Trumble, W. R., Viitanen, P. V, Sarkar, H. K., Poonian, M. S., & Kaback, H. R. (1984) *Biochem. Biophys. Res. Commun. 119*, 860.
Viitanen, P., Garcia, M. L., Foster, D. L., Kaczorowski, G. J., & Kaback, H. R. (1983) *Biochemistry 22*, 2531.
Viitanen, P., Garcia, M. L., & Kaback, H. R. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1629.
Waggoner, A. S. (1979) *Methods Enzymol. 55*, 689.
Watson, J. D. (1977) *The Molecular Biology of the Gene*, W. A. Benjamin, Menlo Park, CA.
West, I. C., & Wilson, T. H. (1973) *Biochem. Biophys. Res. Commun. 50*, 5551.
Wilson, T. H., & Kutsch, M. (1972) *Biochim. Biophys. Acta 255*, 786.
Wong, P. T. S., Kashket, E. R., & Wilson, T. H. (1970) *Proc. Natl. Acad. Sci. U.S.A. 65*, 63.

Truncated Repeated Sequences Generated by Recombination in a Specific Region

Francine C. Eden*

Laboratory of Molecular Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205
Received March 8, 1984

ABSTRACT: Structural relationships within a family of long repeated DNA sequences have been determined by molecular cloning of individual family members. About half of the family members are truncated at one end. There is a short, tandemly repeating region flanked by direct repeats associated with truncation. Recombination in a region near the tandemly repeating segment has apparently generated much of the diversity in this family.

For the vast majority of families of repeated DNA sequences, the structural relationships among family members have not been determined beyond the basic fact that they can base pair with reasonable fidelity. Through molecular cloning, these relationships are beginning to be explored. In some instances, the data go beyond description of a particular family to suggest mechanisms whereby families originate, incur variability, and are dispersed in the genome [reviewed by Brutlag (1980), Jelinek & Schmid (1982), and Singer (1982)].

In rodent genomes, there are families of long repeated DNA sequences some of whose members are truncated at the 5′ end (Fanning, 1983; Gebhard et al., 1982; Wilson of Storb, 1983). The presence of a poly(A) stretch at the 3′ boundary of these elements suggests a mechanism whereby truncated variants could be generated. The steps would be (1) RNA transcription beginning or ending internal to the element, (2) polyadenylation, (3) reverse transcription, and (4) reintegration of the shortened DNA copy. Mechanisms involving RNA intermediates may also be responsible for the structure and dispersal of short repeated sequence families like the Alu family and for the generation of processed genes and pseudogenes (Jagadeeswaran et al., 1981; Van Arsdell et al., 1981; Sharp, 1983). Thus, the repeated component of the genome seems to incur variability by some of the same mechanisms that generate change within structural genes.

Here a family of chicken repeated DNA sequences is described where some members are apparently shortened at one end. This family, which has been studied previously by molecular cloning of individual family members and by genomic blotting (Eden et al., 1980, 1981; Musti et al., 1981), has about 500 genomic copies, each several kilobases in length. In the chicken genome, large repetitive regions occur where an individual member of this family is clustered with other long repeated sequences. The family includes structural variants distinguishable by restriction analysis. It has now been determined how some of this variability arose. A short, tandemly repeating region has been discovered within some family members. It is apparently associated with a specific region of recombination responsible for truncation of some family members at one end.

* Correspondence should be addressed to this author at Meloy Laboratories, 6715 Electronic Drive, Springfield, VA 22151.

Experimental Procedures

Cloned DNA Segments. A 3.6-kilobase (kb)[1] *Eco*RI fragment of chicken DNA cloned in the plasmid pBR322 represents one member of a family of long repeated sequences (Eden et al., 1980) and was used as a source of probes for all of this work. The segments derived from it (Figure 1) were isolated either by separation of restriction fragments in agarose or acrylamide gels (probes I, IV, V, and VI) or by recloning in pBR322 (probes II and III). Recombinant phages containing sequences homologous to the 3.6-kb fragment were selected by screening a library of chicken DNA prepared and provided by D. Engel. It was constructed by the method of Maniatis et al. (1978). Positive phases were plaque purified twice prior to amplification and DNA isolation. Phage and plasmid DNAs were prepared as described previously (Eden et al., 1980).

