denaturation is obtained by changing the protein to DNA ratio[86] or by lowering the ionic strength.[45,88] Comparison of alkali denaturation maps with *E. coli* unwinding protein-induced denaturation maps suggested that the *E. coli* protein has a preference for expanding denaturation loops, probably as a result of its highly cooperative binding.[87] As a result, the denaturation pattern does not show as much fine structure. Similar patterns on phage PM2 DNA were obtained with the gene-32 protein denaturation and the alkali denaturation method.[45]

Instead of formaldehyde, Johnson[89] used glyoxal for the fixation of heat-denatured regions. The partial denaturation maps were reproducible and comparable to the alkaline denaturation maps. The use of glyoxal allowed monitoring the extent of denaturation with a fluorescence assay based on the enhanced fluorescence of ethidium bromide intercalated into the remaining duplex region.[89]

In general, where different techniques have been compared, the alkaline formalde-

hyde method gave best differentiation between AT- and GC-rich regions.[14,78,80] Even then, individual molecules often show variations in the denaturation pattern, and only by constructing a denaturation histogram for a whole population of molecules is it possible to get a characteristic pattern.

Partially denatured molecules can be mounted for electron microscopic examination by any of the methods allowing spreading of single-stranded DNA and distinction of single- and double-stranded parts of a molecule. Usually the formamide spreading in combination with cytochrome c film[14] or with the BAC film[32,34] is used. Inman and Schnös have devised a microversion of the Kleinschmidt spreading technique that minimizes the amount of material used.[14] Samples of 5 μl are spread on a drop of distilled water. The use of distilled water as hypophase results in a much cleaner and smoother background than is the case with buffered formamide; however, the contrast is reduced because less cytochrome collapses around the molecules. The molecules denatured by DNA binding proteins were either spread in cytochrome monolayers[86,87] or adsorbed on positively charged carbon films.[45,88] Before spreading, the complexes are fixed with glutaraldehyde and purified by Sephadex® chromatography to remove the unbound protein. Examples of partially denatured molecules are shown in Figures 3 and 6.

Introduced at a time when other physical mapping methods (e.g., restriction enzyme mapping) were unknown, the partial denaturation technique was initially used mainly for orientation of DNA molecules. Partial denaturation of heteroduplex molecules allowed construction of a physical map which could be correlated with the genetic map.[82,83] The replication origins of phage λ, P2, and T7 DNAs have been mapped;[80,83,90] the direction of replication and structure of branch points were studied on λ DNA.[81] Initiation site and direction of transcription have been mapped on partially denatured transcription complexes.[78] Nowadays it is probably easier and less time consuming to orient certain features of a molecule (enzyme binding site, replication fork, deletions or insertions, transcription initiation site) with respect to restriction enzyme cleavage sites.

The denaturation mapping methods may still be of value in other contexts where biological functions are to be correlated with the local base composition of a DNA molecule. In order to determine whether *E. coli* RNA polymerase binding sites were correlated with AT-rich regions, denaturation maps were compared to enzyme binding maps.[34,44] It was shown that the single-stranded regions that can be detected in superhelical DNA are the same as the earliest melting regions in linear or relaxed circular phage PM2 DNA molecule.[45] Comparing the denaturation map of phage P1 DNA with a genetic map revealed that AT-rich regions may be preferential transposition target regions for IS_1 and IS_1-flanked transposons.[91]

V. DNA:DNA HYBRIDIZATION

A. Heteroduplex Method

The heteroduplex method is based on DNA:DNA hybridization and allows one to detect and map regions of homology between two DNA molecules.[15,22] The principle of the method is the following: when two DNA molecules with partial homology are denatured and renatured together, homologous regions will reanneal to form duplex DNA, whereas regions which are not complementary in base sequence do not pair and remain single-stranded. Initially, these regions had been visualized as bushes of collapsed single-stranded DNA on the heteroduplex molecule.[15] The position, but not the accurate length, of deletions or insertions in phage λ heteroduplex molecules could be measured.

The introduction of formamide for the hybridization reaction and the spreading

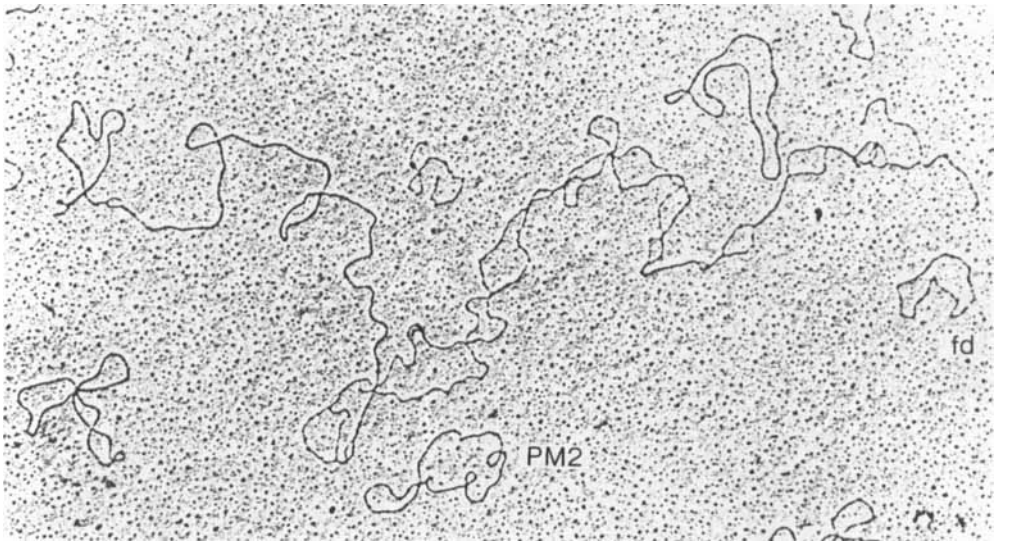

FIGURE 6. Phage λ DNA molecule partially denatured in alkali and spread according to the method of Inman and Schnös.[14] Reference molecules are PM2 double-stranded and fd single-stranded circles. (Courtesy of J. Meyer, Biozentrum, University of Basel, Basel, Switezerland, 1979.)

procedure improved both the condition for heteroduplex formation and for visualization of single-stranded DNA.[12] Discrimination between double- and single-stranded DNA made it possible to map very accurately the size and position of insertions and deletions and of branch points between homologous and nonhomologous regions (Figure 7).

A detailed description of the original heteroduplex method was given by Davis et al.[9] In a recent review, Ferguson and Davis[1] again treat the heteroduplex method extensively, covering improvements in the experimental procedure as well as problems arising in construction and interpretation of heteroduplex molecules and emphasizing the quantitative approach to the technique.

