Stewart, J. M., & Young, J. D. (1969) *Solid Phase Peptide Synthesis*, p 57, W. H. Freeman, San Francisco.
Tam, J. P., Kent, S. B. H., Wong, T. W., & Merrifield, R. B. (1979) *Synthesis*, 955.
Wang, S.-S., Kulesha, I. D., & Winter, D. P. (1979) *J. Am. Chem. Soc. 101*, 253.
Wang, S.-S., Makofske, R., Bach, A. E., & Merrifield, R. B. (1980) *Int. J. Pept. Protein Res. 15*, 1.

Histone 2A, a Heteromorphous Family of Eight Protein Species†

Michael H. P. West and William M. Bonner*

ABSTRACT: The histone 2A family of proteins is shown to consist of eight protein species. In addition to the previously described mammalian 2A variants H2A.1 and H2A.2, we describe two variants which are separable from each other and from variants 1 and 2 on both sodium dodecyl sulfate and acetic acid–urea gels. These two proteins H2A.X and H2A.Z are termed heteromorphous variants to distinguish them from the predominating form and its homeomorphous variants which require nonionic detergents for their resolution. The two heteromorphous variants are present in nucleosomal core particles isolated from mouse L1210 cells. In addition, these variants are found in normal mouse tissues, human HeLa cells, and chicken erythrocytes. On sodium dodecyl sulfate gels, one variant, H2A.X, has an apparent molecular weight ~1000 larger than H2A.1 and comprises ~11% of the total 2A in mouse L1210 cells. The second variant, H2A.Z, has an apparent molecular weight ~600 smaller than H2A.1 and comprises ~4% of the total 2A in mouse L1210 cells. The two heteromorphous variants have the same arginine/lysine ratio as H2A.1. In addition, a fraction of each of the four variants (~11% in L1210 cells) is combined with ubiquitin. The molar sum of these eight H2A species approximately equals the number of moles of H4, H2B, or H3 in chromatin.

The remarkable progress in elucidating chromatin structure and function over the past few years has implicated histones in the primary condensation of deoxyribonucleic acid (DNA) into a nucleosomal structure (Felsenfeld, 1978). This more recent appraisal of the structural role of histones contrasts with earlier notions of histones in gene regulatory roles. However, histones are known to undergo many modifications including acetylation (Dixon et al., 1975), phosphorylation (Dixon et al., 1975), methylation (DeLange et al., 1969, 1973), condensation with ubiquitin (Goldknopf et al., 1978), and poly(adenosine 5'diphosphate) ribosylation (Wong et al., 1977; Giri et al., 1978). These many modifications presumably reflect alterations in chromatin which relate to one or more of its various functions.

Another level of complexity involves the sequences of histones. In vertebrates, histones 3, 2B, and 2A can have nonallelic variants which are related by simple amino acid substitutions (Franklin & Zweidler, 1977). Newrock et al. (1977) have shown that patterns of nonallelic variants may change during embryogenesis. A second type of variation, which includes changes in length due to insertions and deletions, has been reported (von Holt et al., 1979) in plants and invertebrates (Rodrigues et al., 1979). These latter heteromorphous variations are separable on acetic acid–urea or sodium dodecyl sulfate ($NaDodSO_4$) gels.

In some cases heteromorphous variants have been reported in the same organism. Wheat embryo contains heteromorphous H2A variants (Rodrigues et al., 1979), and sea urchin sperm contains heteromorphous H2B variants (Strickland, M., et al., 1977, 1979; Strickland, W. N., et al., 1977). In this paper we report the existence of two heteromorphous variants of H2A in mouse L1210 cells. These two variants, H2A.X and H2A.Z, contain peptides which comigrate with peptides from H2A.1 and H2A.2. In addition, H2A.X and H2A.Z have the same arginine/lysine ratio as variants 1 and 2 and have the same proportion bound to ubiquitin[1] as do variants 1 and 2. H2A.X and Z are also found in HeLa cells and chicken erythrocytes, findings which suggest that the heteromorphous nature of the H2A family may be a general phenomenon.

Materials and Methods

N^{α}-Tosylmethylalanine chloromethyl ketone (TPCK)–trypsin and micrococcal nuclease were obtained from Worthington Biochemicals. Sodium dodecyl sulfate was from BDH chemicals. Agarose (Type 1:low EEO), cysteamine hydrochloride (2-mercaptoethanolamine), Triton N-101, Triton X-100, and hexadecyltrimethylammonium bromide (CTAB) were from Sigma. ^3H- or ^{14}C-labeled amino acids were from New England Nuclear or ICN Pharmaceuticals; $^{32}PO_4$ was from New England Nuclear.

Cell Culture. Mouse L1210 leukemia cells were maintained in Roswell Park Memorial Institute (RPMI) Medium 1630 supplemented with 20% fetal calf serum without antibiotics as a static culture. When required, cells were diluted to 2×10^5 cells/mL in the presence of penicillin and streptomycin and were harvested at approximately 1×10^6 cells/mL after 18–24 h.

Labeling. For fingerprinting, cells were labeled with [^{14}C]arginine at 5 μCi/mL in RPMI 1630 prepared with arginine at 14% of the normal level and supplemented with 20% fetal calf serum. Cells were labeled with [^{35}S]methionine,

† From the Laboratory of Molecular Pharmacology, Developmental Therapeutics Program, Division of Cancer Treatment, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205. *Received November 30, 1979.* M.H.P.W. is a Fogarty Fellow.

[1] The ubiquitin–H2A complex was discovered and named A_{24} by Goldknopf et al. (1975) and Olson et al. (1976). In this paper we use the prefix u to denote the ubiquitin (Cook et al., 1979) adduct; hence u2A.1, u2A.2, u2A.X, and u2A.Z are the ubiquitin adducts of 2A.1, 2A.2, 2A.X, and 2A.Z, respectively. Albright et al. (1979) have also reported that the A_{24} of Goldknopf & Busch (1977) is a mixture of u2A.1 and u2A.2.

[^{35}S]cysteine, [^{14}C]tryptophan, or [^{14}C]phenylalanine in complete medium.

Cells at approximately 10^6 cells/mL were centrifuged and washed with and suspended in 0.15 M NaCl containing 20 mM Hepes, pH 7.6, glucose (2.5 g/L), $NaHCO_3$ (0.35 g/L), and phenol red as an indicator to label cells with $^{32}PO_4$. The cell density was maintained at 10^7 cells/mL, and $^{32}PO_4$ was added at either 50 or 100 μCi/mL. Incubation at 37 °C was continued, and samples were removed into ice to stop further incorporation.