DNA Labeling and Sequencing. DNA used in hybridization experiments was labeled with ^{32}P by nick translation (Rigby et al., 1977). End labeling of DNA was performed by addition of [γ-^{32}P]ATP to dephosphorylated termini using T4 polynucleotide kinase. DNA sequencing was performed according

[1] Abbreviations: SSC, 0.15 M NaCl and 0.015 M trisodium citrate; SDS, sodium dodecyl sulfate; kb, kilobase; bp, base pair; EDTA, ethylenediaminetetraacetic acid.

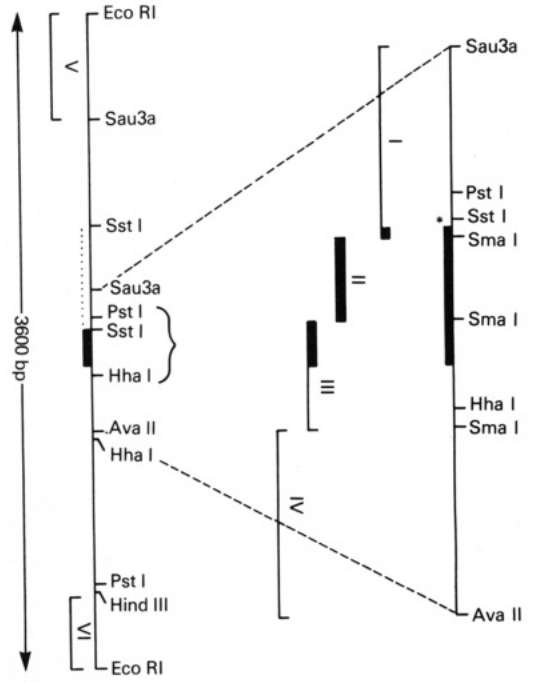

FIGURE 1: Restriction map of the 3.6-kb fragment. A cloned *Eco*RI fragment of chicken DNA is shown together with its restriction sites and probes (I–IV) generated from it. The location of an internal tandemly repeating region (dark box) is shown. The bracket indicates the *Pst*I/*Hha*I fragment described in the text. The important 560-bp *Sst*I fragment (dotted underline) and the specific recombination site (asterisk) are also indicated.

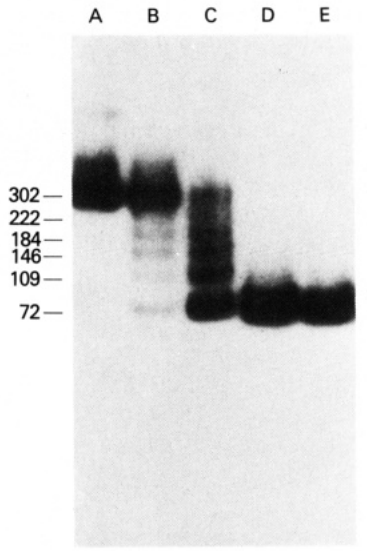

FIGURE 2: Tandemly repeating sequence lies within the 3.6-kb fragment. To prepare the internal 300-bp *Pst*I/*Hha*I fragment labeled only at the *Pst*I site (Figure 1, bracket), the 1.5-kb *Pst*I fragment was isolated, phosphatased, end labeled with ^{32}P, and recleaved with *Hha*I. After purification through an acrylamide gel, the 300-bp fragment was mixed with unlabeled carrier DNA and digested with *Hpa*II using 0.5 unit of enzyme per μg of DNA. Aliquots were removed at intervals, and digestion was stopped by additon of 0.1% SDS and 0.02 M EDTA. The partial digestion products (lanes B–D) were resolved in a 10% polyacrylamide gel together with undigested (lane A) and fully digested (lane E) samples. DNA fragments were detected by autoradiography. Fragment lengths, in base pairs, are indicated at the left and were derived by comparison with labeled DNA fragments of known length (not shown) separated in the same gel.

to Maxam & Gilbert (1980) except that the G + A reaction was in 60% formic acid. The full-length member of the repeated family was sequenced as follows (refer to Figure 3): rightward from *Sau*3a to *Pst*I, from *Pst*I to *Sma*I, and from *Sma*I to *Hha*I; leftward from *Sma*I to *Hha*I, from *Hha*I to *Sst*I, and from *Sma*I to *Ava*II. The homologous region of the truncated family member was sequenced rightward from *Sau*3a to *Sst*I and also from an internal *Hae*III site through the *Sst*I site and beyond.