The original procedure, particularly adapted for lambdoid phages, involved denaturation of DNA (or intact phages) in 0.1 *M* NaOH, 20 m*M* EDTA at room temperature, followed by neutralization with 0.1 *M* Tris and renaturation in 50% formamide. For λ DNA these conditions are about Tm-30° C and give the optimal rate for DNA reassociation.[9] For plasmid DNAs or eukaryotic DNA with higher G + C content it may be necessary to increase the denaturation conditions, either by increasing the NaOH concentration (0.2 *M*) or by increasing the temperature (37°C).

When two DNAs of similar molecular weight are hybridized, it is important to use equimolar amounts of both molecules for optimal heteroduplex formation. This is particularly recommended in hybridization reactions involving DNA molecules with short homologies, where homoduplex formation is favored. If one starts directly with phage,[9] it is more accurate to calculate the concentration by optical density than by titration on plates; phage particles may be damaged and not form plaques, although their DNA is still intact. When isolated molecules of different size are hybridized, however, it is advantageous to use an excess of the smaller fragment to favor heteroduplex formation,[20,92] the reannealing of homologous longer molecules being facilitated by higher probability of nucleation. When the homology region lies inside the long molecule the ratios of small to long fragment has to be increased further because of unfavorable conditions due to the excluded volume effect.[9]

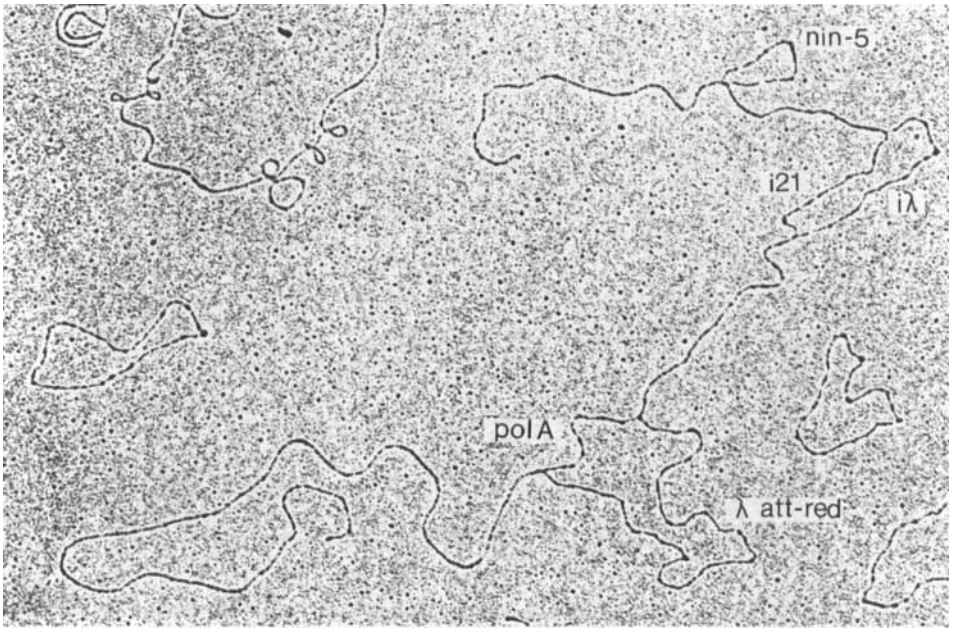

FIGURE 7. Heteroduplex molecule between λ DNA and λ polA (att-red) ∇ imm^{21}nin5, a phage carrying the *E. coli* polA gene.[192] The sample was prepared according to Davis et al.[9] (This picture was taken during a European Molecular Biology Organization, course on recombinant DNA, Biocenter, Basel, 1977.)

To obtain unambiguous products in a heteroduplex preparation, it is very important to start off with intact DNA molecules. Davis et al.[9] therefore recommend beginning with DNA isolated directly from purified bacteriophage. Alkali treatment results in simultaneous lysis of the phage and denaturation of the DNA. It is more difficult to obtain intact, nick-free DNA after chemical extraction, and particularly so after restriction enzyme cleavage. The presence of fragments in the hybridization may lead to branched structures and reassociation of several molecules. Even when intact molecules have long regions of nonhomology, the single-stranded part of one heteroduplex can hybridize to unreacted single strands or to other heteroduplexes leading to complicated aggregates. In order to minimize such problems, it is advisable to under-renature the DNA; renaturation to only ~50% is optimal.[1,9] Under the standard conditions and for λ DNA, this is achieved in 1 to 2 hr at room temperature. The conditions for renaturation to 50% depend on the size, the concentration, and the base composition of the DNA sample. Wellauer et al.[92] have hybridized short rDNA fragments from *Xenopus laevis* with high molecular weight rDNA under the following conditions: short DNA threefold excess, renaturation for 2 hr at 30°C in 70% formamide, 0.1 *M* Tris, 0.02 *M* EDTA, pH 8.5.

The heteroduplex method has been successfully applied to a variety of problems: mapping the position and length of deletions, substitutions, insertions, and duplications in phage DNAs[9,15,22] and animal virus DNA;[93] demonstration of the circular permutation and terminal redundancy in certain phage DNAs;[94] and arrangement of genes and spacers in rDNA of *Xenopus laevis*.[92] The phenomenon of branch migration has been discovered in heteroduplex molecules,[94] and coding regions for an immunoglobulin variable region gene have been mapped by hybridizing Ig cDNA to a cloned mouse DNA fragment.[95]

Ferguson and Davis[1] have introduced a terminology for the description of different types of branched molecules. This code, consisting of two numbers, represents the

number of single strands in a branch, followed by the number of double strands. Thus, the branch structure of a deletion loop in a heteroduplex molecule would be a 2:2 branch, where two single strands meet with two double strands. The possible existence of swivel points and the resulting stability of different branch structures has been discussed.[1]

B. Mapping of Restriction Fragments

A particular application of DNA:DNA hybridization that does not involve two different molecules, but DNA fragments and whole molecules of a same type, is the mapping of restriction fragments. Two different electron microscopic methods have been described for localizing fragments on phage DNAs, ranging in size from several kb to ~200 nucleotides.