Nuclear Isolation. L1210 nuclei were isolated at 4 °C by a detergent method (Thomas et al., 1978). Pellets containing 10^8 cells were washed in 10 mL of buffer A (1 mM potassium phosphate, 1 mM $CaCl_2$, 10 mM dithiothreitol (DTT), and 0.32 M sucrose, pH 6.7) and centrifuged for 5 min at 170*g*. Cells were suspended in buffer A at 10^8 cells/mL and were lysed by the addition of buffer A containing 0.3% Triton N-101 at a final cell concentration of 10^7 cells/mL. Nuclei were centrifuged for 5 min at 675*g* and washed once with buffer A containing Triton.

Acid Extraction. Nuclei were washed once with buffer A and suspended in that solution at 10^8 cells/mL. HCl was added to a final concentration of 0.4 N, and samples were kept at 4 °C for 30–60 min. Samples were spun in a microfuge and the supernatants removed. Urea was added to 8 M, followed by phenolphthalein (0.002%), and the HCl was neutralized by adding NH_4OH until the solution turned pink. DTT was added to a final concentration of 50 mM, and the samples were incubated at room temperature for 5 min, after which time the samples were made 1 M in acetic acid and were loaded onto gels for electrophoresis. In some experiments, 1 mM phenylmethanesulfonyl fluoride (PMSF) was included in all buffers used for the preparation and extraction of nuclei, but no change in the histone pattern was apparent. L1210 nuclei can be incubated for 1 h at 37 °C in buffer A without PMSF with no detectable histone degradation.

Salt Treatment. Nuclei were washed once with buffer A and suspended in buffer A at 10^8 cells/mL. They were diluted to 10^7 cells/mL by the addition of 0.7 M NaCl, 1 mM EDTA, and 2.5 mM potassium phosphate, pH 6.9. This chromatin suspension (5 mL) was layered onto 34 mL of 0.7 M NaCl, 1 mM EDTA, and 2.5 mM potassium phosphate, pH 6.9, with 15% sucrose. The preparation was centrifuged at 4 °C for 3 h at 82000*g* in a swing-out rotor. The pellet was suspended in 0.5 mL of distilled water prior to acid extraction or in 0.5 mL of 0.2 M sucrose and 5 mM Tris-HCl, pH 7.5, prior to nuclease digestion.

Gel Electrophoresis. First-dimension acetic acid–urea–Triton X-100 (AUT) gels were prepared as described by Bonner et al. (1980). The resolving gel contained 15% acrylamide (Eastman Chemical Co., used after filtering), 0.1% *N*,*N*′-methylenebis(acrylamide), 1 M acetic acid, 0.5% *N*,*N*,*N*′,*N*′-tetramethylenediamine, 0.05 M NH_4OH, 8 M urea, 0.5% (8 mM) Triton X-100, and 0.00027% riboflavin and was polymerized by light.

The stacking gel was prepared in the same way as the resolving gel except that the acrylamide was changed to 3.3% and the *N*,*N*′-methylenebis(acrylamide) to 0.16% and the Triton X-100 was eliminated. The upper and lower reservoir buffers were 1 M acetic acid and 0.1 M glycine.

Second-dimension acetic acid–urea–CTAB (AUC) gels were prepared in the same way as the AUT gels with the following exceptions: the urea in both the resolving and stacking gel was lowered to 6 M; the Triton X-100 was eliminated; 0.15% CTAB was added to the upper reservoir buffer. The samples for AUC gels were pieces of other gels which had been stained with 0.1% Coomassie Brilliant Blue R in 40% ethanol and 5% acetic acid containing 0.1% cysteamine hydrochloride and destained in 20% ethanol and 5% acetic acid containing 0.1% cysteamine hydrochloride. The gel pieces were soaked for at least 1 h in 1 M acetic acid–50 mM NH_4OH–1% cysteamine hydrochloride before loading.

Chromatin fragments were separated in a 5% acrylamide gel as described by Varshavsky et al. (1976). The gel was stained with ethidium bromide in the Tris–EDTA gel buffer. Excised nucleosome bands or whole tracks were loaded on top of a discontinuous AUT gel and embedded in 1% agarose–1 M acetic acid–50 mM NH_4OH–0.1% cysteamine hydrochloride. One milliliter of 1% protamine in 1 M acetic acid–50 mM NH_4OH–0.1% cysteamine hydrochloride was layered on top of the sample (Shaw & Richards, 1979).

Extraction of Proteins and Peptide Mapping. Proteins were isolated from gels and trypsinized according to the method of Zanetta et al. (1970). We found that tryptic peptides could be conveniently separated on modified discontinuous acetic acid gels. The resolving gel contained 50% acrylamide, 0.03% *N*,*N*′-methylenebis(acrylamide), 1 M acetic acid, 0.5% *N*,*N*,*N*′,*N*′-tetramethylenediamine, 0.05 M NH_4OH, and 0.00027% riboflavin and was polymerized by light. The stacking gel and the reservoir buffers were the same as for AUT gels. The dried trypsinized protein samples were dissolved in 15 μL of 1 M acetic acid, 0.05 M NH_4OH, and 8 M urea and contained approximately 800 cpm per labeled amino acid residue. Loading the samples to be compared in adjacent slots allowed one not only to see the pattern of each sample separately but also to determine if bands in adjacent samples comigrated. Therefore, a separate mixing experiment was not necessary. The gels were run at 5 mA overnight, immediately dried, and autoradiographed. Model experiments show that this gel system will separate peptides containing three or more amino acids. A more detailed protocol will be reported elsewhere.

Nuclease Digestion of Chromatin. Chromatin was prepared by the high-salt extraction method described above and was dialyzed overnight against 0.2 M sucrose and 5 mM Tris-HCl, pH 7.5. $CaCl_2$ was added to a final concentration of 1 mM, and the suspension was warmed at 37 °C for 30 s. Micrococcal nuclease (2000 units/mL) was added to 1 unit/mL, and digestion was for 1.5 min at 37 °C. Digestion was terminated by the addition of EDTA to 1 mM, and glycerin was added to facilitate sample loading. The sample was brought to 20 mM in Tris-HCl, pH 7.6, and 2 mM in EDTA and loaded onto a chromatin gel (Varshavsky et al., 1976). Electrophoresis was for 4 h at 250 V, and the lower reservoir buffer was changed after 2 h. The gel was stained with ethidium bromide at 3 μg/mL, and bands were visualized under UV light.

Results

Histone Separation. An acetic acid–urea–Triton X-100 (AUT) gel of the acid extract from mouse L1210 nuclei contained a great many bands (Figure 1A). Salt treatment (0.7 M NaCl) prior to acid extraction removed the vast majority of nonhistone proteins, leaving only the known histones and several minor components, two of which were labeled X and Z (Figure 1B).

