of eucaryotes and procaryotes are also seen. Using a peptidyl-tRNA affinity-label analogue to covalently label 70S *E. coli* ribosomes, Pellegrini et al. (1974) found the major reactive species to be a small, basic protein of the large subunit (L27). This reagent requires nucleophilic attack for the labeling to occur. The most likely residue for alkylation is the cysteine sulfhydryl. L27 is known to contain a reactive cysteine group in the intact particle (Moore, 1971; Bakardjieva & Crichton, 1974). These results suggest the presence of a sulfhydryl group at the P site of both eucaryotic and procaryotic peptidyl transferases. The labeling of a small, basic protein in *Drosophila* ribosomes suggests conservation of at least size and electrophoretic behavior and, possibly, some specific amino acid residues within the P site of both procaryotic and eucaryotic ribosomes. In collaboration with P. Butler and R. Traut, we are undertaking studies using this mercurated reagent and *E. coli* ribosomes to further explore these possibilities.

References

Bakardjieva, A., & Crichton, R. (1974) *Biochem. J. 143*, 599–606.
Bermek, E., Mönkemeyer, H., & Berg, R. (1971) *Biochem. Biophys. Res. Commun. 45*, 1294–1299.
Bonner, W. M., & Laskey, R. A. (1974) *Eur. J. Biochem. 46*, 83–91.
Carrasco, L., & Vâsquez, D. (1975) *Eur. J. Biochem. 50*, 317–323.
Czernilofsky, P. A., Collatz, E., Gressner, A. M., & Wool, I. G. (1977) *Mol. Gen. Genet. 153*, 231–235.
Dale, R. M. K., Martin, E., Livingston, D. C., & Ward, D. C. (1975) *Biochemistry 14*, 2447–2457.
Gosâlbez, M. P., Vâsquez, D., & Ballesta, J. P. G. (1978) *Mol. Gen. Genet. 163*, 29–34.
Harris, R. J., & Pestka, S. (1977) in *Molecular Mechanisms of Protein Biosynthesis* (Weissbach, H., & Pestka, S., Eds.) p 413, Academic Press, New York.
Howard, G. A., & Traut, R. R. (1973) *FEBS Lett. 29*, 177–180.
Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265–275.
Miller, D. L., & Weissbach, H. (1977) in *Molecular Mechanisms of Protein Biosynthesis* (Weissbach, H., & Pestka, S., Eds.) p 324, Academic Press, New York.
Monro, R. E. (1971) *Methods Enzymol. 20*, 472–481.
Moore, P. B. (1971) *J. Mol. Biol. 60*, 169–184.
Pellegrini, M., Oen, H., Eilat, D., & Cantor, C. (1974) *J. Mol. Biol. 88*, 809–829.
Pestka, S. (1971) *Methods Enzymol. 20*, 502.
Pestka, S. (1977) in *Molecular Mechanisms of Protein Biosynthesis* (Weissbach, H., & Pestka, S., Eds.) p 468, Academic Press, New York.
Reyes, R., Vâsquez, D., & Ballesta, J. P. G. (1977) *Eur. J. Biochem. 73*, 25–31.
Sherton, C. C., & Wool, I. G. (1974) *Mol. Gen. Genet. 135*, 97–112.
Stockmar, R., Ringer, D., & McAllister, H. (1972) *FEBS Lett. 23*, 51–55.

Sequence of Histone 2B of *Drosophila melanogaster*†

Sarah C. R. Elgin,* James Schilling, and Leroy E. Hood

ABSTRACT: The complete sequence of histone 2B of *Drosophila* has been determined by using an improved Beckman sequenator. Comparing these data with those previously published by other investigators on the histone 2B of calf [Iwai, K., Hayashi, H., & Ishikawa, K. (1972) *J. Biochem. (Tokyo) 72*, 357–367], trout [Koostra, A., & Bailey, G. S. (1978) *Biochemistry 17*, 2504–2510], and *Patella* (a limpet) [van Helden, P. D., Strickland, W. N., Brandt, W. F., & von Holt, C. (1979) *Eur. J. Biochem. 93*, 71–78], it is possible to assess the evolutionary stability of this protein. There is little conservation of sequence in the N-terminal portion of the molecule (residues 1–26 numbering according to calf H2B), while the remainder of the protein, which we designate the C-terminal portion, is highly conserved. In the region of 27–125 residues, there are 9 substitutions in the composite data among the 98 positions, 8 of them conservative. These data indicate that very different selective pressures operate on the two different portions of the H2B molecule, implying the existence of two well-defined regions. Studies on the structure of the nucleosome by others have suggested that the C-terminal portion of H2B is involved in histone–histone interactions while the N-terminal portion is a relatively free "tail" binding to DNA. The sequence data indicate that the function of the C-terminal region of H2B requires considerable sequence specificity while that of the N-terminal region does not.

During the last few years considerable evidence has been obtained which has led to and supported the nucleosome or ν body model of chromatin structure. The chromatin fiber is visualized as a string of beads, each bead made up of eight molecules of the smaller histones, 2A, 2B, 3 and 4, around which the DNA is wrapped. There are ~200 base pairs of DNA associated with each unit. Histone 1 and the nonhistone chromosomal proteins are apparently associated with the DNA on the outside of the core structure. For reviews of the evidence leading to this model and a more detailed discussion, see Elgin & Weintraub (1975), Kornberg (1977), and Felsenfeld (1978).

The pioneering studies of Fambrough, DeLange, and their colleagues (Fambrough & Bonner, 1968; DeLange et al.,

† From the Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138 (S.C.R.E.), and the Division of Biology, California Institute of Technology, Pasadena, California 91125 (J.S. and L.E.H.). *Received April 30, 1979; revised manuscript received September 27, 1979.* Supported by Grant PCM 75-15259 from the National Science Foundation to S.C.R.E. and Grant GM 06945 from the National Institutes of Health to L.E.H. S.C.R.E. is the recipient of a Research Career Development Award from the National Institutes of Health.