Restriction Endonuclease Digestion, Gel Electrophoresis, Blotting, and Hybridization. Restriction endonuclease digestions were under conditions specified by the vendor using 2 units/μg of DNA. Gel electrophoresis was performed as described previously (Eden et al., 1980). Blotting was according to Southern (1975). Hybridization was performed in 6 × SSC, 1 × Denhardt's solution (Denhardt, 1966), and 0.1% SDS at 68 °C for 15–20 h. Filters were washed repeatedly in 0.1 × SSC at 52 °C, dried, and autoradiographed using Kodak X-O-Mat AR film and intensifying screens at −70 °C.

RESULTS

Short, Tandemly Repeating Region Internal to a Long Repeat Element. A cloned, 3.6-kb *Eco*RI fragment that is part of a long repeated sequence in the chicken genome (Eden et al., 1980) is shown in Figure 1. A short, tandemly repeating sequence has now been found within the same 3.6-kb fragment (Figure 1, dark box). The tandemly repeating region can be detected by partial digestion of a 300-bp *Pst*I/*Hha*I fragment (Figure 1, bracket) with the restriction endonuclease *Hpa*II. Figure 2 shows the products generated when this fragment is end-labeled at the *Pst*I site. The first *Hpa*II cleavage site is located 72 bp from the labeled end and is followed by a series of sites spaced at intervals of about 38 base pairs. At least four 38 base pair units can be visualized on this gel.

DNA sequencing has confirmed the presence of a tandemly repeating region and determined its structure. The sequence of a 500 base pair segment containing the tandem repeat and a schematic diagram of the region are shown in Figure 3. Five tandem repetitions of a 38-bp monomer unit occur. Each unit contains a single cleavage site for *Hpa*II (Figure 3; positions 232, 269, 306, 344, and 382). The basic unit is strongly conserved, with only four variable positions. The length of the basic unit varies slightly due to single base insertion or deletion, and the 3′-most unit is missing the last three nucleotides. The five 38-bp units are in turn flanked by imperfect, 12-bp direct repeats (asterisks, Figure 3) which differ in two positions.

Tandemly Repeating Region Is Absent from Some Family Members. The association of this tandemly repeating sequence with the longer repeat element has been further explored by isolation of additional family members from a chicken DNA library in bacteriophage λ. The entire 3.6-kb fragment was used as a probe. Each of these additional family members was then tested for hybridization to the tandemly repeating region or to its 5′- or 3′-flanking DNA segments (probes I–IV, Figure 1). Figure 4 shows that while most members hybridize to the 5′-flanking segment (probe I), many failed to hybridize to the tandem repeat probes (II and III) and to the 3′-flanking DNA (probe IV). No family members were found that were missing the tandem repeats only (positive with probes I, III, and IV; negative with probe II) or were missing the 5′-flanking region only (positive with probes II, III, and IV; negative with probe I). Thus, members of this repeated family are principally of two types. Either they have the tandemly repeating region together with its 5′- and 3′-flanking DNA, or they are missing the tandem repeats and 3′-flanking DNA.