When these two gels were subjected to electrophoresis in a second dimension without Triton X-100, several interesting patterns appeared. The two-dimensional gel of the acid extract prior to salt treatment shows that many proteins lie on the diagonal (Figure 2A); the core histones, however, lie below the diagonal (Bonner et al., 1980). Salt treatment (Figure

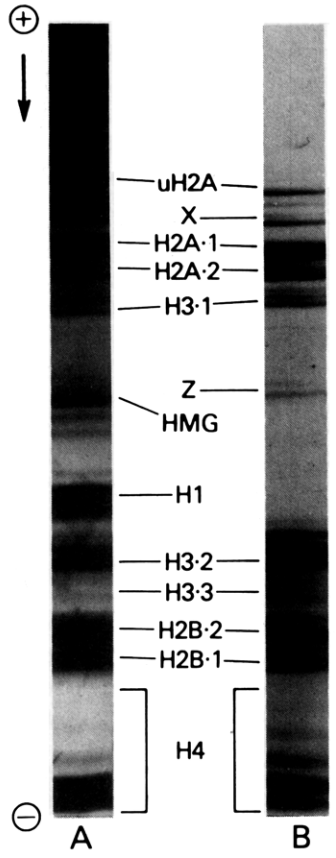

FIGURE 1: Separation of acid-extracted proteins in an acetic acid–urea gel containing Triton X-100 (AUT). Nuclei prepared from L1210 cells were HCl extracted either without (A) or after (B) treatment with 0.7 M NaCl. Migration is from top to bottom in a 15% AUT slab gel as described under Materials and Methods.

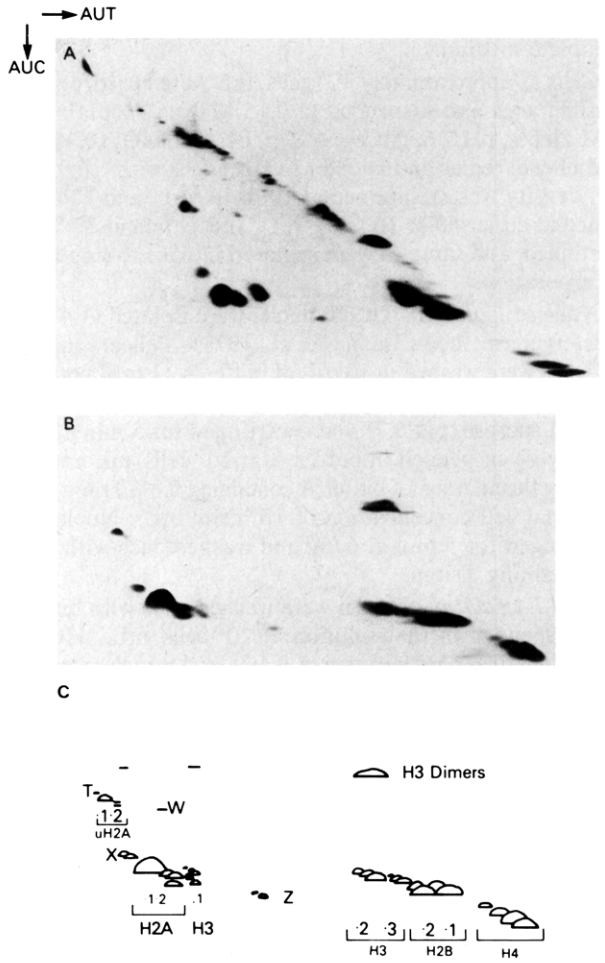

FIGURE 2: AUT–AUC two-dimensional electrophoresis of acid-extracted nuclear proteins. AUT gels similar to those shown in Figure 1 were placed on the stacking gel of a 15% AUC gel, and proteins were electrophoresed from top to bottom as described under Materials and Methods. (A) HCl extract of whole nuclei. (B) HCl extract of 0.7 M NaCl-treated nuclei. (C) Diagram of gel B. The first dimension of gel A was stained and destained in the dark with 0.1% cysteamine, while the first dimension of gel B was exposed to light and was not stained and destained with 0.1% cysteamine. As explained elsewhere (Bonner et al., 1980) these differences in conditions lead to H3 dimer formation, indistinct H3 spot morphology, and fast (photooxodized) forms of H2A.2 and uH2A.2. H2A.1 and uH2A.1, -X, -Z, -T, and -W are unaffected by these differences.

2B) removes the proteins which lie on the diagonal. In addition to the core histones, the spots marked T, W, X, and Z remain, and a repetition of pattern is noticeable between X, 2A.1, 2A.2, and Z and T, u2A.1, u2A.2, and W. The pattern repetition is also found if the second dimension is an $NaDodSO_4$ gel (Figure 3) except that 2A.1 and 2A.2 cannot be resolved. Their Triton binding and pattern repetition suggest that T, W, X, and Z may be related to H2A; therefore, we decided to investigate these eight protein species by comparative peptide mapping.

Characterization of Proteins. Since T, W, X, and Z are cleanly resolved from H2A.1, H2A.2, uH2A.1, and uH2A.2 and from each other in the gel systems used here (Figures 2 and 3), electrophoretic methods were used for their purification. For identification by peptide mapping, the proteins from [^{14}C]arginine-labeled nuclei were separated first on AUT gels. Single bands were excised and electrophoresed on AUC gels. Single bands were again excised and electrophoresed on $NaDodSO_4$ gels, where they migrated as single bands with no visible contaminating protein species.

We have analyzed the [^{14}C]arginine peptides on discontinuous acidic 50% acrylamide gels. In 1 M HOAc virtually all tryptic peptides, even those with several carboxyl groups, should have a net positive charge and therefore migrate into the gel. Figure 4 shows the [^{14}C]arginine peptide pattern of the eight proteins of interest. Putative H2A.X has seven arginine peptides comigrating with H2A.1 arginine peptides, while putative H2A.Z has two. By comparing the fingerprints of u2A.1 and u2A.2, one can pick out four ubiquitin-specific peptides; these four peptides are also present in the putative u2A.X (T) and putative u2A.Z (W). Since putative u2A.X and u2A.Z also contain X and Z peptides, respectively, this confirms the assignment of these two spots as u2A.X and u2A.Z.

To further confirm the relationship of H2A.X and H2A.Z to H2A.1, we decided to concentrate on the peptide Ala-Gly-Leu-Gln-Phe-Pro-Val-Gly-Arg (positions 21–29 in the calf sequence) which has several interesting properties. This peptide is the only arginine peptide, larger than a dipeptide, that remains unchanged in all the H2A's that have been sequenced, from calf, rat, trout, sea urchin, and wheat germ (von Holt et al., 1979). It also has the only phenylalanine in the mammalian sequence. Figure 5 shows that H2A.1, -X, and -Z have one peptide (peptide 1) containing phenylalanine and arginine. Further experiments show that peptide 1 is also labeled with leucine, proline, and valine in all three proteins but is not labeled with isoleucine, histidine, tyrosine, or lysine. Therefore, this peptide, found in all sequenced H2A's, is also found in H2A.X and H2A.Z. We conclude that these eight proteins are comprised of four H2A variants and their respective ubiquitin adducts.