0006-2960/79/0418-5679$01.00/0

1969) demonstrated that the histones are among the most conserved proteins yet analyzed. Similarly, the nucleosome organization of the chromatin fiber is a highly conserved structure, having been observed for essentially all eucaryotes and for all forms of chromatin (interphase, metaphase, polytene) studied to date [e.g., Woodcock et al. (1976)]. It is of interest to determine and compare the primary structures of the histones, the building blocks of nucleosome structure, for several reasons. First, the pattern and degree of sequence conservation should indicate which portions of the molecules are necessary in an invariant form for nucleosome structure and function. Differences in the degree of conservation within a given histone may indicate the positions of different functional domains having different evolutionary constraints. Second, the availability of different histones whose sequences are completely determined should aid in studying the interactions of the histones with each other and with other components of the chromatin complex. Third, the availability of different histones whose structures have been completely determined will allow one to test the specificity of components of the chromatin system, such as histone modification enzymes. It is of particular interest to work with organisms where genetic manipulation is possible and where studies of the histone gene sequences are underway (Lifton et al., 1978). For these reasons, we have determined the complete sequence of histone 2B of *Drosophila*.

Methods

Isolation of Histone 2B. Histones were isolated by extraction with 1.6 N NaCl–0.2 N HCl from crude chromatin of 6–18-h-old *Drosophila melanogaster* Oregon R embryos prepared by the method of Elgin & Hood (1973). The histones were then dialyzed against 0.01 M acetic acid, lyophilized, and stored at −20 °C for future use. Eighty milligrams of histone was redissolved in 0.01 M acetic acid, and histone 1 was removed by extraction with 5% perchloroacetic acid. This step is necessary since histone 1 of *Drosophila* contains cysteine (Alfageme et al., 1974; S. C. R. Elgin, unpublished experiments). The pellet was dissolved in 8 M urea–0.05 M Tris-HCl (pH 8.0)–1% β-mercaptoethanol and dialyzed extensively against 0.1 M Tris-HCl (pH 8)–0.3 M NaCl—1 mM ethylenediaminetetraacetic acid (column buffer, CB). The proteins were applied in this buffer to a 25 mm × 12 cm column of preequilibrated activated thio-Sepharose 4B resin (Pharmacia). Histones 2A and 2B were eluted in the runoff peak (flow rate 4.5 mL/h) with some H3 contamination. The bulk of the H3 was covalently bound to the column. H4 remained in association with H3 and could be eluted with CB + 5 M urea, indicating the expected hydrophobic interaction. H3 was eluted with CB + 5 M urea + 20 mM L-cysteine. The column could be regenerated by washing with CB + 1.5 mM dithiouracil. Recovery of purified histone was ~75%. Histone 2B was purified from the runoff peak by repeated extraction with 80% ethanol–0.26 M HCl (Oliver et al., 1972). Electrophoresis on acid–urea gels indicated that the product was 85–90% H2B, the single major contaminant being H2A. No subfractions of histone 2B were observed in this work, nor have any been reported for *Drosophila* in other studies (Oliver & Chalkley, 1972; Alfageme et al., 1974).

Sequencing Strategy. The N-terminal sequence of the H2B protein was determined for residues 1–62 on an improved Beckman sequenator (data shown in Figure 1). A portion of the H2B protein was blocked at its N-terminal α-amino and lysine ε-amino groups by succinylation. This "blocked" protein was cleaved with cyanogen bromide at positions 56 and 59, and the resulting mixture of three peptides (1–56, 57–59, and 60–122) was loaded in the sequencer without further purification. The two unblocked peptides (57–59 and 60–122) were sequenced simultaneously for 59 residues (data shown in Figure 2). The H2B protein was contaminated with ~10% H2A protein, and the sequence of a cyanogen bromide fragment from H2A also could be followed. We also carried out a tryptic digestion of succinylated H2B and sequenced the "arginine" peptides 31–69 and 97–122 simultaneously as a mixture (data not shown). The arginine peptides were sequenced to within four residues of the C terminus of the C-terminal peptide.

Succinylation. The H2B was dissolved at 10 mg/mL in 8 M guanidine hydrochloride and reacted on a pH stat at pH 9.0 with a 100-fold excess of succinic anhydride over protein (w/w) at room temperature.

Cyanogen Bromide Cleavage. The H2B protein was dissolved at 10 mg/mL in 70% formic acid and reacted for 18 h at 4 °C with a fivefold excess of cyanogen bromide (w/w).

Tryptic Digestion. The H2B protein was dissolved at 10 mg/mL in 0.2 M ammonium bicarbonate and digested at 37 °C for 2.5 h with 250 μg/mL trypsin.

Preparation of Arginine Fragments. The mixture of peptides resulting from tryptic digestion of succinylated H2B was separated on Sephadex G-15 in 0.2 M ammonia into several peaks. The first peak eluted from the column contained arginine fragments extending from residues 31 to 69 and 97 to 122. This mixture of two peptides was analyzed on the automatic sequenator.

Automated Sequence Analysis. Polypeptides were sequenced on a modified Beckman 120B sequencer as described by Hunkapiller & Hood (1978). The phenylthiohydantoin amino acids were identified by high-pressure liquid chromatography on Du Pont Zorbax ODS columns.

Carboxypeptidase Digestion. The H2B protein was dissolved in 0.2 M ammonium bicarbonate. A mixture of carboxypeptidases A and B (both DFP-treated; Sigma) was added at time 0 at a histone/enzyme ratio of 100:1. Digestion was at room temperature for 5 or 60 min. After being boiled for 2 min to stop the digestion, samples were lyophilized and analyzed on a Durrum D-500 amino acid analyzer.

