the dominant contribution to J_2 comes from low latitudes, the region where we have measured ν_{rot}.

According to parametrized, post-newtonian theories of gravitation, the predicted advance of the perihelion of Mercury is

$$\dot{\omega} = 42.95[(2+2\gamma-\beta)/3+2.9\times10^3 J_2] \quad \text{arc s per 100 yr} \quad (6)$$

where γ and β are parameters which are both unity in general relativity. Combining our value of J_2 with the planetary data of Shapiro *et al.*[21] and that of Anderson *et al.*[22], we have

$$(2+2\gamma-\beta)/3 = 1.002\pm0.005 \text{ and } 1.006\pm0.005 \quad (7)$$

Our value of J_2 makes an essentially negligible contribution to equation (7). In particular, the contribution of the J_2 term from equation (6) to equation (7) is a factor of 10 less than the errors quoted in equation (7), which are due to the planetary data. Therefore, at the present level of accuracy of the planetary data our value of J_2 is not in conflict with general relativity.

As noted above, there is an apparent discrepancy between observational estimates of the internal rotation. We calculate a markedly slower internal rotation rate than Hill *et al.*[9]. The discrepancies between the two results are much larger than the measurement uncertainties appear to allow. Unfortunately, there are no modes common to both data sets, so a direct comparison of splittings is not possible. The theory by which the interior rotation rate is calculated from a set of rotational splittings is sufficiently well understood that the two different sets of splittings should yield comparable rotation rates, at least if ν_{rot} is a slowly varying function of r.

It is mathematically possible (see ref. 23) to find a function ν_{rot} that is consistent with all of the data, but we find it implausible that the details of the rapid rotation variation that would then be required would be such as to impart high rotational splitting just to the modes that happen to have been identified in one set of observations. This suggests that there are systematic errors in at least one set of results. The most likely source of systematic errors is mode misidentification, that is, either a feature in a spectrum that is not a mode is wrongly identified as a mode or a feature is assigned incorrect values of n, l, m. The relative simplicity and high resolution in l in the spectra of Duvall and Harvey[1] compared with Bos and Hill[8], plus the unexpected appearance (Fig. 1 of ref. 9) of zonal modes in the latter analysis, lends confidence to the present work.

This work was done while T.L.D. was a visiting astronomer at the National Solar Observatory in Tucson. The National Solar Observatory is operated by the Association of Universities for Research In Astronomy, Inc. under contract with the NSF. The work of P.R.G. was partially supported by the NSF Astronomy Division and the Air Force Office of Scientific Research.

Received 16 March; accepted 30 April 1984.

1. Duvall, T. L. Jr & Harvey, J. W. *Nature* **310**, 19–22 (1984).
2. Cowling, T. G. & Newing, R. A. *Astrophys. J.* **109**, 149–158 (1949).
3. Ledoux, P. *Astrophys. J.* **144**, 373–384 (1951).
4. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Nature* **293**, 443–445 (1981).
5. Isaak, G. R. *Nature* **296**, 130–131 (1982).
6. Dicke, R. H. *Astrophys. J.* **228**, 898–902 (1979).
7. Woodard, M. & Hudson, H. *Bull. Am. astr. Soc.* **15**, 951–952 (1983).
8. Bos, R. J. & Hill, H. A. *Sol. Phys.* **82**, 89–102 (1983).
9. Hill, H. A., Bos, R. J. & Goode, P. R. *Phys. Rev. Lett.* **49**, 1794–1797 (1982).
10. Gough, D. O. *Nature* **298**, 334–339 (1982).
11. Delache, P. & Scherrer, P. *Nature* **306**, 651–653 (1983).
12. Hansen, C. J., Cox, J. P. & Van Horn, J. M. *Astrophys. J.* **217**, 151–159 (1977).
13. Gough, D. O. *Mon. Not. R. astr. Soc.* **196**, 731–745 (1981).
14. Christensen-Dalsgaard, J., Gough, D. O. & Morgan, J. G. *Astr. Astrophys.* **73**, 121–128 (erratum **79**, 260) (1979).
15. Backus, G. & Gilbert, F. *Phil. Trans. R. Soc.* A**266**, 123–192 (1970).
16. Gilman, P. A. in *Physics of the Sun* (Reidel, Dordrecht, in the press).
17. Zahn, J.-P. *IAU Symp.* **59**, 187–194 (1974).
18. Goldreich, P. & Schubert, G. *Astrophys. J.* **150**, 571–587 (1967).
19. Fricke, K. *Z. Astrophys.* **68**, 317–344 (1968).
20. Spruit, H. C., Knobloch, E. & Roxburgh, I. W. *Nature* **304**, 520–522 (1983).
21. Shapiro, I. I., Counselman, C. C. & King, R. W. *Phys. Rev. Lett.* **36**, 555–558 (1976).
22. Anderson, J. D., Keesey, M. S. W., Lau, E. L., Standish, E. M. & Newhall, X. X. *Acta astronaut.* **5**, 43–61 (1978).
23. Gough, D. O. *Mem. Soc. astr. Ital.* (in the press).

Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins

Allen Laughon & Matthew P. Scott

Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder, Campus Box 347, Boulder, Colorado 80309, USA

Mutations in the fushi tarazu (ftz) *locus of* Drosophila *result in embryos with half the usual number of body segments. The sequences of the wild-type gene, a temperature-sensitive allele and a dominant mutant allele are presented. A portion of the conserved protein domain present in* ftz *and several homoeotic genes resembles the DNA-binding region of prokaryotic DNA-binding proteins, and is also similar to products of the yeast mating-type locus.*

TWO well-defined processes in the development of *Drosophila melanogaster* are the establishment of a segmented body pattern and the specification of segmental identity (see ref. 1 for discussion). A large number of genes involved in establishing proper segmentation have been identified[2–7]. The homoeotic genes of the bithorax complex (BX-C) are involved in determining the identity of thoracic and abdominal segments[8–10]. A second cluster of genes, the antennapedia complex (ANT-C)[11], functions primarily in head and thoracic development and contains homoeotic loci, for example *Antennapedia* (*Antp*), *Sex combs reduced* (*Scr*) and *proboscipedia* (*pb*), and a segmentation gene *fushi tarazu* (*ftz*)[3]. Recessive lethal *ftz* mutant alleles result in embryos having half the normal number of segments. This phenotype has also been interpreted[4] as having homoeotic character as the pairwise fused segments appear to have the identity of the more anterior segment, that is, (T1 + T2) → T1, (T3 + A1) → T3, etc. Temperature shift experiments with a temperature-sensitive allele, ftz^{f47ts}, indicate that *ftz* expression is required only during the cellular blastoderm stage between 2 and 4 h after fertilization[4]; this is the time at which *ftz* transcripts are most abundant[12], long before segments become visible, suggesting that *ftz* functions in establishment but not maintenance of the pattern.

Molecular characterization of the ANT-C has included identification of the *Antp* and *ftz* transcript units[12–14]. One important finding is that the coding regions of the 3′ exons of *Antp* and *ftz* and of the *Ultrabithorax* (Ubx) gene of the BX-C share an ~180-base pair (bp) (60-codon) region of homology[13–16]. Over 90% of the amino acid sequences in this region are conserved[16]. The amino acid composition of the region is 28–30% lysine and