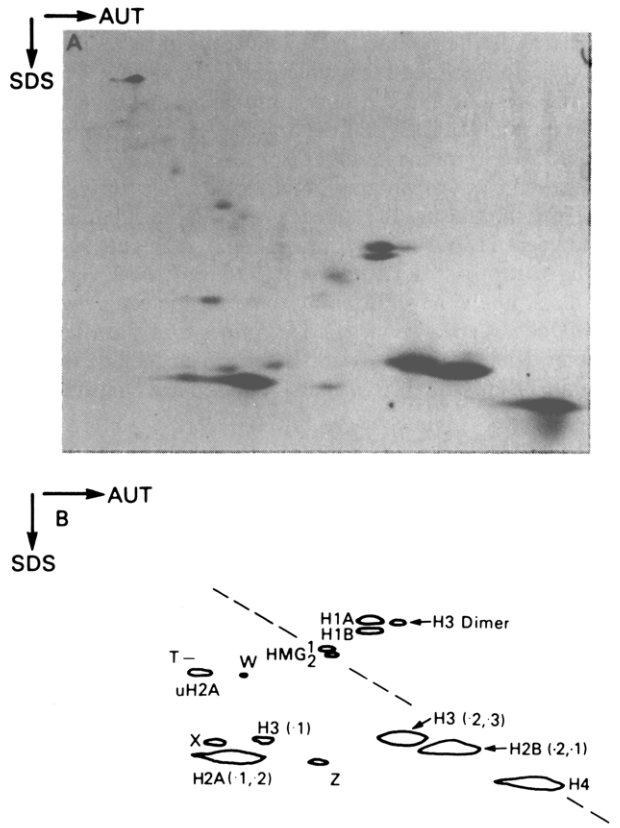

FIGURE 3: AUT–NaDodSO$_4$ two-dimensional electrophoresis of acid-extracted nuclear proteins. (A) An HCl extract of L1210 nuclei was electrophoresed first on a 15% AUT gel and then on a 15% NaDodSO$_4$ gel. (B) Diagram.

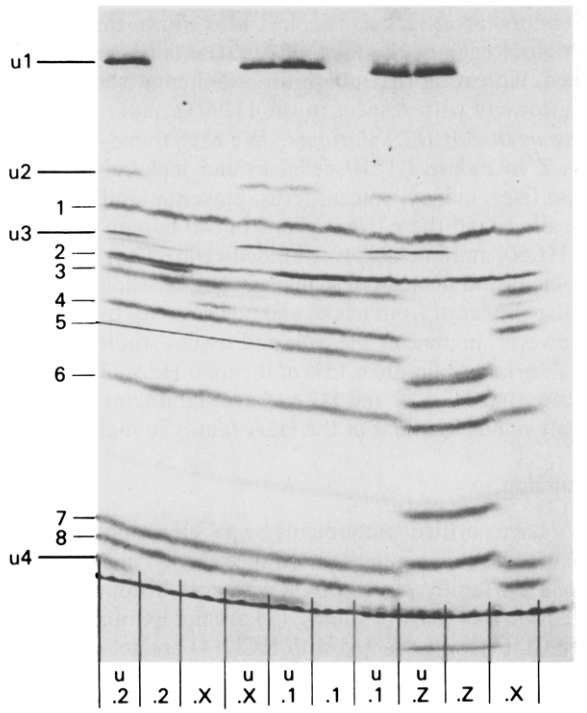

FIGURE 4: Autoradiograms of [^{14}C]arginine peptides from known H2As and putative H2As after complete trypsin digestion. Prominent H2A.1 peptides are numbered 1–8. Ubiquitin-specific peptides are labeled u1–u4.

We are currently using gels such as that in Figure 4 to make detailed comparisons of H2A.1, -2, -X, and -Z. Band 3 has been identified as AcSer-Gly-Arg on the basis that it is the only H2A arginine peptide that comigrates with an arginine

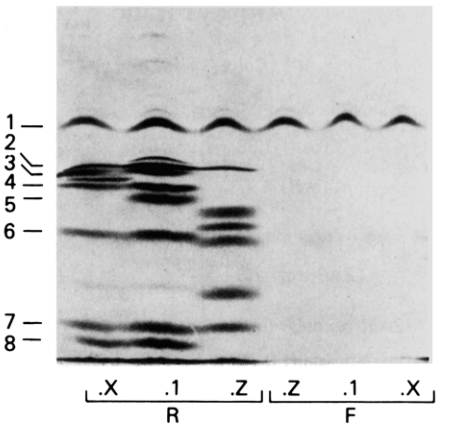

FIGURE 5: Autoradiogram of [^{14}C]phenylalanine (F) peptides in H2A.1, -X, and -Z. The [^{14}C]arginine (R) patterns are included for reference.

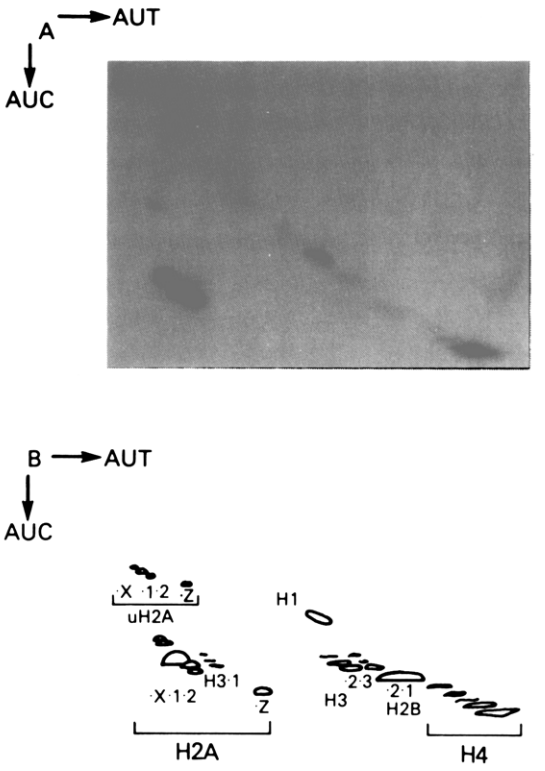

FIGURE 6: Phosphorylated histones analyzed by two-dimensional electrophoresis. Proteins labeled with $^{32}PO_4$ as described under Materials and Methods were electrophoresed in a 15% AUT gel in the first dimension and then in a 15% AUC gel in the second dimension as indicated. (A) Autoradiograph of labeled proteins. (B) Diagram of stained spots. The leading mass spots of H4 and the H2A's are not labeled with $^{32}PO_4$. Z is not labeled with $^{32}PO_4$.

peptide from H4, which also begins with AcSer-Gly-Arg, and that it is depleted in modified H2A (data not presented) as one would expect since this serine is also the site for H2A phosphorylation (Dixon et al., 1975). The results in Figure 4 suggest that H2A.X does contain this N-terminal peptide, but H2A.Z lacks it.

