8-1988

# Drosophila has a single copy of the gene encoding a highly conserved histone H2A variant of the H2A.F/Z type

Angela van Daal
*Washington University in St. Louis*

E White

M Gorovsky

Sarah C.R. Elgin
*Washington University in St. Louis*, selgin@wustl.edu

# *Drosophila* has a single copy of the gene encoding a highly conserved histone H2A variant of the H2A.F/Z type

Angela van Daal, Elizabeth M.White[1], Martin A.Gorovsky[1] and Sarah C.R.Elgin

Department of Biology, Washington University, St Louis, MO 63130 and [1]Department of Biology, University of Rochester, Rochester, NY 14627, USA

Received April 12, 1988; Revised and Accepted June 21, 1988 Accession no. X07485

ABSTRACT

**The *Tetrahymena* histone H2A variant designated hv1 is localized exclusively in the transcriptionally active macronucleus and is absent from the quiescent micronucleus (1). A cDNA clone of the hv1 gene (2) was used to screen a *Drosophila* cDNA library. A cross-hybridizing clone was recovered and shown by sequence analysis to code for a protein homologous to hv1 as well as to the chicken H2A variant, H2A.F (3), the sea urchin H2A variant, H2A.F/Z (4) and the mammalian H2A variant H2A.Z (5). Southern analysis of *Drosophila* genomic DNA indicates that the H2AvD (H2A variant *Drosophila*) gene is present in one copy. *In situ* hybridization places the locus at 97CD on chromosome 3, while the S-phase regulated histone genes are on chromosome 2 (6). Thus the *Drosophila* H2A variant should be accessible to genetic analysis, which will enable its function to be determined.**

## INTRODUCTION

Since the development of the nucleosome model of chromatin structure, questions have arisen as to whether or not all nucleosomes are structurally equivalent. In addition to those histone genes frequently present in multiple identical copies, whose expression is linked to the S-phase of the cell cycle, variants of histones H2A, H2B and H3 have been identified which are not so regulated. These non-allelic variants of the nucleosome core histones, which differ in primary sequence from their counterparts, have been reported in a wide variety of organisms (7). The variant sequences are frequently highly conserved, the variants being more similar to one another than each variant is to its own major histone counterpart (4). The majority of the core histone variants are synthesized throughout

the cell cycle (8) and their gene structure characteristically differs from that of the S-phase regulated histone genes in that the coding sequences are interrupted by introns and the mRNAs are polyadenylated (9,10). There has been much speculation that the utilisation of such variants could alter the fundamental properties of the nucleosome, with profound biological implications.

cDNA clones for H2A variants from a variety of organisms have been identified and characterized. They all have long 3' untranslated regions and are polyadenylated. The first H2A variant characterized at the DNA level was that of the chicken, H2A.F (3). The protein encoded by this cDNA sequence was only 60% similar to the S-phase regulated H2A of chicken. In addition, the mRNA had a limited tissue distribution in the chicken embryo. H2A.F was identified as an H2A variant by the presence of the 9 amino acid peptide (amino acids 23-31) known as the H2A box, which is conserved in all H2A proteins examined to date. A sea urchin cDNA recently isolated has been shown to contain an open reading frame which is 74% similar to that of the chicken H2A.F cDNA coding region (4). The similarity of these two sequences is much more dramatic at the amino acid level. The two proteins derived by conceptual translation of the cDNA sequences are 97% similar with a total of only five amino acid changes. Three of these changes occur in the last five amino acids. This carboxyl terminal region is highly divergent in the S-phase regulated H2A and is thought not to be involved in the histone/histone interactions necessary to generate the octamer (11). The message for the sea urchin variant, H2A.F/Z, was found to be present throughout embryogenesis. However there was a dramatic increase in the level of mRNA at the mesenchyme blastula stage (4). The cDNA sequences of a mammalian H2A variant, H2A.Z, from three species (human, calf and rat) have also been reported (5). These variants are identical to each other at the amino acid level and show extremely high similarity at the DNA sequence level (94-97%). Interestingly, the degree of similarity at the DNA sequence level is greater in the 3'untranslated region (98%) than it is in the coding region (94-97%). However, there is no detectable similarity between the DNA sequences of the mammalian variants

and the chicken or sea urchin variants outside of the coding region, suggesting that any control mechanisms dictated by the 3' untranslated region which are common to the mammalian H2A variants have arisen as part of the evolution of mammals.