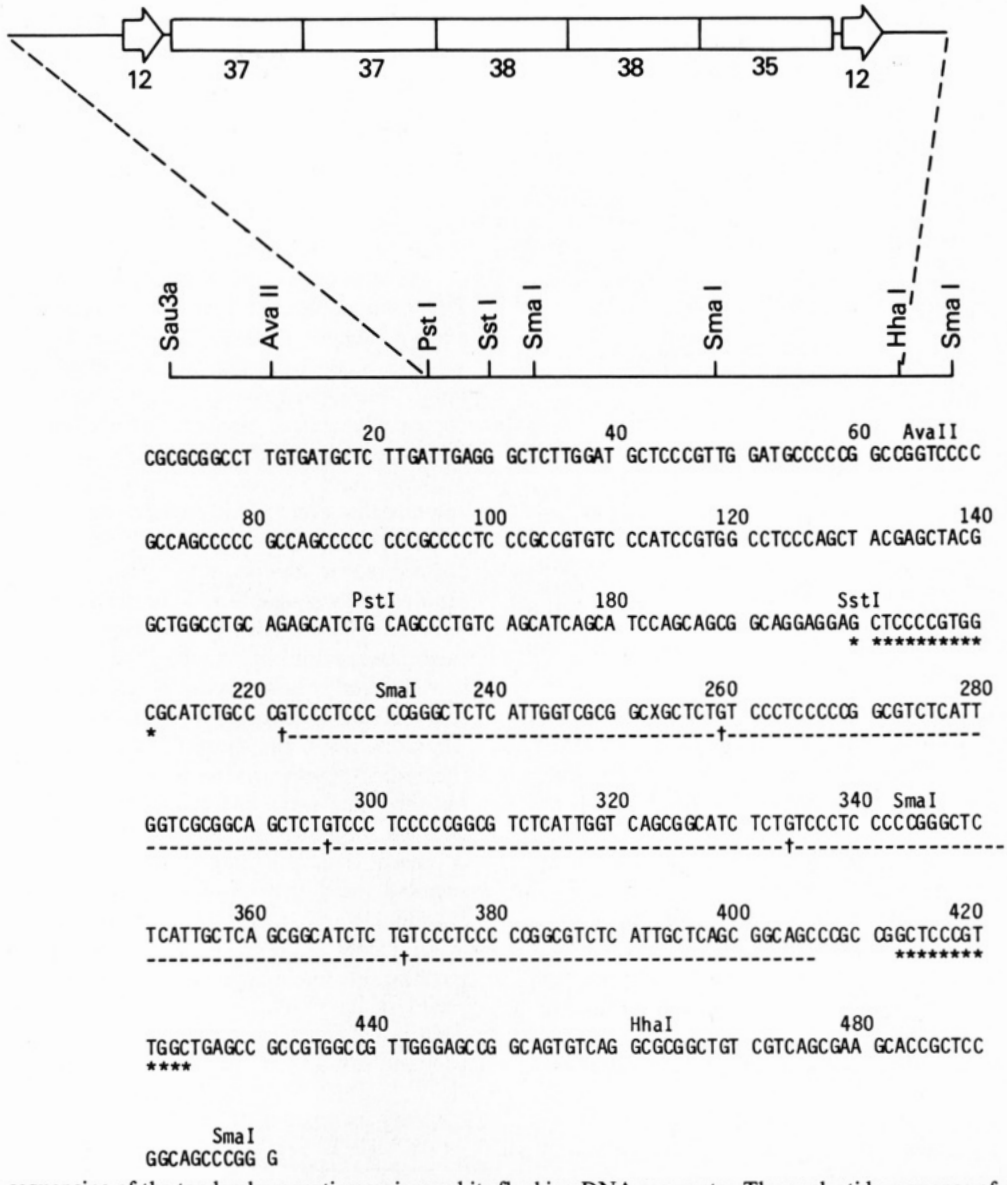

FIGURE 3: DNA sequencing of the tandemly repeating region and its flanking DNA segments. The nucleotide sequence of a 500-bp segment containing the tandemly repeating region is shown. It begins at position 222 and extends to position 406 and is composed of five units (†) approximately 38 bp in length. Direct repeats 12 bp in length (*) flank the tandemly repeating region. A schematic representation of this region is shown at the top.

These data suggest that some of the variation in this family is mediated by recombination near the 5′ end of the tandemly repeating region. Thus, it would be expected that in the chicken genome the distal ends of the long repeat element would lose their association with some frequency. To test this, a library of chicken DNA in bacteriophage λ, where about 1% of the plaques are positive with the 3.6-kb probe, was plated to form separated plaques. Duplicate plaque lifts were hybridized to probes V and VI (Figure 1). Only 43% of the 158 positive plaques hybridized to both probes (data not shown). Thus, the pattern of variability established by using cloned family members seems to be representative of the family as a whole.