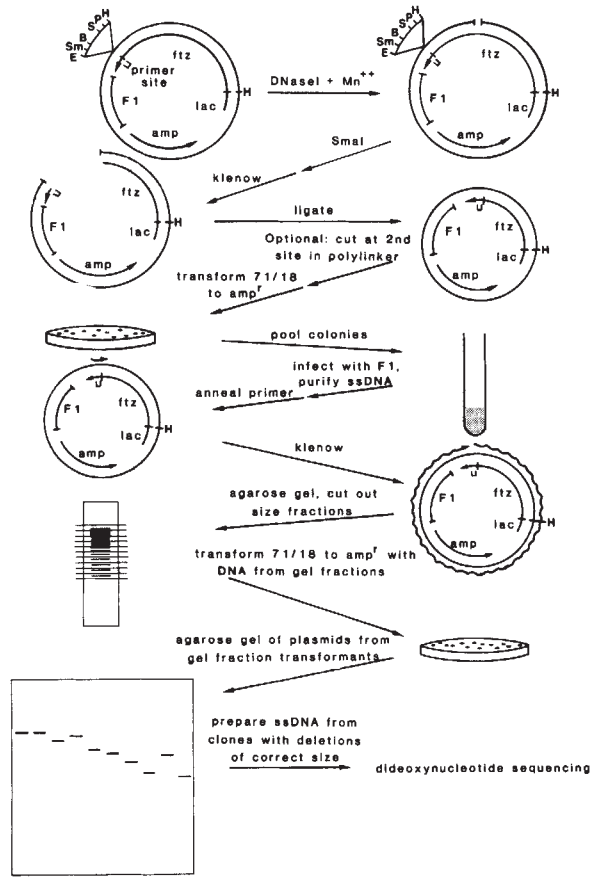

Fig. 1 Outline of a method for generating staggered DNA subclones for dideoxynucleotide sequencing. The key to this procedure is that relatively large DNA inserts (up to 6 kb in our experience) are stable in pEMBL single-stranded plasmids[38], which has allowed us to adopt the sequencing strategy of Frischauf *et al.*[39] for use with dideoxynucleotide sequencing. pEMBL plasmids contain the F1 replication origin and are replicated as single strands and packaged as phage when F1 infection provides replication functions in *trans*[38]. DNA purification and subsequent enzymatic steps were performed according to standard protocols[40]. Preparation of single-stranded DNA was as described elsewhere[38]. DNA to be sequenced (in this case a 3.5-kb *Hin*dIII fragment containing *ftz*) is cloned into a site in the pEMBL polylinker as far as possible from the primer site (in this case, the *Hin*dIII site of pEMBL9$^+$). An average of one random or quasi-random cut per molecule is made with DNase I in the presence of Mn^{2+} (ref. 41) or with restriction enzymes. A second cut is introduced with a restriction enzyme that cleaves as close to the primer site as possible, but which fails to cleave the insert DNA (*Sma*I in this case). The ends of the DNA are repaired with the Klenow fragment of DNA polymerase I followed by ligation of the DNA at concentrations which favour recircularization (10 μg ml^{-1}). Thus a series of deletions is generated, each having one end at the polylinker site. To eliminate clones deleted in the direction opposite to the insert DNA, the ligated DNA is cleaved with a second enzyme which cuts only at a site in the polylinker between the first restriction site and the insert DNA. DNAs with deletions extending in the wrong direction are cleaved while those with deletions into the insert (thus lacking this second restriction site) remain uncut. The DNA is then used to transform *E. coli* strain 71-18 (ref. 38), selecting for ampicillin resistance (Amp^r)[25]. Circular molecules transform at a much higher frequency than linear ones, eliminating most of the clones which have deletions extending in the wrong direction. About 500 *amp^r* colonies are pooled in liquid culture and infected with F1 strain IRI. Single-stranded F1 and pEMBL-derived DNA is purified from the resulting phage mixture. This step selects against deletions which destroy the F1 replication origin of the pEMBL plasmids. The mixture of single-stranded DNAs are then converted to double strands by annealing of the pentadecamer sequencing primer (New England Biolabs) and synthesis of the second strand with Klenow fragment. 2 ng of primer are annealed to 0.1 μg single-stranded DNA in 15 μl of 20 mM Tris-HCl *p*H 8.5, 10 mM $MgCl_2$ at 42 °C for 1 h; to this mixture is then added 2 μl of each 2 mM dXTP, 1 μl of 0.1 M dithiothreitol, 4 μl H_2O, 0.5 μl Klenow (5 units μl^{-1}). The reaction is carried out at 37 °C for 30 min, after which an additional 0.5 μl of Klenow is added, followed by 30 min more at 37 °C. The DNA is then fractionated in a 0.7% low-melting temperature agarose gel and slices between the sizes of pEMBL (deletion of entire insert) and pEMBL+insert (not deleted) cut out. Only DNAs which contain the primer site will be converted to double-stranded molecules, again selecting against deletions extending in the wrong direction. DNA is extracted from the gel slices and used to transform strain 71-18, selecting for ampicillin resistance. The sizes of deletions in individual *amp^r* clones are determined by agarose gel electrophoresis of double-stranded plasmid DNAs. Dideoxynucleotide sequencing is done using single-stranded DNA from clones with deletion end points 200–300 bp apart (the length of sequence determined per end being 250–350 bp).

arginine; the region may be involved in DNA or RNA binding. This region of homology also hybridizes with several other positions within the ANT-C and BX-C, suggesting that the genes of the two complexes are functionally and evolutionarily related and that the region of homology defines a common function or property of the products encoded by these complexes. In particular, it suggests a relationship between at least some homoeotic (*Antp* and *Ubx*) and segmentation (*ftz*) loci.