1. D-Loop Mapping

The cleavage sites for the restriction endonuclease *HindIII* on PM2 DNA have been determined by Brack et al. in the following way.[20] Individual restriction fragments purified by agarose gel electrophorsis were hybridized with full length linear DNA (obtained by cleavage with *HpaII*). A molar excess (10- to 100-fold) of the small fragment was mixed with the linear DNA in the presence of 50% formamide in a buffer containing 50 m*M* ammonium acetate, 5 m*M* Tris-HCl pH 8.0, 0.5 m*M* EDTA. This mixture was heated for 10 min to 60°C to denature the DNA, and hybridization was carried out at 37°C overnight. These conditions were calculated to be 14°C under Tm for PM2 DNA.[26] A small proportion of the renatured molecules (1 to 5%) contained loops at the position of the restriction fragment (Figure 8). These D-loops are formed in the following way: hybridization is driven by the excess of fragments, and hybrids between one strand of a fragment and one strand of a full length molecule are rapidly produced. Two such partially double-stranded hybrids then hybridize with each other to form fully double-stranded loop structures, or a hybrid strand anneals to another single-strand forming a loop that is double-stranded on one side only.

This method allowed mapping of fragments as small as 200 nucleotides long and also discrimination between a mixture of two fragments of nearly identical size by their hybridization to different positions. The disadvantage of the method is the rather low yield of hybrids. The D-loop molecules are probably rather unstable because the fragment can be easily displaced by branch migration. Few measurements are, however, sufficient to give a very accurate localization of restriction fragments.

The term "D-loop molecule" was chosen in analogy to the natural "D-loop molecules" that had been observed as intermediates of DNA replication in mitochondrial DNA.[96,97] The D-loop method has been used also to demonstrate the homology between two independently isolated, presumably identical, DNA clones. The easiest way to determine whether two independently cloned fragments are identical seems to be to make a heteroduplex experiment, either with the DNA fragments or with the hybrid molecules of vector DNA containing the cloned fragment. Interpretation of the results may, however, not be so straightforward. If, for instance, one observes only double- and single-stranded fragments, the DNAs may either be completely homologous or they have no homology at all, in which case the double-stranded fragments are homoduplexes. On the other hand, the two hybrid molecules can pair if the cloned segments are identical and inserted in opposite directions. If they are inserted in the same orientation, they form a nonhomology loop, which is again ambiguous. The interpretation becomes straightforward when one fragment is hybridized with one intact hybrid molecule; D-loop molecules can form independently of the direction of insertion of the cloned, identical fragments.

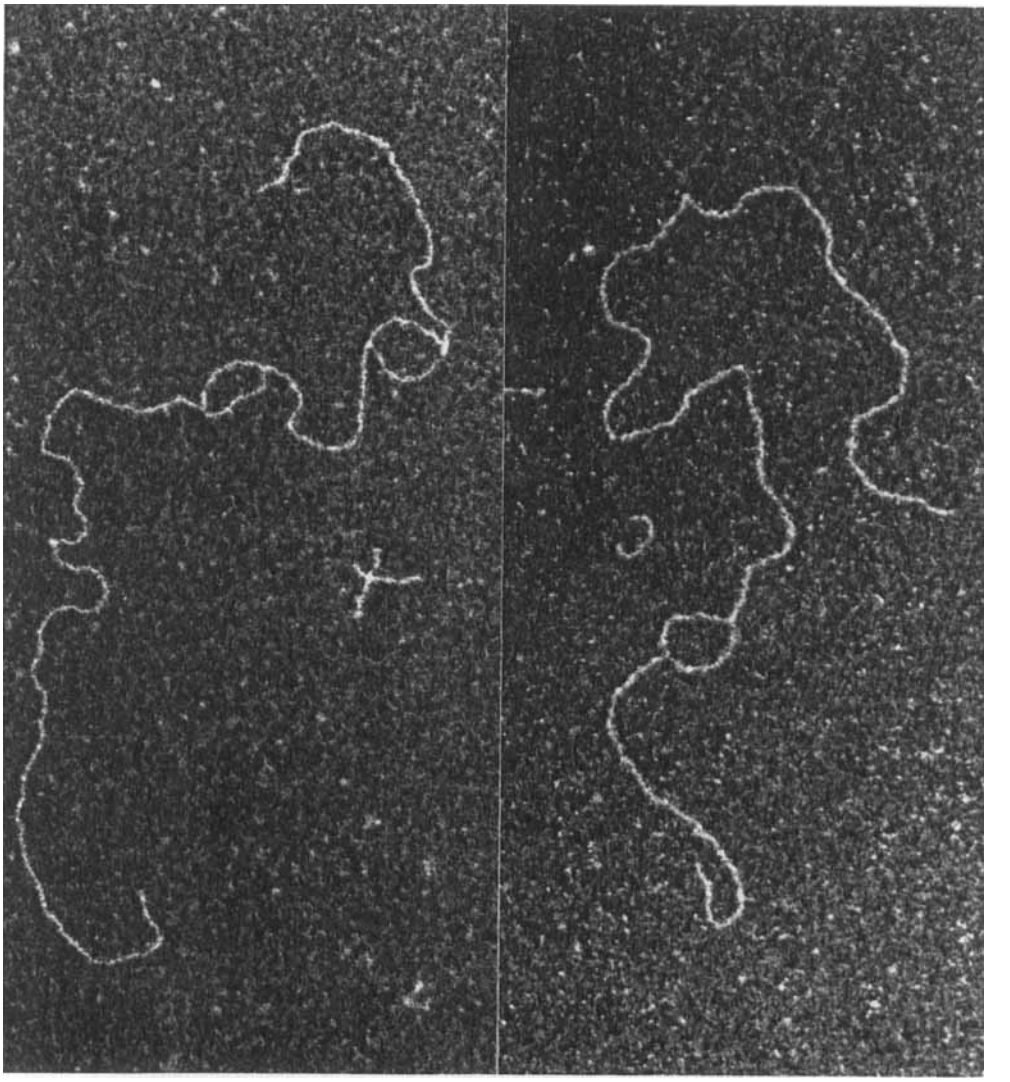

FIGURE 8. D-loop molecules. PM2 restriction fragments *HindIII* f4 and f5 were hybridized to linearized PM2 DNA.[20] Samples were spread with formamide, mounted on carbon films, stained with uranyl acetate, and photographed in dark field.

2. "H-Form" Molecules

To avoid the problem of branch migration in hybrids between restriction fragments and intact T7 DNA molecules, Bick and Thomas[98] have proposed the following procedure:

1. Restriction fragments are purified, denatured, and hybridized to separated intact *l* strands of T7 DNA. Hybridization is carried out in 70% formamide, 1.0 *M* NaCl, 0.2 *M* Tris pH 7.6, and 20 m*M* EDTA, which corresponds to ~Tm-25°C.
2. The hybrids are treated with glyoxal: 0.5 *M* glyoxal, 0.055 *M* sodium phosphate, pH 6.9, incubation at 37°C for 1 hr. Glyoxalation of G residues[99] in the unhybridized single-stranded region prevents subsequent renaturation of these unprotected sequences.
3. The glyoxalated hybrids are melted by heating to 95°C to displace the fragment.
4. An equimolar amount of separated r strand of T7 is added and allowed to renature (in the same formamide buffer as above).