Localization of the Recombination Region. Restriction analysis of the cloned family members has been used to precisely locate the recombination region associated with truncation. The important area lies within a 560-bp *Sst*I fragment, indicated by the dotted line in Figure 1, which hybridizes probe I. If recombination was within this fragment, its length would differ in family members of the two types. Figure 5 shows that this is not the case, as family members of both types have

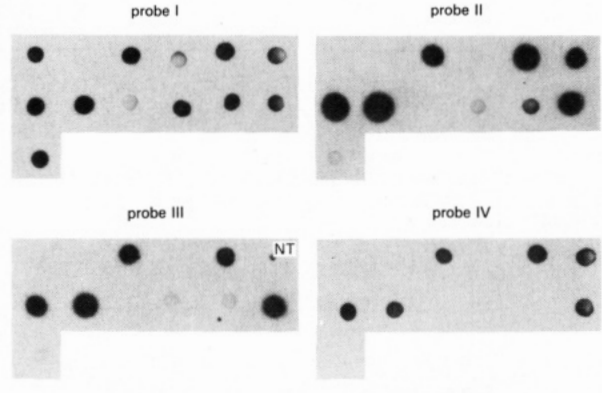

FIGURE 4: Analysis of cloned family members with specific probes. DNA was isolated from many recombinant phages, each containing a different member of the repeated DNA family. The DNAs were denatured and spotted onto nitrocellulose in a patterned array. Duplicate filters were prepared and hybridized to different probes. Probe I, 5′-flanking DNA; probe II, tandemly repeating region; probe III, tandemly repeating region plus 3′-flanking DNA; probe IV, 3′-flanking DNA. NT = not tested.

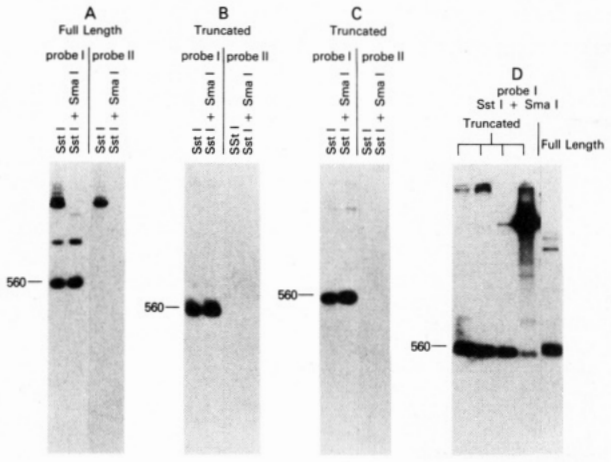

FIGURE 5: Restriction analysis localizes the recombination region in the truncated variants. (Panels A–C) a full-length and two truncated variants of the repeated sequence family were digested with restriction endonucleases. Fragments were separated in 1.8% agarose gels, transferred to nitrocellulose, and hybridized to the indicated probes. Panel D compares four truncated and a full-length family member on the same gel. The diagnostic 560-bp *Sst*I fragment is indicated in each panel.

a 560-bp *Sst*I fragment positive with probe I (panels A–C). Hybridization analysis with probe II shows that the full-length variant (A) has an additional positive *Sst*I fragment which contains the tandem repeats, is cleaved by *Sma*I, and is not present in truncated variants (panels B and C). In panel D, truncated and full–length members are compared on the same gel, establishing further that the 560-bp *Sst*I fragment is not detectably altered in length in any of the truncated variants. Thus, the recombination locale (asterisk, Figure 1) seems to be the same in all of the truncated members, lying between the *Sst*I site and the beginning of the tandem repeats. This region is about 20 bp in length.

The DNA sequence of one of the truncated family members confirms this. Figure 6 compares DNA sequences of corresponding segments of a full-length and a truncated family member in the important region. Homology ends abruptly at the 5′-most direct repeat at the boundary of the tandemly repeating region. A schematic representation of the two types of sequences is included in the figure.

Discussion

One of the more striking features of the family of repeated DNA sequences described here is its limited structural variability. Although there are hundreds of copies each in a separate genomic location, genomic blotting experiments detect only a few restriction fragments, each present in multiple copies. The data presented here confirm that there are only a few specific variants within this family and indicate that most of these could be derived from one another by a specific recombination event. The recombination apparently occurred within a short DNA region. Within the limits of the data, it is clear that it does not vary by more than 50 bp in different family members, and it could be precise to the nucleotide. The presence of the tandemly repeating region may in some way facilitate recombination, although it does not seem to be involved directly because recombination apparently occurred just to the 5′ side of the tandem repeats instead of internal to them. The presence of the short direct repeats would suggest that the tandem repeats themselves constitute an insertion element, but the relationship of the full-length and truncated variants does not support this. When the tandemly repeating element is present, it is associated with a specific 3′-flanking DNA segment which would have to be part of the insertion but is outside the direct repeats. If truncation did occur at the DNA level, it does not seem to be related to commonly accepted mechanisms like unequal crossover in a tandemly repeating region or insertion.