To investigate this finding further, we studied the phosphate incorporation into the H2A's (Figure 6). H2A.X incorporates phosphate, but H2A.Z does not. Presumably, the modified forms of H2A.Z are derived by acetylation. Therefore, the ability of the H2A's to incorporate phosphate correlates with the presence of the AcSer-Gly-Arg peptide.

In Figure 4 several differences in the arginine peptide pattern between H2A.1 and H2A.Z indicate that H2A.Z

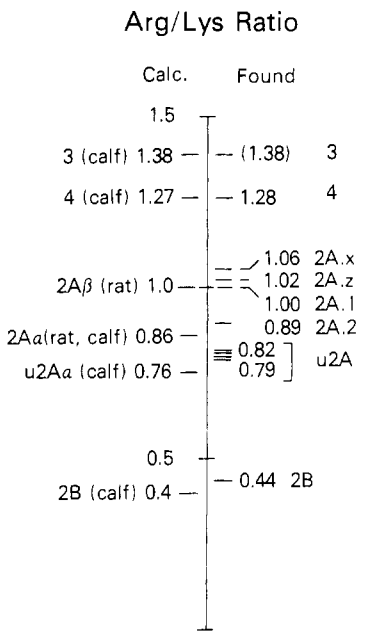

FIGURE 7: Arginine/lysine ratios of the H2As. Stained spots corresponding to each protein were excised from an AUT–AUC two-dimensional gel of L1210 nuclei labeled with [^{3}H]lysine and [^{14}C]-arginine and dried into filter paper. The dried gel pieces were digested overnight in 1.5 mL of a mixture of 95 parts of 30% H_2O_2 and 5 parts of concentrated NH_3. 10 mL of Aquasol was added for scintillation counting. The $^{14}C/^3H$ ratios were normalized to the arginine/lysine ratio of calf H3. H2Aβ has arginine at position 99; H2Aα has lysine at position 99 (Laine et al., 1976).

differs from H2A.1 by multiple internal sites and is therefore not a proteolytic product of H2A.1. The arginine peptide pattern of H2A.X is very similar to H2A.1 with only one difference seen in Figure 4. These comparisons suggest that there is much more divergence between the sequences of H2A.1 and H2A.Z than between those of H2A.1 and H2A.X.

H2A.X and H2A.Z are not labeled with [^{14}C]tryptophan, [^{35}S]cysteine, or [^{35}S]methionine. Their ubiquitin forms are, however, labeled with [^{35}S]methionine which is known to be the N-terminal amino acid residue of ubiquitin (Goldknopf & Busch, 1977). H2A.X and H2A.Z are basic proteins like H2A.1. In nonequilibrium pH-gradient electrophoresis (O'-Farrell et al., 1977) they migrate with H2A.1 and the other histones (data not presented). From its relative migration on $NaDodSO_4$ gels, H2A.X appears to be approximately 1000 daltons larger than H2A.1. By the same criterion, H2A.Z appears to be 600 daltons smaller than H2A.1.

Arginine/Lysine Ratios. The arginine/lysine ratios of H2A.X and H2A.Z substantiate their identification as H2A variants (Figure 7). From sequence data, the arginine/lysine ratio of calf H2A.1 is 0.86, while calf H2B, H3, and H4 have ratios of 0.40, 1.38, and 1.27, respectively. When the $^{14}C/^3H$ ratios for the various histones were normalized to the known arginine/lysine ratio of H3, the ratios of all the H2A variants were found to be between 0.89 and 1.06.

Variant 2 gave a ratio of 0.89, very close to the 0.86 calculated from the calf sequence. However, variant 1 gave a ratio of 1.0 somewhat out of proportion with the other histones. Mouse H2A has not been sequenced, but the sequence of rat chloroleukemia H2A is heterogeneous in position 99, H2Aβ (Arg-99) accounting for 40% and H2Aα (Lys-99) for 60% (Laine et al., 1976). The Arg-99 variant H2A has an arginine/lysine ratio of 1.0.

Amino acid analysis indicates that a similar situation may exist in mouse H2A (Blankstein & Levy, 1976; Blankstein et al., 1977). In parallel analyses, variant 1 from mouse testes had 1.0 less lysine residue and 0.7 more arginine residue than variant 2. In the Friend leukemia line C7D, variant 1 had 0.6 less lysine residue and 0.7 more arginine residue than variant 2. These results suggest that the arginine–lysine polymorphism demonstrated in rat may exist in mouse and that the difference we observed between variants 1 and 2 may reflect the existence of this polymorphism. In reference to this, notice that arginine peptide 5 of H2A.1 is absent in H2A.2. Variants X and Z also have ratios closer to variant 1 than 2; however, since variants X and Z have different apparent molecular weights on $NaDodSO_4$ gels, lysines and arginines may have been inserted or deleted as well as substituted. We conclude that the arginine/lysine ratios of variants X and Z are characteristic of H2A.

H2A Variants in Nucleosomes. In models of chromatin, histones 4, 3, 2B, and 2A are positioned in a core particle surrounded by approximately 1.75 coils of DNA. Histone variants should also be found in core particles. Digestion of 0.7 M NaCl extracted chromatin results in two monosome bands (Figure 8). In the core particles ubiquitin derivatives are absent or extremely deficient; in the slower monosome band more normal amounts of ubiquitin adducts are found. Figure 8 shows that H2A.X and H2A.Z are found in both types of monosomes as well as in larger particles. This result indicates that H2A.X and H2A.Z can substitute for H2A.1 and H2A.2 in the nucleosome.

Stoichiometry of H2A Variants. The nucleosomal model of chromatin structure (Kornberg, 1974) also postulates equimolar amounts of each core histone. When the molar yields of all the known histone variants were calculated, a value of 0.98 was obtained for the sum of all the H2A variants (Table I). These data show that these eight H2A forms satisfy the equimolar criterion. Table I also shows that a similar fraction of each of the four H2A variants has ubiquitin attached, indicating that ubiquitin attachment shows little or no selectivity with respect to the H2A variant.