The *ftz* gene has been delimited to a 3.2-kilobase (kb) DNA fragment (3.5 kb in some *ftz*$^+$ chromosomes)[12]; this fragment contains both a 1.8-kb transcript and the two molecular lesions thus far associated with *ftz* mutations (a 4.9-kb insertion and a translocation breakpoint). Here we report the DNA sequences of three *ftz* alleles (the wild type and two mutations) and discuss the implications of this sequence information for the function of *ftz* and related genes.

Structure of the *ftz* gene

Clones for dideoxynucleotide sequencing of *ftz* were constructed as described in Fig. 1 legend. The regions sequenced are summarized in Fig. 2 and the nucleotide sequences of *ftz* genomic and cDNA clones and of *ftz^{f47ts}* and *ftz^{Rp1}* mutations are given in Fig. 3. The *ftz* genomic clone contains a 1,239-bp open reading frame in the direction of transcription which is split by a 150-bp intron containing stop codons in all three frames. The cDNA clone does not contain this intron. Transcription starts at position −191 as determined from primer extension experiments (data not shown). The −191 start point is also consistent with the structure of the cDNA (see below). Two 'TATA' boxes occur 21 and 31 bp upstream from the transcriptional start; the latter is the more conventional spacing. A series of 'AATAAA' poly(A) addition site sequences occur immediately downstream from the 3′ end of the cDNA, but no data are available to determine which of these is used. A transcript ending within the series of polyadenylation sites, with the intron spliced out, would be approximately 1.8 kb in size, in agreement with the size of the previously reported *ftz* RNA[12].

Structure of the *ftz* cDNA clone

The first 91 bp at the 5′ end of the cDNA sequence do not correspond to the genomic sequence (Fig. 3). The structure of the 5′ end of the cDNA clone could have arisen if, during cDNA synthesis, an inverted repeat complementary to the 5′ end of the transcript (Fig. 3) formed a hairpin structure in the first cDNA strand (Fig. 4). The hairpin could then serve to prime second strand synthesis. S_1 nuclease was used to remove the foldback loop during preparation of the library and the observed cDNA 5′ end would have been created by S_1 nuclease cleavage at the site indicate in Fig. 4. The cloning of the cDNA was completed by ligation of *Eco*RI linkers, transformation into *Escherichia coli*, and apparent subsequent resolution of the heteroduplex in favour of the bold-lettered strand in Fig. 4. As synthesis of cDNA libraries commonly relies on self-primed synthesis of the second cDNA strand, it is likely that many cDNAs will contain such anomalous sequences at their 5′ ends. Another example has been observed in a *Ubx* cDNA clone from