Results and Discussion

Automated Sequence Analysis. The amino acid sequence for positions 1–62 is given unambiguously in Figure 1. The results have been presented by plotting nanomoles of phenylthiohydantoin amino acids recovered vs. residue position for each amino acid. The samples for steps 39–42 were accidently pooled and analyzed together. The sequence of this region was determined by the sequence analysis of a tryptic (arginine) peptide containing residues 31–69 (data not shown).

The sequence analysis of the cyanogen bromide fragment starting at position 60 is unambiguous out to the threonine at position 119 (Figure 2). Indeed, these data suggest that positions 120 and 121 are Ser-Ser, although the signal-to-noise ratio makes this latter assignment tenuous. Arginine peptides 31–69 and 97–122 have sequences which confirm completely the results given in Figures 1 and 2.

Carboxypeptidase Analysis. A 5-min digestion with carboxypeptidases A and B resulted in the release of lysine, thus identifying this residue as the C-terminal amino acid. A longer digestion time resulted in the release of additional amino acids which are consistent with the sequence assignment of the C-terminal region of histone 2B given in Figure 3.

Summary of Sequence Data. The final sequence of histone H2B of *D. melanogaster* with a summary of the evidence used is given in Figure 3. The assignment of Ser-Ser at positions

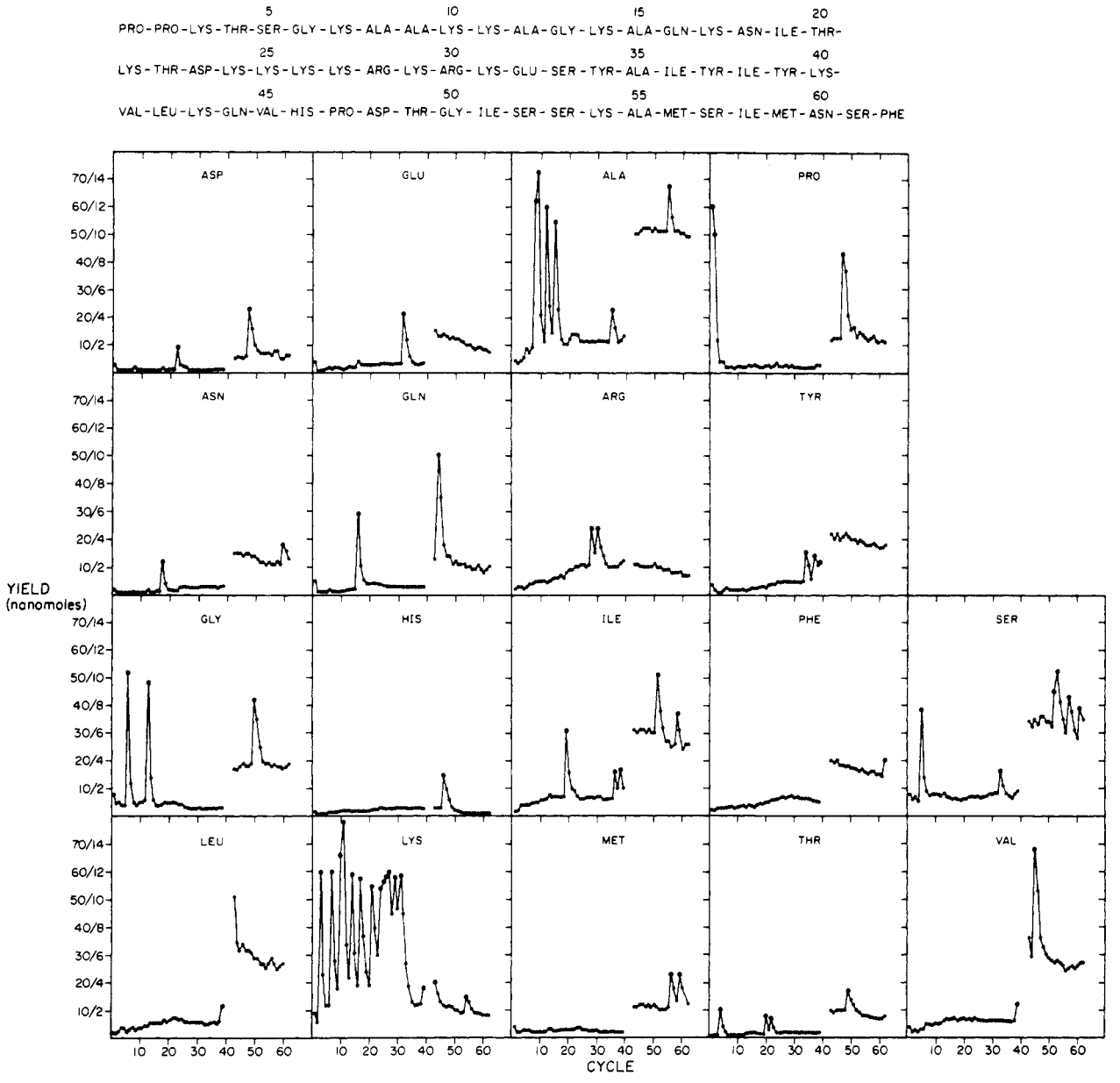

FIGURE 1: Nanomoles of each phenylthiohydantoin amino acid recovered at each step for the H2B N-terminal sequence. The amino acid identified at each step is indicated by a heavier dot. Samples for steps 39–42 were accidentally pooled, and the four amino acids were therefore identified together. The scale is expanded fivefold at cycle 43, except for lysine which is not expanded and histidine and arginine which are plotted throughout on the expanded scale.

120 and 121 remains tentative. The sequence is in complete agreement with the sequence of the H2B gene from the DNA clone cDm 500 from *D. melanogaster* which has been completed for residues 1–117 (M. Goldberg and D. S. Hogness, personal communication). In addition, the data obtained indicate that *Drosophila* H2A, like that of calf, has a blocked N-terminal residue, but unlike calf has a methionine at position 55. The H2A sequence from residues 55–95 (numbering according to calf thymus H2A) can for the most part be detected as the minor peaks in Figure 2.