In other repeat families, truncation has been related to propagation and dispersal through RNA intermediates.

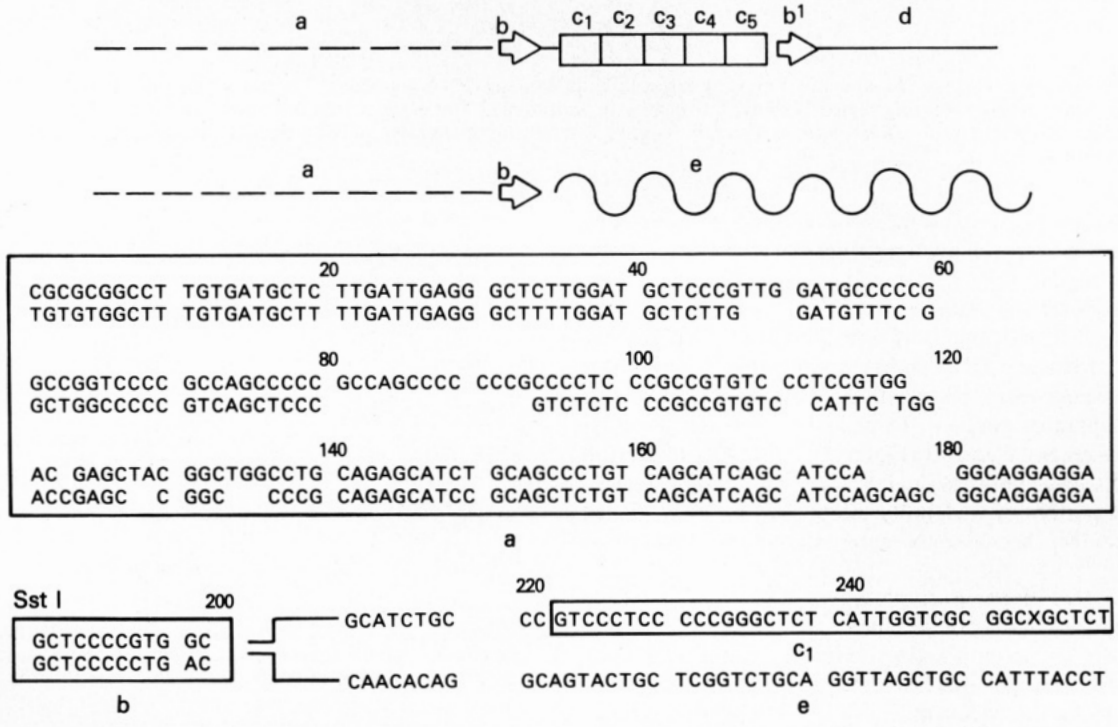

FIGURE 6: Sequencing of a recombination site. Corresponding regions of a full-length (upper) and a truncated (lower) family member are compared. The sequence homology in region "a" is evident. Within the direct repeat, "b" homology ends, with the tandemly repeating region c_1, c_2, etc. and segment "d" replaced by segment "e".

Specific truncated variants would result from strong transcription initiation or termination sites internal to the repeat element. In this context, it is interesting to compare the internal structure and pattern of transcription of the long repeat element *Copia* in the *Drosophila* genome (Fouts & Manning, 1981) with the repeated family described here. There are two principal RNA transcripts from *Copia*, one full length and the other terminating near the center of the element within a tandemly repeating region strikingly similar to the one described here; it too is composed of a few repetitions of a 35-37-bp element. Thus, the truncated variants described here might derive not from recombination at the DNA level but by specific termination of RNA transcripts followed by reinsertion of the corresponding DNA copies.

In some measure, the data presented here serve to explain the limited number of structural variants detected in this, and perhaps in other, long repeated sequence family. There are apparently internal structural features that facilitate the appearance of certain variants. Like "hot spots" for mutation or recombination, these special regions may cause certain specific variants to predominate against a background of random variants generated by other mechanisms.