Perhaps the most interesting H2A variant is the *Tetrahymena* hv1 protein. hv1 is found only in the transcriptionally active macronucleus and is absent from the transcriptionally inert micronucleus (1). It therefore appears to be associated with active chromatin. A cDNA clone for hv1 has been isolated and sequenced (2). The derived protein sequence is 84% similar to the chicken and sea urchin variants in the central region (amino acids 18-120). The amino and carboxy terminal portions, however, have diverged and show no sequence similarity. The extremely high degree of conservation of the protein coding sequences amongst these widely diverse species, however, suggests an important biological function for this protein. The localization of the H2A variant in *Tetrahymena* to the transcriptionally active macronucleus suggests that these H2A variants may play a role specific to the euchromatic structure, perhaps differentiating between active and inactive genes. The isolation of this variant in a genetically well characterised organism such as *Drosophila* should allow its function to be clearly demonstrated for the first time.

## MATERIALS AND METHODS

### Screening of the cDNA Library

A *Drosophila* cDNA library (12) constructed with polyA$^+$ mRNA from 0-3 hour *Drosophila* embryos in λgt10 was screened with the 666bp insert of the *Tetrahymena* hv1 cDNA clone. The probe was labelled by nick translation with $\alpha$-$^{32}$P dATP, dCTP, dGTP and dTTP, and purified through Sephadex G-50. Filters were prehybridized at 55°C in 5xSSC, 0.2% SDS, 1% powdered milk for 3-4 hours. Hybridization was carried out in 5xSSC, 0.2% SDS, 1% powdered milk with the radioactive probe at a final concentration of approximately 2x10$^6$cpm/ml. Filters were washed in 2xSSC, 0.2% SDS two times at room temperature followed by two times at 55°C.

### Sequencing Strategy

The inserts of two putative H2A variant cDNA clones, $9_1$ and

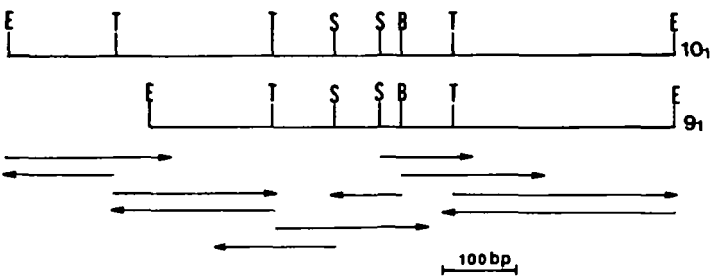


**Figure 1 Sequencing Strategy of CDNA Clones**
The EcoRI inserts of cDNA clones $9_1$ and $10_1$ were sequenced as indicated by the arrows. The leftward and rightward pointing arrows indicate sequencing of the non-coding and coding strands respectively. E, S, T and B represent EcoRI, Sau3A, TaqI and BamHI respectively.

$10_1$, were isolated from the EcoRI site of λgt10. The insert fragments were digested with Sau3A, TaqI or PstI and the internal fragments were subcloned into the BamHI, AccI or PstI site of M13mp18 or M13mp19. The end fragments were cloned into both M13mp18 and M13mp19 digested with EcoRI and BamHI, AccI or PstI. Both strands of all subfragments from both clones were completely sequenced by the dideoxy chain termination method (see Figure 1). One clone, $9_1$, was completely contained in the second longer clone, $10_1$.

**Southern Analysis**

*Drosophila* 6-18 hour embryo DNA was digested with BglII, EcoRI or HindIII, electrophoresed on 0.5% agarose and transferred to nitrocellulose. The filter was hybridised with clone $10_1$ DNA, which had been labelled by nick translation, at 65°C in 5XSSC, 0.1% SDS, 1% powdered milk and 100ug/ml salmon sperm DNA. The filter was washed at 65°C with 0.2XSSC, 0.1% SDS. Lambda DNA digested with BamHI and EcoRI was used to generate molecular weight markers.