Conservation of Sequence. A comparison of the sequence of *Drosophila* H2B with complete H2B protein sequences previously determined by others is presented in Figure 4. In preparing Figure 4, an attempt was made (by visual inspection) to align the proteins for maximum homology, using insertions and deletions as necessary. In the following discussion we will refer to amino acid positions numbered according to the calf protein, as in Figure 4. While H2B's of calf and trout are fairly homologous (1 deletion, 7 substitutions), it is readily apparent that H2B of *Drosophila* differs more extensively from its calf counterpart [6 sequence gaps (deletions or insertions), 15 substitutions]. In fact, there is very little sequence homology for the region 1–26 (numbered according to the calf sequence) when the composite data for these three proteins are examined. Within the region 27–125, homology is considerable; there are no sequence gaps and only eight substitutions of which seven are conservative. While this manuscript was in preparation, the complete amino acid sequence of H2B from *Patella granatina*, a limpet, was reported (van Helden et al., 1979). Comparison with calf H2B indicates a minimum of 5 gaps, all before position 27, and 21 substitutions, 15 conservative. As would be expected from the established phylogenetic relationships, *Patella* H2B shows considerably greater homology with *Drosophila* H2B; there are only 2 gaps, both before position 27, and 12 substitutions, 10 conservative, in this comparison.

The results suggest that there are two quite different regions in H2B, possibly indicative of different functions with very

```
                5                  10                  15                  20
ASN-SER-PHE-VAL-ASN-ASP-ILE-PHE-GLU-ARG-ILE-ALA-ALA-GLU-ALA-SER-ARG-LEU-ALA-HIS-
                25                  30                  35                  40
TYR-ASN-LYS-ARG-SER-THR-ILE-THR-ARG-GLU-ILE-GLN-THR-ALA-VAL-ARG-LEU-LEU-LEU-PRO-
                45                  50                  55
GLY-GLU-LEU-ALA-LYS-HIS-ALA-VAL-SER-GLU-GLY-THR-LYS-ALA-VAL-THR-LYS-TYR-THR
```

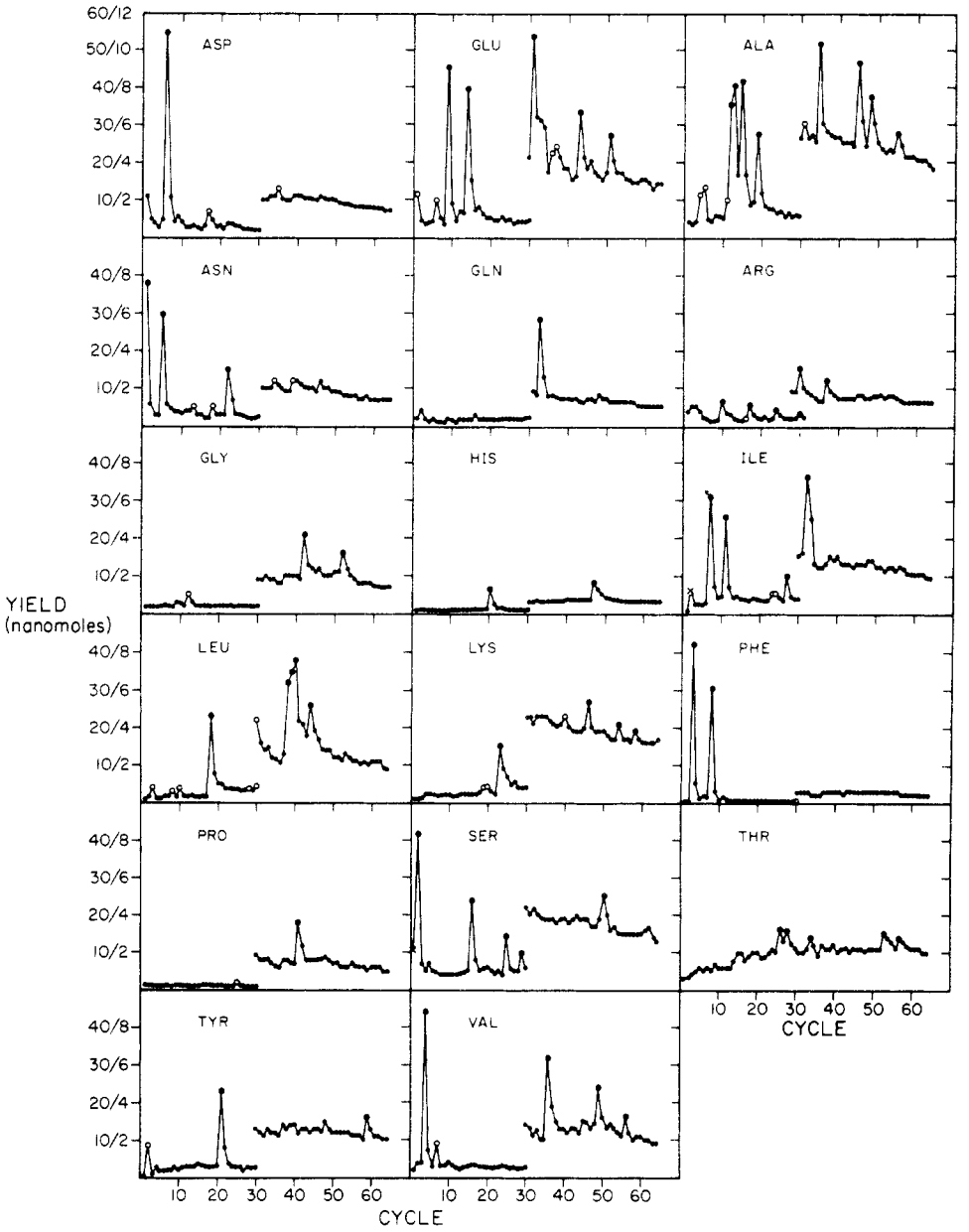

FIGURE 2: Nanomoles of each phenylthiohydantoin amino acid recovered at each step of the sequence of H2B cyanogen bromide peptides. Three sequences can be seen: (●) H2B 60–122; (×) H2B 57–59; (○) contaminating H2A sequences (see Results and Discussion). The N-terminal peptide was blocked. The scale is amplified fivefold at cycle 30 for all amino acids, except threonine which is plotted throughout on the expanded scale.