Occurrence of H2A Variants. We have found H2A.X and H2A.Z in mouse L1210 cells, Friend leukemia cells, and mouse liver, kidney, spleen, fetus, placenta, and brain. We have also found these two variants in two human lines, HeLa and HL60, and in chicken red blood cells. They can be acid extracted from nucleosomes, nuclei, cells, or whole tissues and are also displaced from nuclei and nucleosomes by protamine.

However, in none of the cells and tissues studied are the X and Z variants more than 15% of the total H2A. These results indicate that H2A.X and H2A.Z are significant but quantitatively minor fractions of the H2A family in higher animals.

Discussion

We have purified two proteins by gel electrophoresis which appear to be core histones, more specifically, members of the histone 2A family. In common with core histones these two proteins (1) are basic proteins, (2) are not extracted with 0.7 M NaCl, (3) are extracted with HCl, (4) are not labeled with [^{14}C]tryptophan, (5) are found in nucleosome cores, (6) migrate like core histones in $NaDodSO_4$ gels, and (7) migrate like core histones in acetic acid–urea gels.

In addition to these seven characteristics shared with all core histones, these two proteins are shown to belong to the H2A family by the following characteristics. (1) Their [^{14}C]arginine peptide maps are related to that of H2A but not to those of the other core histones. (2) They contain the peptide Ala-Gly-Leu-Gln-Phe-Pro-Val-Gly-Arg. (3) The arginine/lysine ratios of H2A.X and H2A.Z are typical of H2A and different

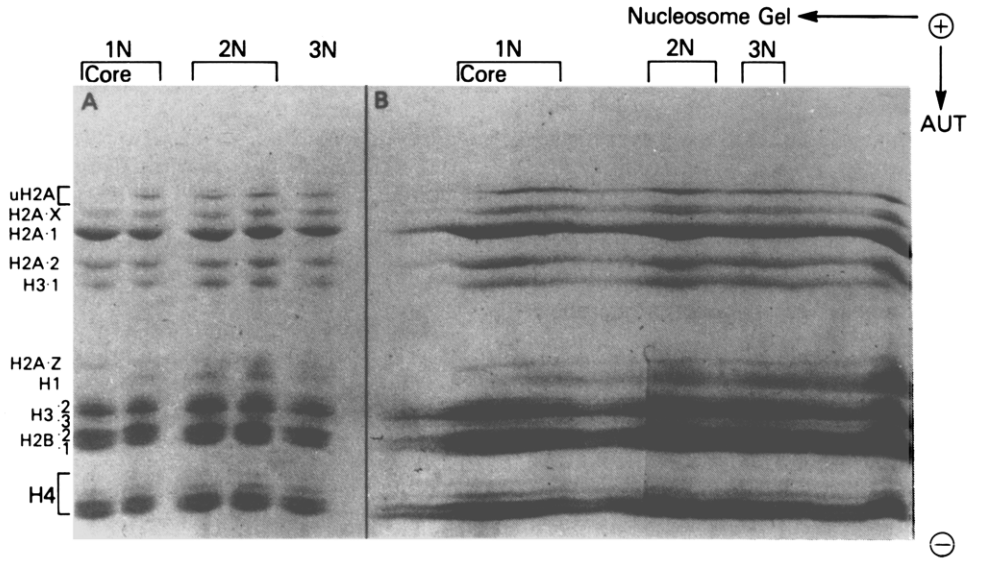

FIGURE 8: Analysis of the histone complements of different nucleosome classes. Nucleosomes prepared from salt-washed nuclei by micrococcal nuclease digestion were electrophoresed in a 5% acrylamide gel as described under Materials and Methods. After being stained with ethidium bromide, individual bands (A) or a section of the gel (B) was excised and the histone content analyzed by electrophoresis in a 15% AUC gel. Measured relative to *Hae*III restriction fragments of ϕX 174RF DNA (Bethesda Research Laboratories), the monosome fraction was found to have an average length of 160 bp DNA. In this preparation some H1 remained after salt treatment, although not in core particles.

Table I: Histone Stoichiometry

(A) H2A Family

variant	cpm	cpm/Arg	molar ratio	variant	cpm	cpm/Arg	molar ratio	uH2A/H2A
H2A.1	140 980	11 748	0.509	uH2A.1	23 239	1453	0.063	0.110
H2A.2	64 241	5 353	0.232	uH2A.2	9 967	623	0.027	0.104
H2A.X	29 594	2 466	0.107	uH2A.X	5 219	326	0.014	0.117
H2A.Z	11 972	998	0.043	uH2A.Z	2 021	126	0.006	0.112
								av: 0.111

(B) Total Histones

histone	cpm	no. of Arg	cpm/Arg		molar ratio
H2A	246 787	12	20 565	23 093	0.977
uH2A	40 446	16	2 528		
H2B.1	92 655	8	11 582	25 215	1.067
H2B.2	109 066	8	13 633		
H3.1	89 631	18	4 979	21 500	0.910
H3.2,.3	297 385	18	16 521		
H4	346 387	14	24 742	24 742	1.047
				av: 23 638	1.000

from the other core histones. (4) H2A.X and H2A.Z have a high affinity for Triton X-100 as do H2A.1 and H2A.2. (5) H2A.Z and H2A.X have ubiquitin adducts as do H2A.1 and H2A.2. (6) The ubiquitin adduct is the same percentage of each H2A variant. (7) In addition, the ubiquitin adducts also have (a) similar [^{14}C]arginine peptide maps and (b) arginine/lysine ratios typical of uH2As. Martinson et al. (1979) have already shown that uH2A.1 replaces H2A.1 as an integral component of the nucleosome core. These findings, taken together with the finding that the sum of the relative molar ratios of these eight species very closely equals the relative molar ratios of these eight species very closely equals the relative molar ratios of the other core histones, lead us to the conclusion that H2A.X and H2A.Z and their ubiquitin adducts are integral components of the nucleosome.

H2A.X and H2A.Z are not labeled with [^{35}S]methionine or [^{35}S]cysteine. This lack of sulfur in these two proteins confirms their unrelatedness to H2B, H3, and H4 since these histones all contain methionine in integral sequences that have been conserved since the divergence of plants and animals (von Holt et al., 1979).

One H2A variant, H2A.2, does contain methionine at position 51. The existence of this variant allows us to rule out posttranslational modification as a probable origin for H2A.X and H2A.Z, since the known posttranslational H2A modifications, phosphorylation and addition of ubiquitin, affect the two homeomorphous derivatives H2A.1 and H2A.2 equally. If H2A.X and H2A.Z were derived by posttranslational modification, one would expect to find sulfur-containing analogues of H2A.X and H2A.Z from H2A.2. We have looked specifically for such analogues without success. Another piece of evidence which strongly argues against a posttranslational origin of H2A.X and H2A.Z is that some internal arginine peptides of H2A.1 are not found in H2A.X or H2A.Z. These two results, which eliminate posttranslational modification as a explanation, also eliminate isolation artifacts as a source of H2A.X and H2A.Z since sulfur-containing analogues of these two variants should also result in this case.