***In situ* Hybridisation**

Polytene chromosomes were prepared from the salivary glands of third instar Oregon-R *Drosophila* larvae and fixed in 45% acetic acid. The genomic clone $5_1$G, which contains the entire $10_1$ sequence, was labelled by nick translation using biotin-UTP 11-mer from Bethesda Research Laboratories (Gaithersburg, Md.). The

hybridisation was done in 45% formamide, 26% dextran sulphate, 5XSSC, 2mg/ml salmon sperm DNA at 37°C for 16 hours and hybrids were visualised using streptavidin conjugated alkaline phosphatase according to the procedure of the Clontech Laboratories (Palo Alto, Ca.) staining kit.

## RESULTS AND DISCUSSION

### Isolation and sequence analysis of H2AvD cDNA clone

An hv1 cDNA clone (2) was used to screen a *Drosophila* cDNA library made from 0-3 hour embryo poly $A^+$ mRNA (12). The major S-phase regulated histone gene transcripts are not polyadenylated and should therefore not be represented in this library. Two clones, designated $9_1$ and $10_1$, were isolated. The sequence of $9_1$ was contained within the sequence of $10_1$. Both clones contained the same 3' end, but $9_1$ did not extend as far 5' as $10_1$ and, in fact, did not contain the entire coding region. The sequence of clone $10_1$ is shown in Figure 2. The cDNA sequence contains 169 nucleotides of 5' untranslated sequence and a 3' untranslated region of 339 nucleotides, including two polyadenylation signals separated by three nucleotides at nucleotide positions 859 and 868. The sequence also contains a 31 nucleotide long poly A tail.

### Comparison of H2AvD and H2A.1

Figure 3 shows a comparison of the deduced protein sequence of H2AvD (second line) with the sequence of the *Drosophila* H2A encoded in the repetitious gene cluster at 39D2-E2 (6), which we will designate H2A.1 (first line). The H2AvD sequence contains the nonapeptide, which is conserved in all H2A's, at position 23-31, indicating that this cDNA encodes an H2A-like protein. However, it is clear that the sequence differs markedly from that of the H2A encoded in the *Drosophila* histone gene cluster. There are 52 amino acid changes in the first 124 amino acids, which indicates a similarity of only 59%. This is approximately the same degree of similarity shown by the H2A variants of chicken (60%), sea urchin (56%-57%), mammals (56%) and *Tetrahymena* (62-65%) relative to their respective major S-phase regulated H2A's.

The cDNA sequence predicts that H2AvD is a larger molecular weight protein than H2A.1. An antibody to the carboxy terminal

```
                                                GAATTCCGTAGAAACAACTCGCACGCTCCGG  31

TTTCGTGTTGCAACAAAACAGGCATTCCCATCGCGGCAGTTAGAATCACCGAGTCGCCC  90

AGAGTCACGTTCGTAAGCAGGCGCAGTTTACAGGCAGCAGAAAAATCGATTGAACAGAA 169

ATG GCT GGC GGT AAA GCA GGC AAG GAT TCG GGC AAG GCC AAG GCG 196
    ala gly gly lys ala gly lys asp ser gly lys ala lys ala

AAG GCG GTA TCG CGT TCC GCG CGC GCG GGT CTT CAG TTC CCC GTG 239
lys ala val ser arg ser ala arg ala gly leu gln phe pro val

GGT CGC ATC CAT CGT CAT CTC AAG AGC CGC ACT ACG TCA CAT GGA 284
gly arg ile his arg his leu lys ser arg thr thr ser his gly

CGC GTC GGA GCC ACT GCA GCC GTG TAC TCC GCT GCC ATA TTG GAA 329
arg val gly ala thr ala ala val tyr ser ala ala ile leu glu

TAC CTG ACC GCC GAG GTC CTG GAG TTG GCA GGC AAC GCA TCG AAG 374
tyr leu thr ala glu val leu glu leu ala gly asn ala ser lys

GAC TTG AAA GTG AAA CGT ATC ACT CCT CGC CAC TTA CAG CTC GCC 419
asp leu lys val lys arg ile thr pro arg his leu gln leu ala

ATT CGC GGA GAC GAG GAG CTG GAC AGC CTG ATC AAG GCA ACC ATC 464
ile arg gly asp glu glu leu asp ser leu ile lys ala thr ile

GCT GGT GGC GGT GTC ATT CCG CAC ATA CAC AAG TCG CTG ATC GGC 509
ala gly gly gly val ile pro his ile his lys ser leu ile gly

AAG AAG GAG GAA ACG GTG CAG GAT CCG CAG CGG AAG GGC AAC GTC 554
lys lys glu glu thr val gln asp pro gln arg lys gly asn val

ATT CTG TCG CAG GCC TAC TAA GCCAGTCGGCAATCGGACGCCTTCGAAACAT 606
ile leu ser gln ala tyr

GCAACACTAATGTTTAATTCAGATTTCAGCAGAGACAAGCTAAACACGACGAGTTGTAA 665

TCATTTCTGTGCGCCAGATATATTTCTTATATACAACGTAATACATAATTATGTAATTC 724

TAGCATCTCCCCAACACTCACATACATACAAACAAAAAATACAAACACACAAAACGTAT 783

TTACCCGCACGCATCCTTGGCGAGGTTGAGTATGAAACAAAAACAAAACTTAATTTAGA 842

GCAAAGTAATTACACGAATAAATTTAATAAAAAAAACTATAAAAAAAAAAAAAGAAAAAA 901

AAAAAAAAAAAAG
```