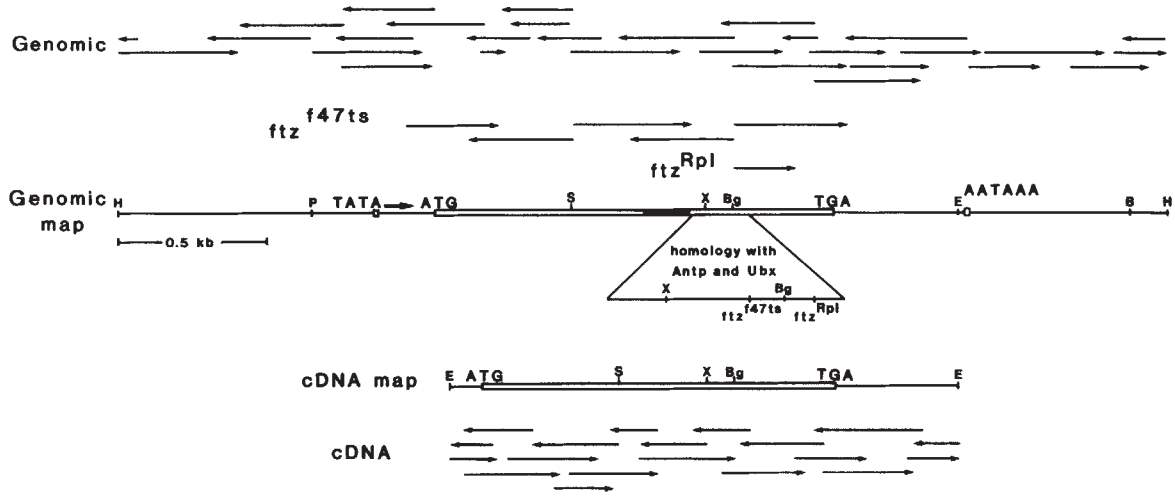

Fig. 2 Summary of dideoxynucleotide sequencing of wild-type and mutant *ftz* genomic clones and a *ftz* cDNA clone. Deleted clones for sequencing were derived as described in Fig. 1 legend, except for the sequences of ftz^{f47ts} and ftz^{Rp1} in which deletions were made from specific restriction sites. The genomic clone is a 3.5-kb *Hin*dIII fragment from a ftz^{+} P^{P} chromosome[12]. The ftz^{f47ts} mutation was induced on this same chromosome. The *ftz* cDNA clone was isolated as a 1.7-kb *Eco*RI fragment from the Oregon R, 0–5-h embryonic cDNA library of Goldschmidt-Clermont, Saint and Hogness (personal communication). Genomic and cDNA sequencing runs are shown above and below the maps of these clones respectively. Open bar, coding region; solid bar, intron. The positions of TATA and AATAAA boxes are shown; arrow, start of transcription. The region of homology with *Antp* and *Ubx*[16] is shown expanded, as are the positions of the ftz^{f47ts} point mutation and the ftz^{Rp1} (refs 2,3) reciprocal translocation breakpoint: material to the right (distal) of the ftz^{Rp1} breakpoint is from chromosome 2, or may be a small insertion of unknown origin[12]. B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sal*I; X, *Xho*I. Single-stranded DNA was prepared for sequencing as described by Biggin *et al.*[42] except that reactions were done in the wells of microtitre plates (D. Peattie, personal communication).

the same library (M. Goldschmidt-Clermont, R. Saint and D. S. Hogness, personal communication). The secondary structure at the 5′ end of the *ftz* mRNA could have a function relating to the translation, processing, localization or stability of the RNA.

Encoded proteins

The first AUG in the *ftz* transcript is 24 bp in from the 5′ end at position −97 in Fig. 3; translation starting at this AUG would result in a 22-amino acid peptide ending at a UAG triplet. This reading frame is out of register with the main 1,239-bp *ftz* open frame. The context of adjacent nucleotides at the −97 AUG has been observed only once in initiator AUGs but is found 14 times in 'nonfunctional' upstream AUGs[17]. However, the codon usage of this short open reading frame is biased correctly for *D. melanogaster* coding regions (P. O'Connell and M. Rosbash, personal communication), suggesting that it may be functional. As the 22-codon open reading frame is within a hairpin loop of the RNA (Fig. 4), its translation might influence RNA secondary structure. It will be necessary to generate mutations within the 22-codon region to determine whether it is functional. The second AUG in the *ftz* transcript is the beginning of the 1,239-bp open reading frame (position 1). The second AUG may also be involved in base pairing if the secondary structure shown in Fig. 4 exists in the mRNA. The third AUG at position 46 is also in the correct frame and is not included in the inverted repeat at the 5′ end of the *ftz* transcript. Translation starting at position 1 would yield a 413-amino acid protein of molecular weight (MW) 45,000 while initiation at position 46 would yield a 398-amino acid protein of MW 43,500.

By several criteria the *ftz* protein seems to be divided into three domains. The 189-bp *Antp*, *ftz* and *Ubx* homology region begins 8 bp downstream from the 150-bp intron in *ftz* and the break in homology at the 3′ end is abrupt. Of the residues encoded by the homology region, 30.6% are basic while the two remaining portions of the protein at the amino and carboxyl ends are 9.9% and 10.4% basic, respectively. Thus, the protein seems to be divided into relatively neutral amino-terminal and carboxy-terminal domains with a very basic homology region domain in between. In addition, the amino and carboxyl domains each comprise more than 10% proline and 10% tyrosine and the carboxyl domain comprises almost 15% glutamine. The homology domain is not rich in any of these amino acids.

Comparison of the *ftz* genomic DNA sequence with that of the cDNA clone (isolated from the Oregon R cDNA library of Goldschmidt-Clermont, Saint and Hogness) reveals six sequence polymorphisms (Fig. 3): four of these are in the third base of a codon and do not change the amino acid sequence. Another third-position difference at nucleotide 1,344 places a histidine, rather than a glutamine, in the cDNA. The sixth polymorphism is the deletion of 9 bp in the cDNA clone encoding Tyr-His-Ser at nucleotides 139–147 in the genomic clone.