Since they are separable from H2A.1 on $NaDodSO_4$ and AUC gels, we have termed H2A.X and H2A.Z heteromorphous variants in order to distinguish them from variants that are not separable from the major forms on $NaDodSO_4$ or AUC gels. Such variants of H2A have also been characterized in wheat germ (Rodrigues et al., 1979). Five variants are

found in three bands on acetic acid–urea gels. Alignment of wheat germ partial sequences with the calf H2A sequence shows that some of the wheat germ H2A variants have a seven amino acid extension on the amino terminus. This result indicates that variants which can be resolved on acetic acid–urea or $NaDodSO_4$ gels are likely to have different lengths. A similar study of the complete sequence of three H2B variants from sea urchin sperm yielded lengths of 143, 144, and 148 amino acids (Strickland, M., et al., 1977, 1979; Strickland, W. N., et al., 1977). Coupled with the above, our results with human, mouse, and chicken indicate that heteromorphous families of H2A or H2B may be a common if not universal characteristic of eukaryotes.

Using $NaDodSO_4$ gel electrophoresis, several authors have determined the relative stoichiometry of the core histones by radioactive amino acid incorporation (Joffe et al., 1977; Rall et al., 1977; Albright et al., 1979). Although all these authors concluded that H2A or the sum of H2A and uH2A was equimolar or nearly equimolar with the other core histones, in all cases the value for H2A was lower than the others. Since in L1210, H2A.X, H2A.Z, uH2A.X, and uH2A.Z account for $\sim$15% of the total H2A and this value may be different for other cell types, the relative stoichiometry for H2A and uH2A obtained by these authors is completely consistent with our results. The relative amount of H2A.X and H2A.Z is in most cases probably less than the experimental uncertainty in those experiments. Likewise, our stoichiometric results cannot rule out the existence of other minor variants of H2A or any other histone. What they do show, however, is that the amounts of the eight H2A proteins are consistent with the equimolar hypothesis of nucleosome structure. In fact, we have recently found ubiquitin adducts of H2B which in L1210 cells account for $\sim$1% of the H2B (M. H. P. West and W. M. Bonner, unpublished experiments).

During the course of our work we became aware of work (Zweidler, 1976, 1978) describing several minor proteins which the author concludes have properties typical for histones but whose amino acid compositions are not closely related to any of the major histone classes. Three of these proteins migrate in AUT and AUC gels similarly to H2A.X, H2A.Z, and uH2A.1. When the approximate amino acid stoichiometry of M1, Zweidler's possible counterpart to H2A.Z, is calculated from its amino acid composition and our molecular weight and compared to that of H2A.1, 11 of 17 amino acids agree within one residue, the other 4 within three. In addition, like H2A.Z and H2A.X, M1 and M2 have no methionine or cysteine. More recently, Urban et al. (1979) have reported that M2 from chicken appears to be an H2A-like variant. Albright et al. (1979) have reported that M1 and M2 are present in monosomes and have proposed that they may replace H2A, which they found to be below equimolar in monosomes. From these studies, we felt that M1 is probably H2A.Z and M2 is probably H2A.X. No counterparts for uH2A.Z and uH2A.X were reported in these studies.

The paramount question concerns the role of H2A.X and H2A.Z in the structure and function of chromatin. So far we have found H2A.X and H2A.Z as minor components in all tissues and species studied. These two variants are found in nucleosome cores in approximately the same relative amounts as in whole chromatin. This result suggests that the presence of H2A.X and H2A.Z in nucleosomes does not alter their sensitivity to micrococcal nuclease. Whether or not nucleosomes containing H2A.X and H2A.Z are of different sizes or have different sensitivities to other nucleases are questions that remain to be answered. Weisbrod & Weintraub (1979) have published studies which indicate that differences in DNase I sensitivity of various regions of the genome, while directly related to the presence of high mobility group (HMG) proteins, are also related to a predisposition of these regions to bind HMG proteins. It is possible that this predisposition is related to the histone content of these regions of the genome. The heteromorphous H2A variants may be important in such functional differentiations of various regions of the genome.

References

Albright, S. C., Nelson, P. P., & Garrard, W. T. (1979) *J. Biol. Chem. 254*, 1065–1073.

Blankstein, L. A., & Levy, S. B. (1976) *Nature (London) 260*, 638–640.

Blankstein, L. A., Stollar, B. D., Franklin, S. G., Zweidler, A., & Levy, S. B. (1977) *Biochemistry 16*, 4557–4562.

Bonner, W. M., West, M. H. P., & Stedman, J. D. (1980) *Eur. J. Biochem.* (in press).

Cook, W. J., Suddath, F. L., Bugg, C. E., & Goldstein, G. (1979) *J. Mol. Biol. 130*, 353–355.

DeLange, R. J., Fambrough, D. M., Smith, E. L., & Bonner, J. (1969) *J. Biol. Chem. 244*, 319–334.

DeLange, R. J., Hooper, J. A., & Smith, E. L. (1973) *J. Biol. Chem. 248*, 3261–3274.

Dixon, G. H., Candido, E. P. M., Honda, B. M., Louie, A. J., MacLeod, A. R., & Sung, M. T. (1975) *Ciba Found. Symp. 28*, 229–248.

Felsenfeld, G. (1978) *Nature (London) 271*, 115–122.

Franklin, S. G., & Zweidler, A. (1977) *Nature (London) 266*, 273–275.

Giri, C. P., West, M. H. P., & Smulson, M. (1978) *Biochemistry 17*, 3495–3500.

Goldknopf, I. L., & Busch, H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 864–868.

Goldknopf, I. L., Taylor, C. W., Baum, R. M., Yeoman, L. C., Olson, M. O. J., Prestayko, A. W., & Busch, H. (1975) *J. Biol. Chem. 250*, 7182–7187.

Goldknopf, I. L., French, M. F., Daskal, Y., & Busch, H. (1978) *Biochem. Biophys. Res. Commun. 84*, 786–793.

Joffe, J., Keene, M., & Weintraub, H. (1977) *Biochemistry 16*, 1236–1288.

Kornberg, R. D. (1974) *Science (Washington, D.C.) 184*, 868–871.

Laine, B., Sautiere, P., & Biserte, G. (1976) *Biochemistry 15*, 1640–1645.

Martinson, H. G., True, R., Burch, J. B. E., & Kunkel, G. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 1030–1034.

Newrock, K. M., Alfageme, C. R., Nardi, R. V., & Cohen, L. H. (1977) *Cold Spring Harbor Symp. Quant. Biol. 42*, 421–431.