The r strand can hybridize only to regions that had been protected from glyoxalation reaction by the annealed restriction fragment. Thus, "H-form" molecules are formed with fragments located inside the molecule. Terminal fragments produce "Y"-form hybrids. A small number of D-loop molecules was also observed, which probably resulted from incomplete glyoxalation of the single strands and reannealing of the fragment. The yield of well-defined "H-form" molecules was between 30 and 40%. The shortest fragment mapped with this method was about 400 nucleotides long.[98]

C. Triple Hybrids, Diheteroduplexes

A triple hybrid, also called heterotriplex, double heteroduplex, or diheteroduplex, is a structure formed by hybridizing together three different molecules and shows the sequence relation among these molecules. It can consist either of three DNA, or of DNA and RNA molecules. The term "heterotriplex" was first used for hybrids between T7 mRNA and T7 DNA which actually corresponded to R-loop molecules constructed by a two-step method.[100] Another example of a triple hybrid are the R-hybrids discussed in the following section (VI.A.5); they are heteroduplex molecules containing an R-loop. Also, the D-loop molecules described above are a type of triple or quadruple hybrid. We shall use the term "diheteroduplex" only for hybrids consisting of three different DNA molecules and the more general term, "triple hybrid", for mixed RNA-DNA hybrids.

Construction of diheteroduplex molecules has allowed the determination of the orientation of SV40 DNA sequences in the adeno-SV40 hybrid DNA Ad2 + ND1 and identification of the SV40 sequences in a heteroduplex between Ad2 + ND1 × Ad2 DNAs.[101] Similarly, the orientation of the P4 EcoRI fragments A and D carried in a λ transducing phage were determined by diheteroduplex formation.[1]

Diheteroduplex formation between three cloned mouse DNA fragments carrying immunoglobulin gene sequences — one isolated from myeloma DNA and two isolated from embryonic tissue — provided unequivocal proof that a complete immunoglobulin gene is created by somatic recombination.[25] Figure 9 shows such a diheteroduplex formed between three mouse DNA fragments containing Ig κ genes. The branch point corresponds to the recombination site between the two embryonic DNA fragments.

D. Intramolecular Hybrids

Hybridization of homologous sequences within one DNA strand leads to specific snap-back or fold-back structures. These intramolecular hybrids can be used for a variety of purposes: demonstration of weak homologies or inverted repeat sequences and construction of secondary structure maps.

1. Weak Homologies

Ferguson and Davis[102] have been able to detect a weak sequence homology between SV40 and polyoma DNA by intramolecular hybridization. SV40 DNA was covalently joined to polyoma DNA by ligation of the two *EcoRI* cleaved molecules. When denatured, the single-stranded hybrid molecules formed characteristic snap-back structures. The region of homology was too short (<200 nucleotides) for the hybrid segment to be measurable, but the crossover or contact point was reproducible and could be mapped at a unique position.

2. Underwound Loops, Inverted Repeat Sequences

Broker and co-workers[103] described a structure called the "underwound loop"[103] that occurs in renatured molecules containing inverted repeat sequences. Denatured DNA containing inverted repeat sequences can form specific snap-back or stem-loop

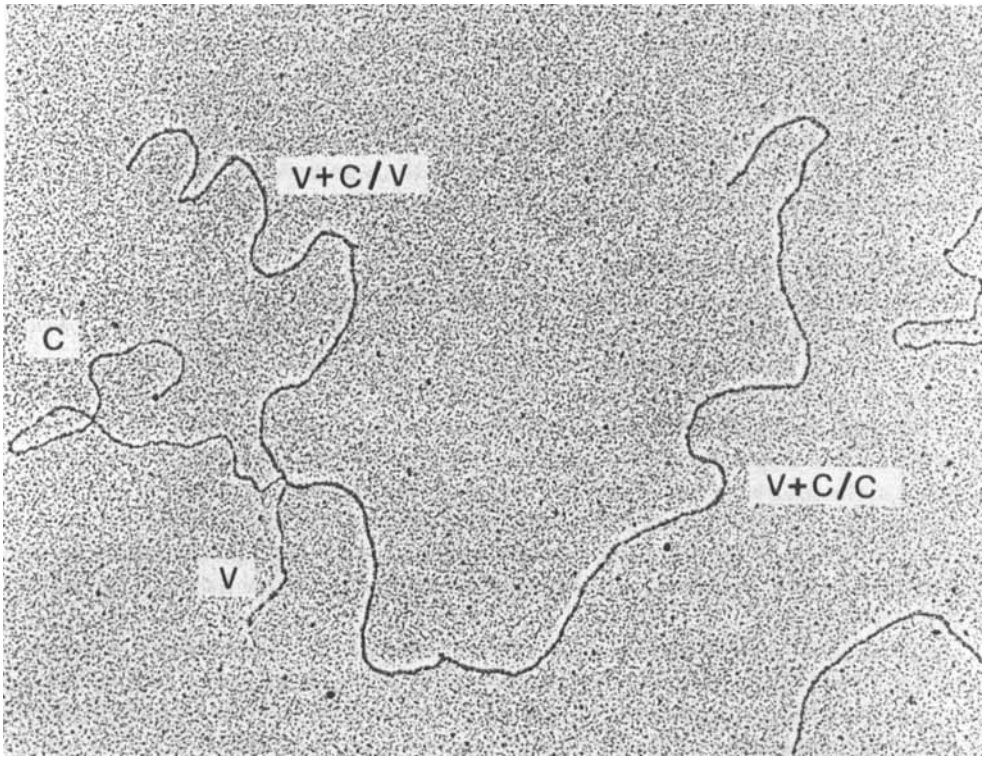

FIGURE 9. Diheteroduplex molecule constructed by hybridizing together three cloned mouse DNA fragments containing immunoglobulin κ chain genes.[91] The branch point defines the J (joining) region[25] or recombination site. The two double-stranded arms of the hybrid correspond to the parts of the myeloma DNA (V + C) that are homologous to the constant (C) or the variable (V) region gene fragments isolated from embryonic DNA.

structures. When two complementary strands, each containing a snap-back structure, pair with one another, steric hindrance may prevent complete reannealing and lead to the formation of underwound loops. These are partially double-stranded, partially single-stranded structures that are produced as follows. During the early stages of reannealing, intrastrand stem-loop structures form with first-order kinetics when the inverted duplications pair. In a slower second-order reaction, complementary strands reanneal. If complementary sequences in the loops start to pair, the stem-loop structures are trapped in a metastable state. Because the two hybridizing loops cannot interwind, they form a variable pattern of single- and double-stranded segments.[103] Each left-handed turn is compensated by a right-handed helical turn. The explanation for the formation of this so-called "form V DNA" was recently presented by Stettler et al.[104]

These structures are very characteristic and much more easily detected than short homology regions (crossovers) in single-stranded molecules. They have been used to identify short inverted duplications on *E. coli* F factor DNA that are only about 20 to 50 nucleotides long.