Structure of mutant *ftz* alleles

The *ftz* locus was first mapped by using chromosome rearrangements associated with *ftz* mutations[12]. One of the mutations, ftz^{Rp1}, originally found by I. Duncan, is associated with a translocation of parts of chromosomes 2 and 3. The ftz^{Rp1} allele is of particular interest because it has both a recessive lethal *ftz* phenotype and a dominant *postbithorax*-like phenotype. The *postbithorax* (*pbx*) phenotype, in which posterior third thoracic structures are transformed into posterior second thoracic structures, is normally a recessive phenotype resulting from mutations in the *pbx* locus of the bithorax complex. Flies that are *pbx*/*pbx* and flies that are ftz^{Rp1}/+ both have posterior halteres transformed into posterior wings (that is, posterior T3 → posterior T2). The dominant effect of ftz^{Rp1} cannot be due to loss of *ftz* function, as flies with a deletion of *ftz* heterozygous with a wild-type *ftz* allele develop normally. Therefore, the dominance of ftz^{Rp1} is due to abnormal activity of *ftz*—this is consistent with the partial loss of function recessive lethal phenotype of ftz^{Rp1}.

A restriction map of the ftz^{Rp1} allele showed the breakpoint within the *ftz* gene to be near the 3′ end[12]. The relevant part of the mutant gene was sequenced (Fig. 3) and reveals that ftz^{Rp1} encodes a truncated protein in which the C-terminal 100 amino acids of the normal protein have been replaced by 10 novel amino acids. The breakpoint of the chromosome rearrangement is within the sequence that is similar to the other homoeotic loci. The homology region is largely but not completely intact in the ftz^{Rp1} protein product. It seems likely that the shortened