different requirements for sequence specificity. The N-terminal region, residues 1–26, can be defined as that portion of the molecule which lacks extensive sequence homology, which includes sequence gaps in comparisons of H2B sequences, and which includes most of the clusters of basic amino acids. The boundary of the conserved and nonconserved regions has been assessed by examination of all the currently published data. The sequence of a sea urchin (*Strongylocentrotus purpuratis*) embryo H2B has recently been reported from DNA sequencing of the gene (Sures et al., 1978). (Since the histone genes are repeated, one cannot yet be certain that this gene is expressed.) A comparison of this sequence with that of calf H2B indicates 3 gaps, all occurring before position 27, and 22 substitutions, 15 conservative (see Figure 4). In addition to the protein and DNA sequence data from somatic cell H2B histones used in constructing Figure 4, data on the N-terminal region of the analogous protein from sea urchin sperm may be considered. Five H2B proteins from *Parechinus angulosus* and *Pasammechinus miliaris* sperm have recently been isolated and sequenced (Strickland et al., 1977a,b, 1978). The sea urchin sperm histone 2B's are ~15% larger than those of calf, trout, and *Drosophila*. An excellent alignment, however, can be obtained between these proteins and calf H2B starting at residue 27. The common sequence for these proteins at

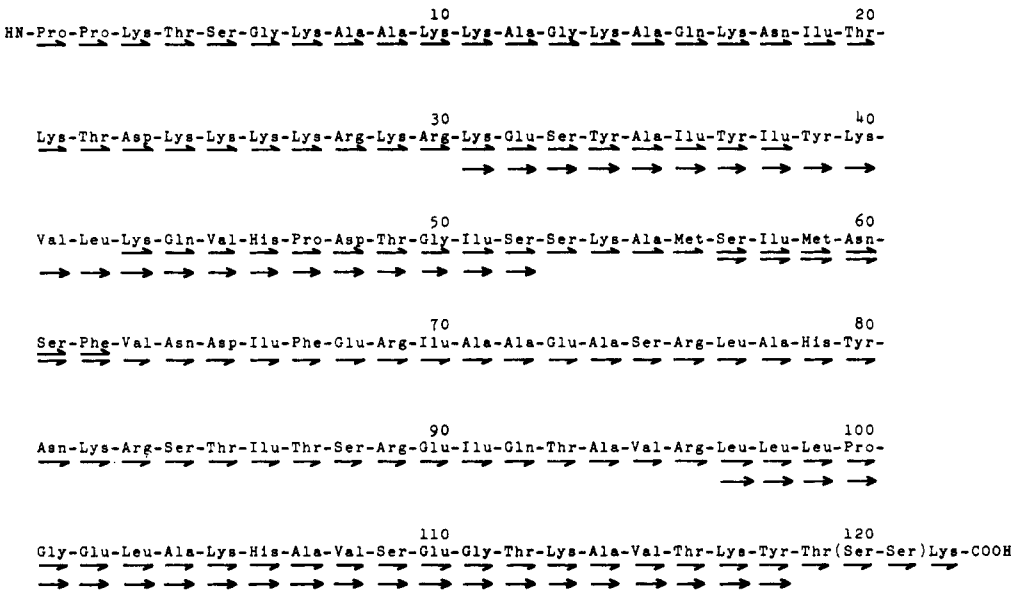

FIGURE 3: Summary of evidence for the sequence of *Drosophila* histone H2B. Residues determined by sequence analysis from the N-terminal residue are indicated by —; those determined by analysis of the cyanogen bromide peptides are indicated by —; those determined by analysis of the arginine peptides are indicated by →. Parentheses indicate tentative assignments. The C-terminal residue was identified as Lys by using carboxypeptidase digestion.

	10 20
Calf	HN-P-E-P-A-K-S-A-P---A-P-K-K---G-S-K-K-A-------V-T-K-A-Q-K-K-D-G-
Trout	HN-P-E-P-A-K-S-------A-P-K-K---G-S-K-K-A-------V-T-K-T-A-G-K-G-G-
Sea urchin	HN-A-P-T-A-Q---V-A-K-K---G-S-K-K-A-V---K-G-T-K-T-A-X---G-G-
Drosophila	HN-P-P-K-T-S-G-K-A-A-K-K-A-G-----K-A-Q-K-N-I-T-K-T-D---------
Patella	HN-P-P-K-V-S-S-K-G-A-K-K-A-G-----K-A-----K-A-A-R-S-G-D-------

	30 40 50
Calf	K-K-R-K-R-S-R-K-E-S-Y-S-V-Y-V-Y-K-V-L-K-Q-V-H-P-D-T-G-I-S-S-K-A-
Trout	K-K-R-K-R-S-R-K-E-S-Y-A-I-Y-V-Y-K-V-L-K-Q-V-H-P-D-T-G-I-S-S-K-A-
Sea urchin	K-K-R-N-R-K-R-K-E-S-Y-G-I-Y-I-Y-K-V-L-K-Q-V-H-P-D-T-G-I-S-S-R-A-
Drosophila	K-K-K-K-R-K-R-K-E-S-Y-A-I-Y-I-Y-K-V-L-K-Q-V-H-P-D-T-G-I-S-S-K-A-
Patella	K-K-R-K-R-R-R-K-E- V
Sea urchin sperm	V-K-R-R-R-R-R-R-E-
Others	H V