O'Farrell, P. Z., Goodman, H. M., & O'Farrell, P. H. (1977) *Cell 12*, 1133–1142.

Olson, M. O. J., Goldknopf, I. L., Guetzow, K. A., James, G. T., Hawkins, T. C., Mays-Rothberg, C. J., & Busch, H. (1976) *J. Biol. Chem. 251*, 5901–5903.

Rall, S. C., Okinaka, R. T., & Strniste, G. F. (1977) *Biochemistry 16*, 4940–4944.

Rodrigues, J. De A., Brandt, W. F., & von Holt, C. (1979) *Biochim. Biophys. Acta 578*, 196–206.

Shaw, B. R., & Richards, R. G. (1979) *NATO Adv. Study Inst. Ser., Ser. A 21a*, 125–136.

Strickland, M., Strickland, W. N., Brandt, W. F., & von Holt, C. (1977) *Eur. J. Biochem. 77*, 263–275.

Strickland, M., Strickland, W. N., Brandt, W. F., von Holt, C., Wittmann-Liebold, B., & Lehmann, A. (1979) *Eur. J. Biochem. 89*, 443–452.

Strickland, W. N., Strickland, M., Brandt, W. F., & von Holt, C. (1977) *Eur. J. Biochem. 77*, 277–286.
Thomas, C. B., Kohn, K. W., & Bonner, W. M. (1978) *Biochemistry 17*, 3954–3958.
Urban, M. K., Franklin, S. G., & Zweidler, A. (1979) *Biochemistry 18*, 3952–3960.
Varshavsky, A. J., Bakayev, V. V., & Georgiev, G. P. (1976) *Nucleic Acids Res. 3*, 477–492.
von Holt, C., Strickland, W. N., Brandt, W. F., & Strickland, M. S. (1979) *FEBS Lett. 100*, 201–218.
Weisbrod, S., & Weintraub, H. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 630–634.
Wong, N. C. W., Poirier, G. G., & Dixon, G. H. (1977) *Eur. J. Biochem. 77*, 11–21.
Zanetta, J. P., Vincendon, G., Mandel, P., & Gombos, G. (1970) *J. Chromatogr. 51*, 441–458.
Zweidler, A. (1976) *Life Sci. Res. Rep. 4*, 187–196.
Zweidler, A. (1978) *Methods Cell Biol. 17*, 223–233.

Cycling of Ribonucleic Acid Polymerase To Produce Oligonucleotides during Initiation in Vitro at the *lac* UV5 Promoter†

Agamemnon J. Carpousis* and Jay D. Gralla

ABSTRACT: High-resolution gel electrophoresis has been used to detect and quantitate promoter-specific oligonucleotides produced during initiation of transcription in vitro at the lactose operon (*lac*) UV5 promoter. The resolved products are RNA species of various lengths which correspond to the initial *lac* mRNA sequence. Quantitation shows that many oligonucleotides can be formed per preinitiation complex, including species as long as hexanucleotide. Synthesis occurs without dissociation of the enzyme, as evidenced by levels of synthesis in the presence of heparin, a selective inhibitor of free RNA polymerase. Thus, RNA polymerase cycles at this promoter in vitro producing oligonucleotides reiteratively. In general, the yield of oligonucleotides decreases when the total concentration of all four substrates is increased or when a missing nucleoside triphosphate substrate is added. Nevertheless, oligonucleotide synthesis persists under all conditions tested. Strikingly, the dinucleotide always represents 50% of the total of all oligonucleotides, even when conditions are manipulated to cause a 100-fold variation in this total. This shows that, after formation of the first phosphodiester bond at the *lac* UV5 promoter, dissociation of the dinucleotide is as likely as formation of the second phosphodiester bond. As discussed above, after release of a small RNA, RNA polymerase may then begin another RNA chain, which is again subject to premature release. These considerations lead to a model in which RNA polymerase cycles to produce oligonucleotides during initiation of transcription at the *lac* UV5 promoter in vitro. Production of a long RNA transcript is then essentially an escape from this cycling reaction. The drug rifampicin, which drastically inhibits escape to produce RNA, limits, but does not prevent, the cycling reaction.

Production of RNA by the DNA-dependent RNA polymerase is accomplished in a series of steps. These include formation of a promoter-specific preinitiation complex between enzyme and DNA, conversion to an elongation complex, and termination of transcription (for a review, see Chamberlin, 1976a). The rate and mechanism of each of these steps must be determined in order to understand the factors that may contribute the overall yield of RNA over a wide range of conditions.

The conversion of preinitiation complexes to elongation complexes, often termed "initiation", has been studied using a variety of assays which attempt to isolate this step from others along the pathway. Mangel & Chamberlin (1974) used an assay that relies on the properties of the inhibitor rifampicin. This assay relied on an assumed relationship between the mechanism of rifampicin inhibition and the mechanism of initiation. This assumption has now been shown to be at the least an oversimplification (McClure et al., 1978; McClure & Cech, 1978).

An alternative to this method involves limiting the reaction to the formation of a single phosphodiester bond by exclusion of appropriate nucleoside triphosphates (Johnson & McClure, 1976). This "abortive initiation" reaction leads to reiterative synthesis of dinucleoside tetraphosphate. However, the initiation rate and mechanism determined may not accurately reflect the productive initiation of transcript, since artificial steady-state conditions have been imposed by the absence of substrates required for conversion to elongation complexes.

A third assay involves observation of the rate of acid-precipitable RNA production after addition of nucleoside triphosphates to preinitiation complexes (Stefano & Gralla, 1979; Nierman and Chamberlin, 1979). Such rates, however, can be very rapid, so the utility of this assay is restricted to in vitro systems or conditions where initiation is a relatively slow process.

Thus, in part because of these uncertainties and limitations, the mechanism of the initiation process remains unsettled. An important complication is the recent observation that formation of the first phosphodiester bond to produce dinucleoside tetraphosphate is not always followed by extension to complete RNA transcript. Specifically, McClure & Cech, (1978) have shown that under continuous transcription conditions (all four nucleoside triphosphates present, and no inhibitors added) approximately 2 mol of the promoter-specific oligonucleotide pppApApC accumulates per mol of long-chain RNA from the $\lambda P_{R'}$ promoter. Thus, for this promoter the mechanism of conversion of preinitiation complexes to elongation complexes is more complicated than anticipated.

A systematic and quantitative investigation of such potential oligonucleotide products of transcription has not been reported.

† From the Biochemistry Division of the Chemistry Department and Molecular Biology Institute, University of California, Los Angeles, California 90024. *Received December 26, 1979.* Supported by Grant CA 19941 from the National Cancer Institute and a grant from the UCLA Academic Senate. A.J.C. was supported as a trainee of U.S. Public Health Service Grant GM 07185.

0006-2960/80/0419-3245$01.00/0