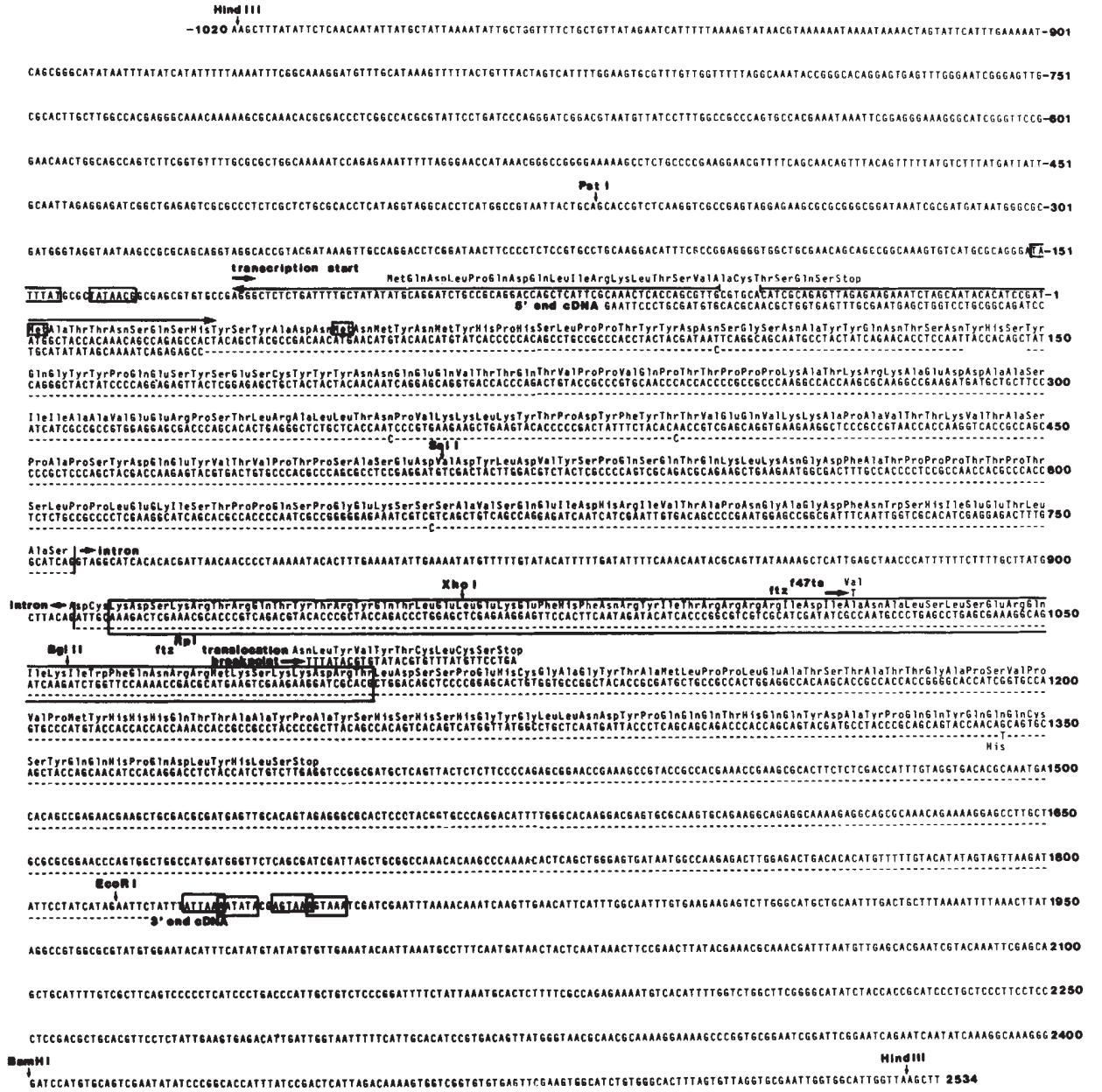

Fig. 3 DNA sequences of *ftz*. The genomic sequence is numbered from −1,020 to +2,534. The amino acid sequence of the encoded protein is shown above the DNA sequences. The cDNA sequence appears below the genomic sequence starting at −65 and ending at the *Eco*RI site at 1,818. Dashes are shown where the cDNA sequence is the same as the genomic sequence. The region of homology with *Antp* and *Ubx* is boxed, as are putative initiator ATGs, TATA and AATAAA boxes. A large imperfect inverted repeat at the 5′ end of the transcription unit is denoted by arrows running in opposite directions. ftz^{f47ts} and ftz^{Rp1} mutations are shown together with the resulting change in amino acid sequence. The cDNA clone ends at the genomic *Eco*RI site, presumably due to incomplete *Eco*RI methylation during cDNA cloning.

protein is responsible for the ftz^{Rp1} dominant phenotype. It is also possible that the mutation affects RNA structure or processing.

We have also determined the DNA sequence of a temperature-sensitive *ftz* allele, ftz^{f47ts}. This mutation has been used previously to determine the developmental stages during which *ftz* function is important[4]. The entire coding region of the temperature-sensitive allele was sequenced. The wild-type genomic sequence in Fig. 3 is from the chromosome in which the ftz^{f47ts} mutation was induced, and was chosen to prevent confusion resulting from sequence differences between strains. Only one change in the DNA sequence was found in the ftz^{f47ts} coding region: a transition of C to T (Fig. 3) which results in the substitution of a valine for the normal alanine, a fairly conservative amino acid change. The temperature sensitivity of ftz^{f47ts} is probably due to the production of a slightly modified, thermolabile protein. The protein encoded by ftz^{f47ts} is altered within the region of homology with other *Drosophila* homoeotic gene products.

Possible DNA-binding domain

Crystallographic structures determined for the cro and repressor proteins of bacteriophage λ and the CAP protein of *E. coli* have revealed the existence of conserved α-helical structures which computer-modelling studies suggest are the DNA-binding sites for these proteins[18–24]. The model is supported by studies of protection from chemical modification by binding of these pro-

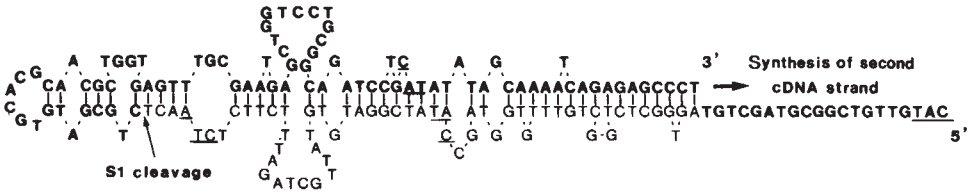

Fig. 4 Formation of the *ftz* cDNA. The figure shows a hypothetical secondary structure of a cDNA strand complementary to the 5′ end of the *ftz* transcript. Nucleotides shown in bold letters correspond to the observed sequence of the cDNA clone. The 3′ end of the hairpin structure presumably serves as primer for the synthesis of the second cDNA strand. S_1 nuclease cleavage at the indicated site followed by ligation of *Eco*RI linkers would give rise to the observed cDNA sequence if the mismatched sequences were resolved by copying the bold-lettered strand. The underlined codon triplets are complementary to the 'upstream' AUG which would begin a 22-amino acid leader peptide, the UAG ending this open reading frame, and the two possible initiator AUGs for the major *ftz* open reading frame.