	60 70 80 90
Calf	M-G-I-M-N-S-F-V-N-D-I-F-E-R-I-A-G-E-A-S-R-L-A-H-Y-N-K-R-S-T-I-T-
Trout	M-G-I-M-N-S-F-V-N-D-I-F-E-R-I-A-G-E-S-S-R-L-A-H-Y-N-K-R-S-T-I-T-
Sea urchin	M-V-I-M-N-S{F}V{N}D-I-F-E-R-I-A-G-E-S-S-R-L-A-Q-Y-N-K-K{S}T{I}S-
Drosophila	M-S-I-M-N-S-F-V-N-D-I-F-E-R-I-A-A-E-A-S-R-L-A-H-Y-N-K-R-S-T-I-T-
Others	T V V S Q T T A L S

	100 110 120
Calf	S-R-E-I-Q-T-A-V-R-L-L-L-P-G-E-L-A-K-H-A-V-S-E-G-T-K-A-V-T-K-Y-T-S-S-K-OH
Trout	S-R-E-I-Q-T-A-V-R-L-L-L-P-G-E-L-A-K-H-A-V-S-E-G-T-K-A-V-T-K-Y-T-S-S-K-OH
Sea urchin	S-R-E-I-Q-T-A-V-R-L-I-L-P-G-E-L-A-K-H-A-V-S-E-G-T-K-A-V-T-K-Y-T-T-S-K-OH
Drosophila	S-R-E-I-Q-T-A-V-R-L-L-L-P-G-E-L-A-K-H-A-V-S-E-G-T-K-A-V-T-K-Y-T-S-S-K-OH
Others	A (V)

FIGURE 4: Comparison of amino acid sequence data available for histone 2B. Sequences have been aligned to obtain maximum homology with the calf sequence. Numbering is according to the calf sequence. Complete sequences from the following references are presented: calf H2B, Iwai et al., 1972; trout (*Salmo trutta*) H2B, Koostra & Bailey, 1976, 1978; sea urchin (*S. purpuratis*) early embryo H2B (from DNA sequence), Sures et al., 1978; *D. melanogaster* H2B, this paper. The sequence for *P. granatina* H2B (van Helden et al., 1979) for residues 35–125 is identical with that of *Drosophila* with the exception of valine at position 54. Positions at which substitutions have occurred in the region 27–125 are underlined. The partial sequence of sea urchin sperm H2B is from Strickland et al. (1978) (see text). Data collectively presented as "others" for the region 27–125 amino acids indicate any additional substitutions noted in the sequences from the following: sea urchin (*P. miliaris*) H2B (from DNA sequencing), positions 20–122 (Birnstiel et al., 1977; Schaffner et al., 1978); sea urchin (*P. angulosus*) embryo H2B, positions 27–35 and 59–79 (Brandt & von Holt, 1978); sea urchin (*P. angulosis*) embryo H2B fraction III, positions 59–70, embryo H2B fraction V, positions 59–71, and somatic H2B, positions 59–80 (Brandt et al., 1979); rat chloroleukemia H2B, residues 121–125 (Martinage et al., 1976); mouse H2B, residues 73–79 (Franklin & Zweidler, 1977); calf minor H2B, residues 73–79 (Franklin & Zweidler, 1977). Parentheses indicate ambiguities or uncertainties in the data.

positions 27–35 is given in Figure 4. A comparison of the C-terminal 90 residues of the 5 sperm H2B proteins with calf H2B indicates no gaps and 20 substitutions, 18 of them conservative. The somatic histone 2B's of pea and tobacco have also been reported to be ~20% larger than that of calf (Gigot et al., 1976; Spiker & Isenberg, 1977). In an analysis of the tryptic peptides and BrCN cleavage fragments of pea H2B, Hayashi et al. (1977) were able to align the peptides with the

calf H2B sequence starting at position 35. No significant alignments were obtained in the N-terminal portion. Thus, in both these cases the additional residues must be part of the N-terminal region. Comparison of the C-terminal 90 residues (from positions 35 to 125) tentatively indicates 1 gap and 19 substitutions for pea H2B relative to calf H2B. Most of the substitutions are conservative. If the sequence gap at position 39/40 for pea H2B implied by the peptide data is substantiated in the complete sequence, one would wish to consider moving the boundary position to 40. Otherwise, the data available indicate that the boundary occurs at position 26/27, with the size and position of the block of basic residues starting at position 27 relatively conserved features (see Figure 4).

It is of interest to compare this result to that obtained for the other histones involved in the nucleosome core. Extensive data on H3 and H4 indicate that these proteins are highly conserved over the entire molecule. No sequence gaps and only five positions where substitutions have occurred have been observed in sequence studies of H3 from vertebrates, invertebrates, and plants [see von Holt et al. (1979), Schaffner et al. (1978), and Sures et al. (1978)]; no sequence gaps and only three positions where substitutions have occurred have been seen in these H4's [see von Holt et al. (1979)]. In both cases, no substitutions have been found in the N-terminal 40 amino acids.[1] In particular, we have examined the N-terminal sequence of *Drosophila* H3 to position 32 and found it to be identical with that of calf (S. C. R. Elgin, unpublished experiments). Histone 2A is more variable; 8 sequence gaps and 27 substitutions may be noted in a composite comparison of the sequences from calf, rat, mouse, chick, trout, sea urchin, and *Drosophila* (partial) [see von Holt et al. (1979), Laine et al. (1978), and R. Goldberg and D. S. Hogness (personal communication)]. Insertions and deletions are mostly clustered in the N- and C-terminal regions (residues 1–21 and 121–129); the substitutions appear randomly distributed. A recent study of the sequences of H2A fractions from wheat germ reports the N-terminal region to be variable in both size and sequence (Rodrigues et al., 1979). Thus, while histones H3 and H4 are conserved structures throughout, both H2B and H2A show size and sequence variability in the terminal portions of the molecule with a conserved central core.