We have observed that certain immunoglobulin gene-containing mouse DNA clones frequently lost specific segments by deletion. When these molecules were denatured and self-annealed, specific snap-back loops and underwound loops were detected (Figure 10). They indicate the presence of inverted repeat sequences which may lead to deletions by intramolecular crossing over.

3. Secondary Structure Maps

Secondary structure mapping of RNA has been used to elucidate the structure and processing pathways of HeLa cell ribosomal RNA.[105] Wellauer and Dawid observed that single-stranded ribosomal DNA showed an equally regular and reproducible arrangement of secondary structure hairpin loops.[106] Regions coding for the 40s rRNA could be identified on single-stranded DNA by its resemblance to the secondary structure of rRNA. By comparative analysis of the secondary structure maps of the amplified rDNA from two related Xenopus species, it was shown that the pattern in the gene region was identical, but that differences existed in the nontranscribed spacer regions.[107] To obtain reproducible secondary structure in denatured DNA, Wellauer has used the following modifications of the cytochrome spreading method: the hyperphase consisted of 0.5 to 1 μg/mℓ DNA, 10 m*M* NaCl, 17 m*M* EDTA (pH 8.5), 10% formamide, 0.5 *M* urea, and 30 μg/mℓ cytochrome c; it was spread on distilled water and picked up on freshly prepared carbon grids.[106]

Characteristic hairpin loops have been mapped on single-stranded SV40 DNA,[108] suggesting the presence of inverted repeat sequences which could cross-hybridize with each other. The hairpin loops could be observed only when molecules were spread in the presence of 1 *M* ammonium acetate in the 50% formamide spreading solution.

Secondary structures in single-stranded polyoma DNA have been demonstrated with the gene-32 protein mounting technique.[109] Three reproducible fold-back structures were observed: a hairpin loop close to one end of the linearized molecule and a small and a large loop bounded by short inverted repeat stems of 40 ± 20 base pairs. The loops and hairpin were distinguishable from random crossovers by their reproducible dimensions and positions. The low frequency of occurrence of the different secondary structural features (around 10% at 37°C and 10 to 50% at 22°C) was thought to be due to the short length of the paired stem region and/or to partial sequence mismatching. The position of the secondary structure loops and the hairpin suggested that the short inverted repeat sequences lie at the splicing sites of late and early messengers, respectively.

4. Psoralen Cross-Linked Secondary Structure

Secondary structure maps vary with the ionic conditions under which the molecules are prepared and reflect the degree of homology between different paired regions. In order to "freeze" the secondary structures of single-stranded DNA molecules at different ionic strengths, Shen and Hearst[110] cross-linked the hairpins with the psoralen derivative Trioxsalen (4,5',8-Trimethylpsoralen). The use of Trioxsalen cross-linking of DNA in combination with electron microscopic examination has been introduced by Hanson et al.[111] After intercalation into the DNA double helix, Trioxsalen and other psoralen derivatives can covalently cross-link pyrimidines on opposite DNA strands when irradiated with UV light of long wavelength. After cross-linking, the DNA can be spread under totally denaturing conditions[111] or under isodenaturing conditions[110] for the observation of the cross-linked structures.

The secondary structure maps of single-stranded phage fd DNA have been observed after cross-linking the DNA at different ionic strengths.[110] As the salt concentration was increased, more hairpin loops appeared, and all of them could be mapped at specific positions corresponding to promoter regions. The different stabilities of hairpins are probably a result of different degrees of base mismatching (see also discussion in Section VII).

Cech and Pardue[112] have approached the question of whether the inverted repeat sequences or palindromes observed in eukaryotic nuclear DNA can exist as "cruciform" fold-back structures in vivo. Mouse tissue culture cells were treated with Tri-

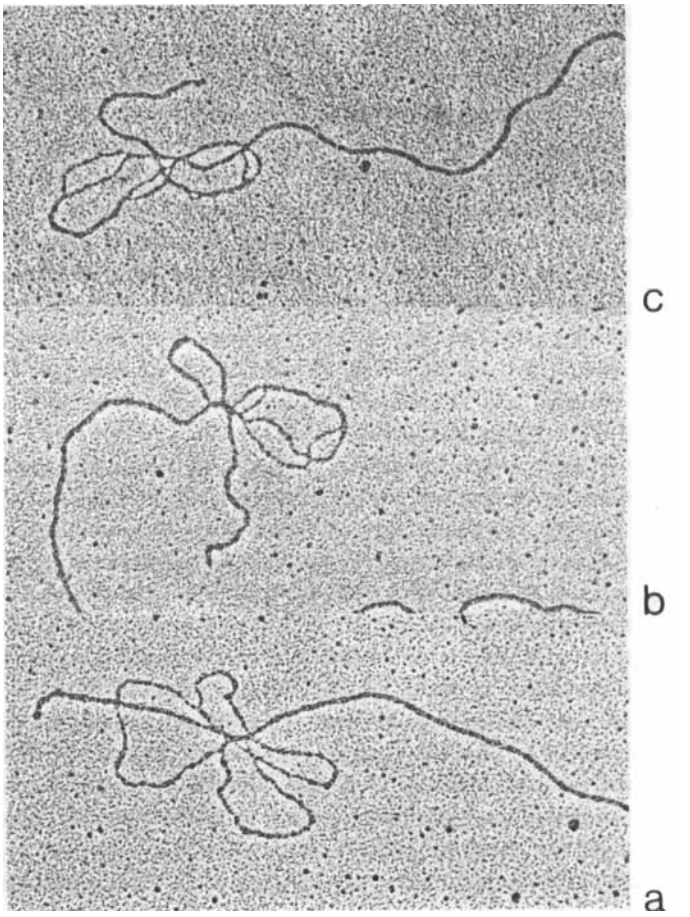

FIGURE 10. Formation of underwound loops. A DNA molecule containing inverted repeat sequences has formed two snap-back loops. The arms of the snap-back molecules can pair normally (a), but the loops cannot fully base-pair because of steric hindrance and form partially single-stranded and double-stranded underwound loops (b, c).

oxsalen and irradiated with UV light in order to cross-link the DNA inside the living cells. DNA was then extracted and spread for electron microscopy either from 50% formamide to observe secondary structures or after complete denaturation with glyoxal in order to check for the cross-linking frequency. Under the conditions used, the frequency of cross-links was about one per 200 to 300 nucleotides, which should have been high enough to cross-link average size inverted repeats in mouse DNA. No secondary structures could be observed in the in vivo cross-linked DNA, and it was concluded that the repeat sequences do not form secondary structures in vivo.[112]