**Figure 2 DNA Sequence of H2AvD cDNA Clone $10_1$**

Both strands of *Drosophila* cDNA clones $9_1$ and $10_1$ were sequenced by the dideoxy chain termination method of Sanger *et al.*(15). Sequence within the open reading frame is printed in triplets. Note the polyadenylation signals at 859 and 868.

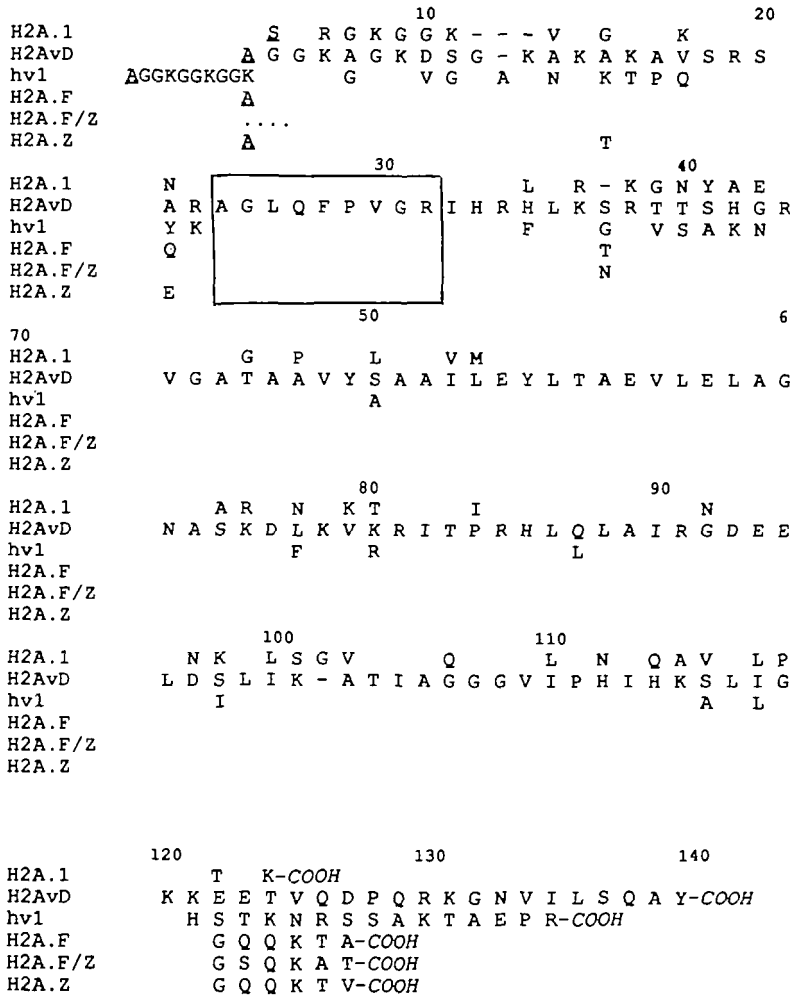


**Figure 3 Comparison of Protein Sequences of *Drosophila*, Chicken, Sea Urchin and *Tetrahymena* H2A Variants and *Drosophila* H2A.1**