It should be noted that while both the size and sequence of the N-terminal region of H2B vary, the overall amino acid composition of the region is relatively constant, with 26–32% basic amino acids for the somatic H2B's (data from calf, trout, *Drosophila*, *Patella*, and sea urchin embryo) and 38–41% basic amino acids for the sea urchin sperm H2B's. There is no conserved, readily identifiable pattern of spacing of either the basic residues or the prolines in the N-terminal region of the somatic 2B histones. These observations suggest that only the very broad structural characteristics predictable from amino acid composition (e.g., net charge) need be conserved for this portion of the protein to function properly. It is known that several of the lysine residues in the N-terminal region can be modified by acetylation. Analysis of the data in the trout system for all the histones has led Dixon et al. (1975) to suggest that sites of the type -X-Lys-X-, where X is glycine, alanine, serine, or threonine, or of the type -Lys-Arg-, -Arg-Lys-, or -Lys-Lys- are preferentially acetylated, although the secondary structure of the protein is no doubt important in limiting acetylation to the N-terminal region. In the composite data of Figure 4 for H2B, 31 of 35 lysines in the N-terminal region occur at such sites. While to some degree this must be a reflection of the amino acid composition of this region, the conservation of this type of lysine site suggests a functional requirement.

Functional Analysis of H2B. Several lines of evidence have suggested the presence of at least two separate functional domains in H2B (and in the other core histones). On the basis of nuclear magnetic resonance and optical–spectroscopic data, Bradbury and his colleagues (Bradbury & Rattle, 1972; Bradbury et al., 1972) have suggested that residues 31–102 constitute a relatively structured part of H2B, while residues 1–30 and 102–125 are primarily involved in DNA binding. It should be noted, however, that these DNA binding studies were carried out in the absence of a full complement of histones. More recent work supports the concept of a H2B structural core formed by residues 55–78 (Lilley et al., 1975; Tancredi et al., 1976). A similar study using proton magnetic resonance and circular dichroism to analyze the H2A–H2B complex indicates that residues 37–114 of H2B are involved in the stable tertiary structure that is formed (Moss et al., 1976). This region is essentially that which is highly conserved in amino acid sequence.

Results of this type have led several investigators to predict that the N-terminal regions of H2B (and those of the other core histones) would project as "tails" from the nucleosome core, available to bind to the associated DNA [e.g., Van Holde et al. (1974)]. Such a model is supported by the observation of Weintraub & Van Lente (1974) that trypsin digestion of chromatin, while degrading H1, H5, and the nonhistone chromosomal proteins completely, removes the N-terminal 20–30 residues from the smaller histones but leaves the core essentially intact. All four of the smaller histones are required to generate the trypsin-resistant complex, which is stable in the absence of DNA in 2 M NaCl (Weintraub et al., 1975).

Studies of histone interactions in solution show strong complex formation between H2B and H2A and H2B and H4 and a relatively weaker interaction between H2B and H3 (D'Anna & Isenberg, 1974; Spiker & Isenberg, 1977). More recently, Spiker & Isenberg (1978) have found that the pattern of pairwise histone–histone interactions is conserved for heterologous complexes of core histones from calf and pea, even though pea H2B is ~20% larger than calf H2B. These interkingdom complexes form with essentially the same binding constants as the regular complexes. Spiker & Isenberg (1978) estimate (given the error limits of their data) that at most one or two residue changes may have occurred at the binding surfaces between two interacting histones. Given the results obtained by protein sequencing, these findings imply that the C-terminal region of H2B is involved in histone–histone interactions, while the N-terminal region is not. Direct studies using chemical cross-linking indicate that the middle one-third of H2B interacts closely with H2A and the C-terminal one-third of H2B interacts closely with H4 (Martinson et al., 1979; DeLange et al., 1979). Similar cross-linking results have been obtained by using chromatin from either plants or animals (Martinson & True, 1979).

Thus, the data suggest that the two very different regions of H2B identified by comparative sequence analysis reflect two very different functional domains of the protein. There is a conserved core of histone 2B (residues 35–125) which is both necessary and sufficient to generate the stable tertiary

[1] Glover & Gorovsky (1979) have recently found that H4 from a protozoa, *Tetrahymena thermophil*, differs substantially from that of calf. In the first 66 residues there are 2 gaps and 13 substitutions. This H4 behaves differently from others in several functional assays, complexing less strongly than calf H4 with both calf and *Tetrahymena* H2B and H3 (Glover & Gorovsky, 1978); consequently, it has not been included in this analysis.

structure of this protein in a complex(es) with histones 2A, 3, and 4. The more basic N-terminal domain 1–26 binds to DNA in a manner requiring considerably less structural specificity. A run of five to eight basic residues occurs at the end of the N-terminal domain; the significance of this cluster is unknown. Why H2B differs from the core histones H3 and H4 in its pattern of conserved structure remains to be resolved. The case of H2A resembles that of H2B. While current evidence suggests that the H3–H4 tetramer is necessary and sufficient for beginning DNA folding (Camerini-Otero et al., 1976; Sollner-Webb et al., 1976; Moss et al., 1977), direct tests using antibodies against H2B indicate that this histone is present in most, if not all, nucleosomes (Bustin et al., 1976; Simpson & Bustin, 1976). Studies on the reactive properties of the histones in chromatin have indicated that the N-terminal region of H2B is relatively exposed [e.g., Malchy (1977)]. A functional analysis of the different constraints placed on the N-terminal domains of the different core proteins must await further data on chromatin structure.

References

Alfageme, C. R., Zweidler, A., Mahowald, A., & Cohen, L. H. (1974) *J. Biol. Chem. 249*, 3729–3736.

Birnstiel, M. L., Schaffner, W., & Smith, H. O. (1977) *Nature (London) 266*, 603–607.

Bradbury, E. M., & Rattle, H. W. E. (1972) *Eur. J. Biochem. 27*, 270–281.

Bradbury, E. M., Carey, P. D., Crane-Robinson, C., Riches, P. L., & Johns, E. W. (1972) *Eur. J. Biochem. 26*, 482–489.

Brandt, W. F., & von Holt, C. (1978) *Biochim. Biophys. Acta 537*, 177–181.