VI. RNA:DNA HYBRIDIZATION

A. R-Loop Formation

The R-loop method developed by White and Hogness[113] and Thomas et al.[114] allows mapping of DNA sequences complementary to specific RNA molecules. These authors made the following observation: when single-stranded RNA is incubated with duplex DNA containing sequences complementary to the specific RNA in the presence of

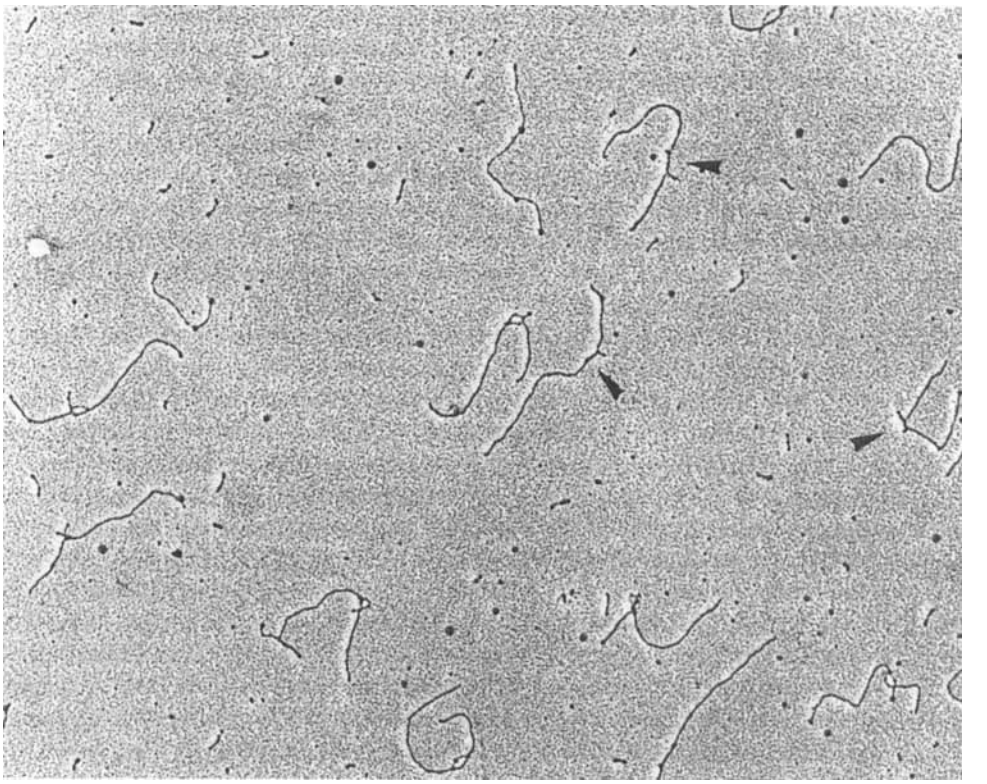

FIGURE 11. DNA fragments saturated with R-loops. A 3-kb fragment (XbaI cleaved) of Ig303λ[5,117,146] was hybridized with Igλ mRNA. Under the high formamide (70%) spreading conditions, some of the R-loops are collapsed by lateral aggregation (→).

formamide and at temperatures close to the Tm of the DNA, a characteristic hybrid structure, called an R-loop, is formed. The RNA hybridizing to its complementary sequence displaces the second DNA strand; this RNA:DNA hybrid forms one part, the displaced DNA strand the second part of the R-loop bubble (examples, see Figures 11 and 16). The R-loop molecules are stable and can be mounted for electron microscopic observation by the formamide spreading method. Measurement of the position and size of the loops allows accurate mapping of the location, structure, and size of specific structural genes or rRNA genes.

The R-loop formation is possible because under the conditions used the RNA:DNA hybrids are more stable than the DNA:DNA duplex. McConaughy et al.[26] had already pointed out the advantage of using formamide solutions for nucleic acid hybridization, because they permit specific DNA:DNA and DNA:RNA associations to occur at lower temperatures. A linear relation exists between the formamide concentration and the Tm of DNA: the Tm of native duplex DNA is reduced by approximately 0.72°C per 1% formamide.[26] Whereas in aqueous solutions DNA:DNA hybrids are more stable than DNA:RNA hybrids,[115] the situation is reversed in formamide solutions.[116,117] In an extensive study, Casey and Davidson[117] have examined the rate and extent of DNA:RNA association and the thermal stability of the hybrids in different formamide concentrations. These experiments show that in high formamide concentrations (70 to 80%) the difference in stability between DNA:DNA and DNA:RNA hybrids is maximal. The thermal stability of different nucleic acid duplexes in the high formamide solution can be summarized as follows: RNA:RNA > RNA:DNA > DNA:DNA. Thus,

RNA:DNA hybridization is maximal at conditions where no DNA:DNA association occurs.[117] The fact that RNA:RNA duplexes are more stable than RNA:DNA hybrids can be important for the interpretation of R-loop molecules, as we shall discuss below.

In general, G + C rich duplexes are more stable than A + T rich regions. As Casey and Davidson have shown, the destabilizing effect of formamide is least in G + C rich RNA:DNA hybrids; i.e., the difference between DNA:DNA and RNA:DNA hybrid stability increases with increasing G + C content. Other denaturing solvents, such as concentrated perchlorate and trichloracetate solutions, also favor RNA:DNA association over DNA:DNA association, the optimal temperature of the former being 5 to 12°C above the Tm of duplex DNA.[118]

This differential thermal stability of hybrids has been the basis for the development of the R-loop method. Using as a model system a hybrid phage carrying a fragment of yeast DNA with ribosomal genes plus yeast rRNA, Thomas et al.[114] studied various conditions for the rate of R-loop formation and the thermal stability of R-loops. They have shown that the temperature for the maximal rate of R-loop formation (Tmax) is within one degree of the strand separation temperature (Tss). This temperature depends on the base composition of the DNA and can be experimentally determined for any given DNA fragment as follows. The test DNA is incubated in the formamide R-loop buffer: 70% formamide, 0.1 *M* Pipes [piperazine-N,N′-bis(2-ethanesulfonic acid)] pH 7.8, 0.01 *M* Na_2 EDTA (+ NaCl). The sample is heated at increasing temperatures and aliquots are taken at 1°C intervals, starting at 45°C, and examined in the electron microscope. The temperature at which half of the duplex DNA is irreversibly converted to single-stranded DNA is defined as the strand separation temperature (Tss). (The Tss is conceptually different from the melting temperature Tm, where half of the base pairs is no longer paired.)[114]

Thomas et al.[114] have also studied the rate of R-loop formation as a function of RNA concentration and ionic strength. As in DNA renaturation, the rate of R-loop formation is a function of the ionic strength. The reaction rate increases about 15-fold in the presence of 0.5 *M* NaCl in the buffer. Above the Tss, the rate of R-loop formation is directly proportional to the RNA concentration. At a molar excess of RNA over DNA, DNA molecules can be saturated with R-loops (see example, Figure 11).