The deduced H2AvD protein sequence is shown and compared with the sequences of the Drosophila H2A.1, chicken H2A.F (3), sea urchin H2A.F/Z (4), mammalian H2A.Z (5) and *Tetrahymena* hv1 (2). The sequence of *Drosophila* H2A.1 is from the cDm500 clone (16). Unless otherwise indicated, the sequences are the same as that of H2AvD. In all cases the sequences are derived by conceptual translation from the cDNA or genomic clones. The first seven amino acids of hv1 are derived from the protein sequence (17). The N-terminal amino acid is underlined. The H2A.F/Z cDNA clone was not full length, so that the N-terminal amino acid is not known. Note that all proteins have a highly conserved region at amino acids 23-31 (boxed sequence).

portion of H2AvD detects a protein of lower mobility than the major histone H2A on electrophoresis in SDS polyacrylamide gels (A. van Daal, unpublished results). An H2A variant (D2) has been previously reported in *Drosophila*(13). D2 has a molecular weight of 13,400 daltons and differs from H2A.1 in that it contains no methionine residues and has an increased histidine content. H2AvD shares these properties.

**Comparison of H2AvD and hv1**

A comparison of H2AvD with the *Tetrahymena* H2A variant hv1 is

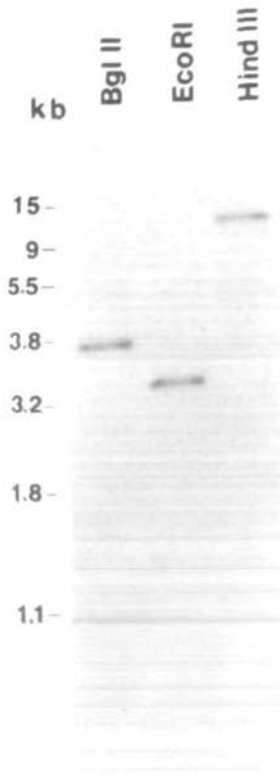


**Figure 4 Genomic Southern Analysis of the H2AvD Gene**
10ug of Drosophila embryo DNA was digested with BglII, EcoRI and HindIII, electrophoresed on 0.5% agarose, transferred to nitrocellulose and probed with the $\lambda gt_{10}10_1$ cDNA clone.

shown in the second and third lines of Figure 3. The amino and carboxy terminal ends of the *Drosophila* and *Tetrahymena* H2A variants have diverged. However, the body of the protein shows remarkably high similarity. There are only 16 amino acid changes in the region of amino acids 18-120, which gives a similarity of 84%. The carboxy terminal end of hv1 (from amino acid 121) is completely different from that of H2AvD in both amino acid composition and length. The N-terminus shows some sequence similarity (less than 50%), but again the length is different.

It should be noted that the major *Tetrahymena* histone H2A differs significantly from the major H2A's of the other species. The H2A.1's of human, cow, rat and chicken are 95-99% similar to each other. The *Drosophila* and sea urchin H2A.1's are 81-87% similar to those of mammals, chicken and each other. However, the


**Figure 5** ***In Situ*** **Hybridisation of an H2AvD Genomic Clone to** ***Drosophila*** **Polytene Chromosomes.**
Polytene chromosomes of third instar larvae were fixed in 45% acetic acid, squashed, denatured and probed with biotin-labelled $5_1$G. Hybridisation was visualised using streptavidin conjugated alkaline phosphatase.

*Tetrahymena* H2A.1 is only 63-69% similar to those of other species, and so it is perhaps not surprising that H2AvD is quite different from hv1, whereas, it is very similar (97%) to the sea urchin variant.