Brandt, W. F., Strickland, W. N., Strickland, M., Carlisle, L., Woods, D., & von Holt, C. (1979) *Eur. J. Biochem. 94*, 1–10.

Bustin, M., Goldblatt, D., & Sperling, R. (1976) *Cell 7*, 297–304.

Camerini-Otero, R. D., Sollner-Webb, B., & Felsenfeld, G. (1976) *Cell 8*, 333–347.

D'Anna, J. A., Jr., & Isenberg, I. (1974) *Biochemistry 13*, 4992–4997.

DeLange, R. J., Fambrough, D. M., Smith, E., & Bonner, J. (1969) *J. Biol. Chem. 244*, 5669–5679.

DeLange, R. J., Williams, L. C., & Martinson, H. G. (1979) *Biochemistry 18*, 1942–1946.

Dixon, G. H., Candido, E. P. M., Honda, B. M., Louie, A. J., Macleod, A. R., & Sung, M. T. (1975) *Ciba Found. Symp. 28*, 229–250.

Elgin, S. C. R., & Hood, L. E. (1973) *Biochemistry 12*, 4984–4991.

Elgin, S. C. R., & Weintraub, H. (1975) *Annu. Rev. Biochem. 44*, 725–774.

Fambrough, D. M., & Bonner, J. (1968) *J. Biol. Chem. 243*, 4434–4439.

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

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

Gigot, C., Phillipps, G., Nicolaieff, A., & Hirth, L. (1976) *Nucleic Acids Res. 3*, 2315–2329.

Glover, C. V. C., & Gorovsky, M. A. (1978) *Biochemistry 17*, 5705–5713.

Glover, C. V. C., & Gorovsky, M. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 585–589.

Hayashi, H., Iwai, K., Johnson, J. D., & Bonner, J. (1977) *J. Biochem. (Tokyo) 82*, 503–510.

Hunkapiller, M. W., & Hood, L. E. (1978) *Biochemistry 17*, 2124–2133.

Iwai, K., Hayashi, H., & Ishikawa, K. (1972) *J. Biochem. (Tokyo) 72*, 357–367.

Kootstra, A., & Bailey, G. S. (1976) *FEBS Lett. 68*, 76–78.

Kootstra, A., & Bailey, G. S. (1978) *Biochemistry 17*, 2504–2510.

Kornberg, R. D. (1977) *Annu. Rev. Biochem. 46*, 931–954.

Laine, B., Kmiecik, D., Sautiere, P., & Biserte, G. (1978) *Biochimie 60*, 147–150.

Lifton, R. P., Goldberg, M. L., Karp, R. W., & Hogness, D. S. (1978) *Cold Spring Harbor Symp. Quant. Biol. 42*, 1047–1051.

Lilley, D. M. J., Howarth, W. W., Clark, V. M., Pardon, J. F., & Richards, B. M. (1975) *Biochemistry 14*, 4590–4600.

Malchy, B. L. (1977) *Biochemistry 16*, 3922–3927.

Martinage, A., Sautiere, P., Kerckaert, J., & Bisertz, G. (1976) *Biochim. Biophys. Acta 420*, 37–41.

Martinson, H. G., & True, R. J. (1979) *Biochemistry 18*, 1947–1951.

Martinson, H. G., True, R. J., Lau, C. K., & Mehrabian, M. (1979) *Biochemistry 18*, 1075–1082.

Moss, T., Cary, P. D., Abercrombie, B. D., Crane-Robinson, C., & Bradbury, E. M. (1976) *Eur. J. Biochem. 71*, 337–350.

Moss, T., Stephens, R. M., Crane-Robinson, C., & Bradbury, E. M. (1977) *Nucleic Acids Res. 4*, 2477–2485.

Oliver, D., & Chalkley, R. (1972) *Exp. Cell Res. 73*, 295–302.

Oliver, D., Sommer, K. R., Panyim, S., Spiker, S., Chalkley, R. (1972) *Biochem. J. 129*, 349–353.

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

Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H. O., & Birnstiel, M. L. (1978) *Cell 14*, 655–672.

Simpson, R. T., & Bustin, M. (1976) *Biochemistry 15*, 4305–4311.

Sollner-Webb, B., Camerini-Otero, R. D., & Felsenfeld, G. (1976) *Cell 9*, 179–193.

Spiker, S., & Isenberg, I. (1977) *Biochemistry 16*, 1819–1826.

Spiker, S., & Isenberg, I. (1978) *Cold Spring Harbor Symp. Quant. Biol. 42*, 157–164.

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

Strickland, W. N., Strickland, M., Brandt, W. F., & von Holt, C. (1977b) *Eur. J. Biochem. 77*, 277–286.

Strickland, M., Strickland, W. N., Brandt, W. F., & von Holt, C. (1978) *Biochim. Biophys. Acta 536*, 289–297.

Sures, I., Lowry, J., & Kedes, L. H. (1978) *Cell 15*, 1033–1044.

Tancredi, T., Temussi, P. A., Paolillo, L., Trivellone, E., Crane-Robinson, C., & Bradbury, E. M. (1976) *Eur. J. Biochem. 70*, 403–408.

van Helden, P. D., Strickland, W. N., Brandt, W. F., & von Holt, C. (1979) *Eur. J. Biochem. 93*, 71–78.

Van Holde, K. E., Sahasrabuddhe, C. G., & Shaw, B. R. (1974) *Nucleic Acids Res. 1*, 1579–1586.

von Holt, C., Strickland, W., Brandt, W. F., & Strickland, M. S. (1979) *FEBS Lett. 100*, 201–218.

Weintraub, H., & Van Lente, F. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4249–4253.

Weintraub, H., Palter, K., & Van Lente, F. (1975) *Cell 6*, 85–110.

Woodcock, C. L. F., Safer, J. P., & Stanchfield, J. E. (1976) *Exp. Cell Res. 97*, 101–110.