The following equation shows the variation of Tmax with base composition, ionic strength, and formamide concentration: Tmax = 81.5 + 0.50 (% G + C) + 16.6 log Na^+ − 0.60 (% formamide).[114] The authors state that this equation may not apply if the conditions are significantly different from the ones they used. Applying this formula, two different conditions were calculated for the R-loop formation with mouse immunoglobulin DNA and immunoglobulin mRNA; similar results were obtained either by incubation at high ionic strength (0.56 *M* Na^+) and high temperature (57°C) or at low ionic strength and low temperature (0.12 *M* Na^+, 51°C).[25,119]

Once formed, the R-loops are quite stable.[114] They can be stored for several days in the formamide solution, or even dialyzed against a buffer without formamide. This allowed further cleavage of the hybrid molecules with restriction endonucleases for more accurate mapping of the R-loop position. One problem occurring during storage or during sample preparation for electron microscopy is the partial displacement of the RNA by branch migration. We shall deal with this problem in a separate chapter.

In theory it should be possible to calculate the optimal conditions for R-loop formation for any given DNA segment. In practice, however, working with rare mRNAs, with genes of unknown base composition, with several genes having different G + C content on the same DNA molecule, and with total mRNA and whole-cell DNA, it may not always be possible to obtain the ideal conditions described in the model system of ribosomal DNA and rRNA.

In the mouse immunoglobulin gene system, for example, the conditions have been worked out fairly accurately.[25] However, the yield of R-loop molecules varied in different experiments from 25 to 85%. In one case the low yield (20 to 30%) could be explained by the fact that the gene isolated was only cross hybridized, because it was not completely homologous to the mRNA used for R-loop formation.[120] But even in homologous hybridization reactions, however, the yield varied considerably. This may have had technical reasons: because the amount of purified mRNA was the limiting factor, we had to scale down the reaction volume to a minimum (17 μl), where accurate pipetting becomes difficult.

Stable secondary structure formation in certain RNAs also prevents complete hybridization to the complementary DNA sequence and reduces the yield and stability of R-loops (see next section).

In the following we describe some of the problems that are frequently encountered in the construction and interpretation of R-loop molecules, together with recent modifications and improvements of the method.

B. Two-Step R-Loop Method; Single-Stranded DNA:RNA Hybrids*

One modification of the R-loop method introduced by Holmes et al.[121] is the "two-step method". Here the DNA is first denatured, and then single-stranded DNA:RNA hybrids are formed and allowed to reanneal, forming R-loops. In a first incubation (in 2 × SSC, 80% formamide, 50°C for 1 hr), RNA:DNA hybrids are formed, but the conditions suppress DNA:DNA association. Subsequent lowering of the temperature in the high formamide concentration (1 hr incubation at 25°C) allows DNA:DNA association to occur without displacement of the mRNA.

In the author's laboratory, the following procedure is routinely used for constructing both single-stranded hybrids and R-loops. The complete R-loop mixture containing RNA (5 $\mu g/ml$), DNA (5 to 10 $\mu g/ml$), and 0.5 *M* NaCl in the formamide buffer[25] is heated to about 10°C above the Tss (65 to 70°C) for 5 min to denature the DNA. The samples are then placed in a 60°C waterbath with controlled temperature shift; the temperature is slowly lowered (by 1°C/ 3 hr) passing from 60°C through the RNA:DNA association temperature to the Tss (56°C) of the DNA. Aliquots can be taken out at any time point and examined in the electron microscope. It is thus possible to observe the hybrids both in the single-stranded and in the double-stranded R-loop forms.[122,131,152] Chow et al.[123-126] used a similar procedure for their extensive study of the complex splicing patterns of late and early mRNAs of adenovirus. After denaturation of the DNA in the R-loop buffer, the samples were incubated at different temperatures between 50 and 60°C. Alternatively, mRNAs were annealed to purified R strand of adenovirus restriction fragments by gradually cooling the R-loop mixture from 52 to 35°C.[127]

Single-stranded DNA:RNA hybrids had been visualized directly long before the introduction of the R-loop method. Messenger RNAs were mapped on T7 DNA[128] and ribosomal RNA genes on several bacterial plasmids[129-130] by mounting the hybrids with the formamide cytochrome monolayer method. Wu and Davidson[58] have developed a method involving complex formation with gene-32 protein and spreading with ethidium bromide for mapping ribosomal genes and tRNA genes, which considerably improved the resolution by allowing better discrimination between duplex regions and single strands. Construction and observation of single-stranded DNA:RNA hybrids presents the following advantages:

* Some confusion exists in the literature concerning the terminology of these hybrid molecules. Single-stranded R-loops, RNA:DNA heteroduplex, or just RNA:DNA hybrids are all synonymous for a hybrid formed between a single-stranded DNA molecule and a complementary RNA molecule.

1. Resolution of the method. Very small hybrid regions and single-stranded DNA loops can be resolved, such as the short exons coding for the hinge region of immunoglobulin heavy-chain genes [122,152] and those coding for the five specific J regions of immunoglobulin κ light chains,[131] all of which are 36 to 39 nucleotides long (Figure 12). Very often, these short hybrid regions could not be observed under conventional R-loop conditions, either because they were not stable enough and were displaced by branch migration, or because the homology region was too short to initiate nucleation of RNA:DNA association in the duplex DNA.
2. Stability. The single-strand DNA:RNA hybrids can be stored at 4°C for several days without much DNA:DNA reassociation.[124] Displacement of the RNA by the second DNA strand is therefore minimized.
3. Hybrid formation in G + C rich and A + T rich regions. Since G + C rich regions of duplex DNA are not completely denatured when R-loop mixtures are heated to the average Tss of the DNA, the effective single-strand concentration of such regions is low relative to that of A + T rich regions, and R-loop formation is negligible.[124] The G + C rich messengers of adenovirus hybridized efficiently when incubated with single-stranded DNA at 59 or 61°C, whereas A + T rich mRNAs hybridized preferentially at lower temperatures, 50 to 53°C. Incubation at gradually decreasing temperatures should therefore allow one to find optimal conditions for any RNA:DNA association, regardless of the base composition. Another way to avoid these difficulties will be discussed below.
4. G + C rich RNA may form secondary structures that are more stable than RNA:DNA hybrids around the Tss and prevent complete R-loop formation.[117] Characteristic secondary structures have been observed in immunoglobulin light-chain mRNAs.[131] When these mRNAs are hybridized with DNA fragments containing related, but not completely homologous sequences under normal R-loop conditions (Tss ± 1°C), R-loops are sparse or incomplete (Fig. 13b). The mRNA maintains its secondary structure and cannot hybridize completely; interpretation and measurement of such R-loops are obviously difficult. When the hybridization is carried out at higher temperature, the mRNA molecules unfold and hybridize completely to the single-stranded DNA, giving measurable RNA:DNA hybrids (Figure 13a).

