The sequence of the Drosophila regulatory gene *Suppressor two of zeste*

Brian P.Brunk and Paul N.Adler
Biology Department, Molecular Biology Institute and Cancer Center, University of Virginia, Charlottesville, VA 22901, USA

Submitted April 26, 1991 EMBL accession no. X56798

The product of the *zeste (z)* gene is a candidate for being a mediator of synapsis dependent gene expression phenomena in Drosophila (1–8). In a fly mutant for the neomorphic z^1 mutation, the expression of the *white (w)* gene is inhibited if there are two copies of the *w* gene that are in close enough proximity to somatically pair (4, 7, 8). The *z* gene encodes a DNA binding protein (9, 10) that has been shown to act as a transcription factor *in vitro* (11). Mutations have been recovered in a number of genes that are dominant modifiers of the z^1-w interaction (12, 13). Interestingly, a number of these have also been found to result in homeotic transformations (14, 15) that presumably arise due to misexpression of the homeotic selector genes of the BX-C and/or Antennapedia Complex (Antp-C). Prehaps the best characterized of these genes is the *Suppressor 2 of zeste (Su(z)2)* gene (12, 14–17), which appears to be a pleiotropic regulatory gene. There are mutant alleles (in most cases gain of function alleles) of this locus that can suppress the z^1-w interaction (23, 14, 15), affect sensory bristle development (16, 17), give rise to a homeotic transformation of arista (distal antenna) to tarsus (distal leg) (16), or result in lethality (15–17). We report here the sequence of an 11 kb segment of the second chromosome that contains the *Su(z)2* gene. Via comparison with *Su(z)2* cDNA clones (18) we conclude that the gene is comprised of six exons which nearly span the entire segment of genomic DNA. The first and the last exon are entirely non-coding.

The *Arp* mutation is a P element mediated inversion that results in a gain of function mutation in the *Su(z)2* gene (16) due to the overexpression of the *Su(z)2* mRNA (17). Previous genomic Southern analyses indicated that the distal breakpoint of this inversion was just upstream of the *Su(z)2* transcription unit (16). We sequenced cloned mutant *Arp* DNA (Figure 1) and found that the P element that mediated the distal breakpoint is inserted at nucleotide 33. Since the *Arp* inversion does not result in a loss of function mutation in *Su(z)2* (16), this nucleotide marks the upstream boundary for any essential *Su(z)2* sequences. The longest cDNA clone starts at nucleotide 160, therefore, at most 127 bp of upstream regulatory sequence are essential for *Su(z)2* function. There are many genes where regulatory information lies within the transcription unit itself (i.e. *sevenless* (19), *troponin I* (20)). Such genes have commonly been found to contain large first or second introns, as is the case with *Su(z)2*. Consistent with the hypothesis that the *Su(z)2* gene may contain regulatory information in such a location is the presence of many repeated sequences (both direct and inverted) in the first intron of *Su(z)2*. Clustered repeated sequences have been shown to be common in regulatory regions and often represent binding sites for trans-regulatory proteins (reviewed in 21).

ACKNOWLEDGEMENTS

Special thanks are due Elizabeth Martin, Charles Vinson, Charlie Emerson, Mary Beth Davis, and Ann Beyer for helpful discussions during the course of the work and comments on the manuscript. Thanks go to Jeannette Charlton, Sharon Conover, and Matthew Allison for technical help, and to Cliff Bishop for isolating the intial *Su(z)2* cDNA clones. This work was supported by National Science Foundation Grants DCB 8502683 and DCB 8812076. BPB was supported by an NIH predoctoral training grant in developmental biology (HDO-7192).

REFERENCES

1. Gelbart,W.M. and Wu,C.-t.Wu (1982) *Genetics* **102**, 179.
2. Jack,J.W. and Judd,B.H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1368.
3. Bingham,P.M. (1980) *Genetics* **95**, 341.
4. Gubb,D., Ashburner,M., Roote,J. and Davis,T. (1990) *Genetics* **126**, 167.
5. Kaufman,T.C., Tasaka,S.E. and Suzuki,D.T. (1973) *Genetics* **75**, 199.
6. Babu,P. and Bhat,S. (1981), *Mol. Gen. Genet.* **183**, 400.
7. Gans,M. (1953) *Bull. Biol. Fr. Belg.* **38**, 1.
8. Green,M.M. (1967) *Biol. Zentraub. (suppl)* **86**, 211.
9. Benson,M. and Pirrotta,V. (1987) *EMBO J.* **6**, 1387.
10. Mansukhani,A., Crickmore,A., Sherwood,P.W. and Goldberg,M.L. (1988) *Mol. Cell. Biol.* **8**, 615.
11. Biggin,M.D., Bickel,S., Benson,M., Pirrotta,V. and Tijan,R. (1988), *Cell* **53**, 713.
12. Kalisch,W.-E. and Rasmuson,B. (1974) *Hereditas* **78**, 97.
13. Persson,K. (1976) *Hereditas* **82**, 111.
14. Wu,C.-t., Jones,R.S., Lasko,P.F. and Gelbart,W.M. (1989) *Mol. Gen. Genet.* **218**, 559.
15. Wu,C.-t. (1984) Ph.D. dissertation, Harvard.
16. Brunk,B.P. and Adler,P.N. (1990) *Genetics* **142**, 145.
17. Brunk,B.P., Martin,E.C. and Adler,P.N. (1991) *Genetics* **128**, 119.
18. Brunk,B.P. (1989) Ph.D. Dissertation, University of Virginia.
19. Basler,K., Siegrist,P. and Hafen,E. (1989) *EMBO J.* **8**, 2381.
20. Konieczny,S.F. and Emerson,C.P. (1987) *Mol. Cell. Biol.* **7**, 3065.
21. McKnight,S. and Tijan,R. (1986) *Cell* **46**, 795.

```
          20        30        40        50        60        70

wt  GAAAATGCGCCCGCGCAAAAGTAGAAAGTTTGTTTGTCGTCTCTCTCGCCGCGGCGAA
      : :           : ::::::::::::::::::::::::::::::::::::::
Arp CCACCTTATGTTATTTCATGGTAGAAAGTTTGTTTGTCGTCTCTCTCGCCGCGGCGAA
    ::::::::::::::::::::
P   CCACCTTATGTTATTTCATG
```

Figure 1. Comparison of the sequence at the breakpoint of the *Arp* inversion with the wild type *Su(z)2* and P element sequences. The 20 nucleotides of P element sequence shown are from the 31 bp repeat present at the end of P elements. The KFMATCH program was used to align the sequences. The numbers for the wild type sequence correspond to the numbers in the submitted sequence (X56798). The *Arp* sequence shows perfect alignment with the wild type sequence downstream of nucleotide 34, and matches the P element sequence upstream of this location.