**Comparison of H2A variant proteins**

Figure 3 also shows a comparison of the deduced protein sequences of the H2A variants of chicken (H2A.F), sea urchin (H2A.F/Z) and mammals (H2A.Z) with that of *Drosophila*. From the comparison of H2AvD with the H2A variants of these other species, it is clear that the amino acid sequence of H2AvD is more closely related to the chicken, sea urchin and mammalian H2A variants than it is to the major *Drosophila* H2A, H2A.1. The conservation in the first 122 amino acids is extremely high. There is 99% similarity between the *Drosophila* and the sea urchin proteins, with only one amino acid change (ser to asn at position 38). There is 98% similarity between H2AvD and the chicken and mammalian proteins. Both have two amino acid changes (ala to gln at position 21 and ser to thr at position 38 in chicken and ala to thr at position 14 and ala to glu at position 21 in mammals). The carboxyl ends of these four variants all differ. However, the *Drosophila* variant is the most divergent. The C terminal tail is longer than the others and its sequence is also dissimilar.

**Southern Analysis**

Southern blot analysis of *Drosophila* genomic DNA with clone $10_1$ indicates that the H2AvD gene is unique (Figure 4). $10_1$ hybridizes to one 3.8kb BglII fragment, one 3.4kb EcoRI fragment, and one 10.6kb HindIII fragment. The cDNA clone has been used to screen a genomic library (14) and two overlapping clones, $4_1$G and $5_1$G, were isolated. A total of eight genomic clones were recovered after screening seven genome equivalents of DNA, supporting the hypothesis that H2AvD is indeed a unique gene.

***In situ* Hybridisation**

One of the genomic clones ($5_1$G) was used to locate the gene on the polytene chromosomes of third instar larvae (Figure 5). $5_1$G DNA was labelled by nick translation and hybridised to denatured chromosomes. The H2AvD locus was mapped to 97CD, towards the end of chromosome arm 3R. The *in situ* hybridisation reveals a unique location, supporting the finding of only one copy of the gene. The histone gene cluster is located at 39D2-E2

(6), making it genetically separable from the H2AvD locus. This opens up, for the first time, the exciting possibility of determining the role of a histone variant in chromatin structure by genetic manipulation.

**Acknowledgements**

This work was supported by grants from the NIH to S.C.R.E and to M.A.G. We are grateful to Ian Duncan and Elizabeth Burgess for their help in the *in situ* localisation of the H2AvD gene.

REFERENCES

1. Allis, C.D., Glover, C.V.C., Bowen, J.K. and Gorovsky, M.A. (1980). Cell **20**, 609-617.
2. White, E., Shapiro, D.L., Allis, C.D. and Gorovsky, M.A. (1988). Nucl. Acids Res. **16**, 179-198.
3. Harvey, R.P., Whiting, J.A., Coles, L.S., Krieg, P.A. and Wells, J.R.E. (1983). Proc. Natl. Acad. Sci. USA **80**, 2819-2823.
4. Ernst, S.G., Miller, H., Brenner, C.A., Nocenta-McGrath, C., Francis, S. and McIsaac, R. (1987). Nucl. Acids Res. **15**, 4629-4644.
5. Hatch, C.L. and Bonner, W.L. (1988). Nucl. Acids Res. **16**, 1113-1124.
6. Pardue, M.L., Kedes, L.H., Weinberg, E.S. and Birnstiel, M.L. (1977). Chromosoma **63**, 135-151.
7. Stein, G.S., Stein, J.L. and Marzluff, W.F. eds. (1984). Histone Genes: Structure, Organization and Regulation. John Wiley and Sons, New York.
8. Wu, R.S. and Bonner, W.M. (1981). Cell **27**, 321-330.
9. Brush, D., Dodgson, J.B., Choi, O.R., Stevens, P.W. and Engel, J.D. (1985). Mol. Cell. Biol. **5**, 1307-1317.
10. Wells, D. and Kedes, L. (1985). Proc. Natl. Acad. Sci. USA **82**, 2834-2838.
11. Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984). Nature **311**, 532-537.
12. Poole, S.J., Kavar, L.M., Drees, B. and Kornberg, T. (1985). Cell **40**, 37-43.
13. Palmer, D., Snyder, L. and Blumenfeld, M. (1980). Proc. Natl. Acad. Sci. USA **77**, 2671-2675.
14. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978). Cell **15**, 687-701.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
16. Goldberg, M.L. (1979). PhD. Thesis "Sequence Analysis of *Drosophila* Histone Genes", Stanford University.
17. Allis,C.D., Richman, R., Gorovsky, M.A., Ziegler, Y.S., Touchstone, B., Bradley, W.A and Cook, R.G. (1986). J.B.C. **261**, 1941-1948.