There are several problems encountered with these methods. Length measuring errors on single-stranded DNA are usually larger than on duplex regions. Adequate single-stranded length standards should be included. It may be difficult to get intact DNA molecules of high molecular weight. The quality of the spreadings should allow good discrimination between single and double-stranded regions.

C. Stabilization of R-Loops: Cross-Linking and Glyoxal Fixation

Two important modifications of the R-loop method have recently been described by Kaback et al.[132] They improve both the formation of R-loops on high molecular weight DNA and the stability of R-loops under various conditions. When one wants to study R-loop formation between very large chromosomal DNA and total cell mRNA, variations in G + C content may require complete melting of DNA duplexes. As discussed above, the chance of obtaining intact (nick-free) duplex molecules decreases with increasing size of the DNA molecule. Kaback et al. have shown that cross-linking of the DNA with Trioxsalen[111] prevents strand separation and does not interfere with R-loop formation.[132] After cross-linking, the RNA can be hybridized at temperatures 1 to 3°C above the Tss, in which conditions all DNA is completely single-stranded. At a cross-linking frequency of about one per 740 base pairs, up to 89% of the expected number of R-loops were obtained. If the cross-links are too frequent, they may result in a

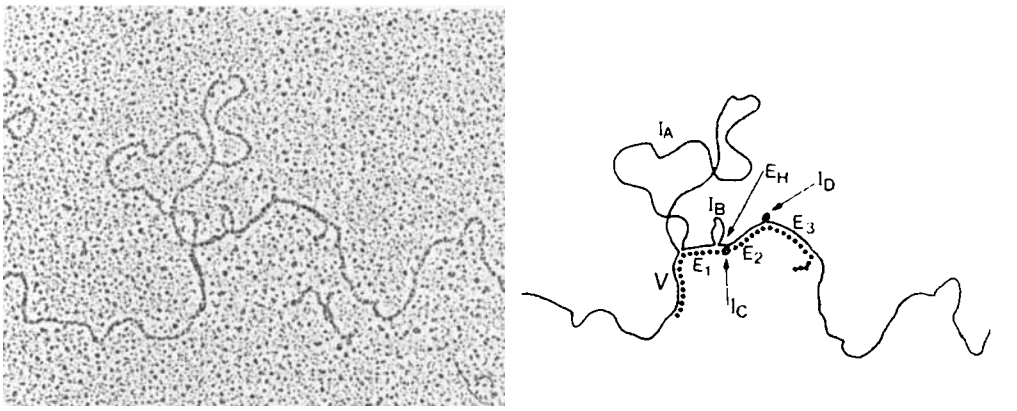

FIGURE 12. Single-stranded DNA:RNA hybrid between a mouse DNA fragment coding for a complete immunoglobulin heavy chain γ_{2b} gene and γ_{2b} mRNA. It is possible to distinguish the small intron loops I_D and I_C that separate the exons E_3 and E_2, and E_2 and the hinge region E_M, respectively. Larger introns I_B and I_A separate the hinge region and E_1 and the constant region gene E_3-E_1 from the variable region gene V. (Courtesy of R. Maki and A. Traunecker.[152])

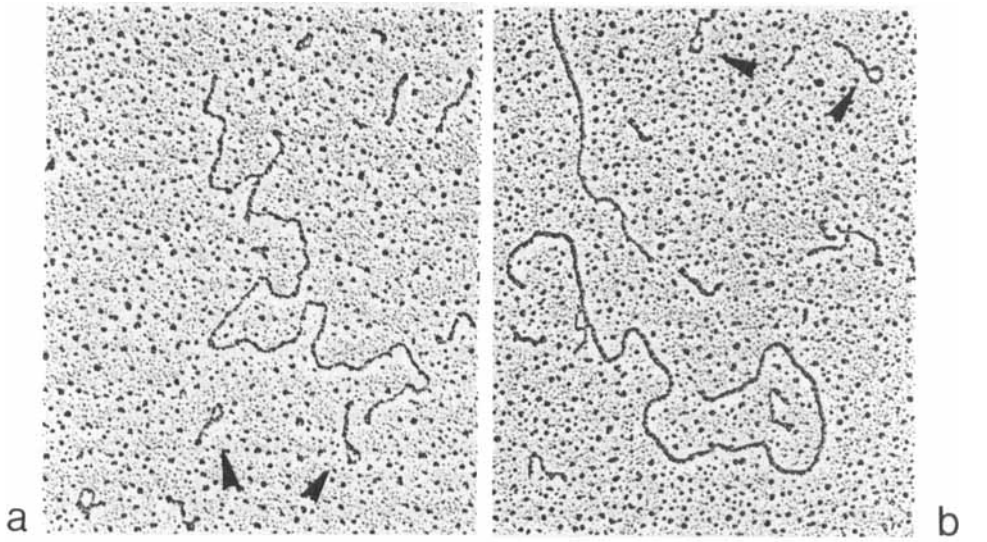

FIGURE 13. Two-step R-loop formation between an immunoglobulin x light chain mRNA (MOPC 321) and a DNA fragment containing MOPC 321-related sequences (Tonegawa and Brack, unpublished results). The x mRNA forms stable secondary structures (→) and cannot hybridize completely to the related sequences under R-loop conditions (Tss ± 1°C); the unfolded mRNA hybridizes to the single-stranded DNA (a). Unhybridized RNA molecules again form secondary snap-back structures under the spreading conditions (→).[131]

shortening of the loops (up to 15% shorter than in untreated samples), or they may interfere with the hybridization of the RNA, and even with the interpretation of the results. Using a cross-linking frequency of one per 2 to 3 kb, the authors were able to study the distribution of ribosomal genes on large molecular weight yeast chromosomal DNA. They observed up to 26 regularly spaced 18s rRNA genes on DNA molecules as long as 236 kb.[132]

Another problem frequently encountered is the stability of R-loops during storage or sample preparation for electron microscopy. Usually, spreading from 50% form-