ACKNOWLEDGMENTS

I thank Raymund Cuevo for help in isolation and analysis of recombinant phages and Tom McCutchan and Dinah Singer for many constructive suggestions during preparation of the manuscript.

REFERENCES

Brutlag, D. L. (1980) *Annu. Rev. Genet. 14*, 121-144.
Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun. 23*, 641-646.
Eden, F. C. Burns, A. T. H., & Goldberger, R. F. (1980) *J. Biol. Chem. 255*, 4843-4853.
Eden, F. C., Musti, A. M., & Sobieski, D. A. (1981) *J. Mol. Biol. 148*, 129-151.
Fanning, T. G. (1983) *Nucleic Acids Res. 11*, 5073-5091.
Fouts, D. L., & Manning, J. E. (1981) *Nucleic Acids Res. 9*, 7053-7064.
Gebhard, W., Meitinger, T., Hochtl, J., & Zachau, H. G. (1982) *J. Mol. Biol. 157*, 453-471.
Jagadeeswaran, P., Forget, B. G., & Weissman, S. M. (1981) *Cell (Cambridge, Mass.) 26*, 141-142.
Jelinek, W. R., & Schmid, C. W. (1982) *Annu. Rev. Biochem. 51*, 813-844.
Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., & Efstratiadis, A. (1978) *Cell (Cambridge, Mass.) 15*, 687-701.
Maxam, A. M., & Gilbert, W. (1980) in *Methods in Enzymology* (Grossman, L., & Moldave, K., Eds.) pp 499-560, Academic Press, New York.
Musti, A. M., Sobieski, D. A., & Eden, F. C. (1981) *Biochemistry 20*, 2989-2988.
Rigby, P. W. J., Dickmann, M., Rhode, S. C., & Berg, P. (1977) *J. Mol. Biol. 113*, 236-251.
Sharp, P. A. (1983) *Nature (London) 301*, 471-472.
Singer, M. F. (1982) *Int. Rev. Cytol. 76*, 67-112.
Southern, E. M. (1975) *J. Mol. Biol. 94*, 51-69.
Van Arsdell, S. W., Denison, R. A., Bernstein, L. B., Weiner, A. M., Manser, T., & Gesteland, R. F. (1981) *Cell (Cambridge, Mass.) 26*, 11-17.
Wilson, R., & Storb, U. (1983) *Nucleic Acids Res. 6*, 1803-1807.

Formation of Radiation-Induced Cross-Links between Thymine and Tyrosine: Possible Model for Cross-Linking of DNA and Proteins by Ionizing Radiation†

Michael G. Simic and Miral Dizdaroglu*

Radiation Physics Division, Center for Radiation Research, National Bureau of Standards, Gaithersburg, Maryland 20899
Received May 8, 1984

ABSTRACT: A model for radiation-induced cross-linking of DNA and proteins has been developed. It is based on initial formation of free radicals on a DNA base, i.e., thymine, and on an amino acid, i.e., tyrosine. It was demonstrated that interaction of these radicals is highly favored as measured by their kinetics and the cross-linked products. The gas chromatography–mass spectrometry methodology used for the identification of the thymine–tyrosine cross-links is suggested as an experimental approach in the measurements of biological cross-links.

Cross-links within DNA and between DNA and proteins induced by different agents and processes, such as drugs (Kohn, 1977; Kohn & Ewig, 1979), autoxidation (Riess & Tappel, 1973), and ionizing radiations (Fornace & Little, 1977; Mee & Adelstein, 1981; Bowden et al., 1982), and intrastrand links caused by UV light (Smith, 1976) are becoming recognized as a common occurrence in vitro and in vivo. However, the cross-linking has not received a deserved attention in the overall consideration of DNA damage and its consequences, although cross-linking is by no means a minor process. In most cases, the kinetics and mechanisms of cross-linking are poorly understood, whereas UV light-induced intrastrand links are well-known to take place mainly via condensation of two double bonds, i.e., the formation of cyclobutane-type dimers (Setlow, 1966). Similarly, UV light causes interstrand

†This work was conducted pursuant to a contract with the National Foundation for Cancer Research, Bethesda, MD.

0006-2960/85/0424-0233$01.50/0