teins to their operator DNAs[24] and by extensive genetic analysis[25,26]. The DNA-binding structure (reviewed in ref. 24) comprises two α-helices: 'helix 3' fits into the major groove of DNA and 'helix 2' lies across helix 3, holding it in position. The two helices are connected by a β-turn. The amino acids which are important for the conformation of this structure are conserved in these proteins and in homologous proteins from other lambdoid phages as summarized in Fig. 5. Alanine is usually found in position 7, isoleucine or valine in position 17, and hydrophobic residues are found in positions 6, 10, 12 and 20. The side chains at positions 10–12 are small to permit a tight turn between the helices. Amino acids in positions 13–16, 18, 19, 21 and 22 are involved in determining DNA sequence specificity (other parts of the protein can have some influence also). Similar patterns of conserved amino acids have been found in several bacterial regulatory proteins and in the *MATa1* gene of yeast (Fig. 5)[21–24,27]. The position of the DNA-binding domain within the bacterial proteins varies (Fig. 5).

The conserved pattern of amino acids found in bacterial DNA-binding proteins is also present near one end of the conserved amino acid sequence common to *ftz*, *Antp* and *Ubx* (Fig. 5). The *ftz*, *Antp* and *Ubx* sequences are also similar to the putative DNA-binding regions encoded by the yeast *MATa1* and *MATα2* genes (19/29 amino acids are homologous to *MATa1*, 13/29 amino acids are homologous to *MATα2)*. Similarity between *MATα2* and the *ftz*, *Antp* and *Ubx* homology regions was originally noted by Shepard *et al.*[28]. The homologous *Drosophila* and yeast sequences shown in Fig. 5 are flanked by very basic regions that are unlikely to form α-helices. The conserved glycine residue following helix 2 is replaced by serine in *ftz* and by cysteine in *Antp* and *Ubx*; this glycine is thought to be important for formation of the tight turn necessary for the two helices to come into close contact. However, a glutamic acid residue can be accommodated at this position in a functional λ repressor[29]. In addition, serine and cysteine are the only substitutions for glycine at this position in putative DNA-binding structures from other bacterial proteins (Fig. 5).

Further evidence that the possible DNA-binding region of *ftz* is in fact important for the function of the protein comes from the temperature-sensitive mutation described above. The ftz^{f47ts} mutation changes the highly conserved alanine in helix 2 to a valine. The alanine normally comes in close contact with the conserved isoleucine/valine in helix 3 (ref. 24). Although the Ala→Val change is fairly conservative, the added bulk of the valine side chain would be predicted to increase the distance between the two helices, perhaps destabilizing the protein in this region. It would be instructive to know the phenotype of a mutation affecting this alanine in λ repressor, cro or lac repressor (in which many mutations have been sequenced[25,26,30]), but unfortunately no mutations at this position have been reported. The ftz^{Rp1} translocation breakpoint occurs exactly at one end of the homology with *MATa1* and *MATα2* and probably results in an abnormally functioning protein.

According to the model outlined above, the specificity of DNA binding would be determined by the side groups protruding from the α-helix which lies in the DNA major groove (helix 3 in Fig. 5). The sequences of *ftz*, *Antp* and *Ubx* are identical in this region, suggesting that they might recognize and bind to the same DNA sequence.

Conclusions

The extraordinary size and complexity of some of the *Drosophila* homoeotic genes, for example the 73-kb *Ubx* unit[31] and the 103-kb *Antp* unit[13,14], contrasts with the relative simplicity of the *ftz* gene[12]. Yet *ftz* shares a sequence encoding a protein domain of ~6,500 MW with *Antp* and *Ubx*[16]. The ftz^{Rp1} allele provides another indication of related function. This *ftz* allele results in a phenotype characteristic of a class of mutations in the bithorax complex. The recessive *ftz* phenotype (pairwise segmental fusions) sets *ftz* apart from the classical homoeotic genes, in which mutations generally cause dramatic transformations of only one to a few body segments. The function of ftz^{+} seems to be to determine the formation of alternate segments; presumably other homoeotic genes further specify the identities of the segments. The location of *ftz* only 30 kb from the homology domain of *Antp* suggests an evolutionary relationship between the *Antp* 3′ exon and the *ftz* 3′ exon[16].

It has been clear from genetic experiments, and has been confirmed by *in situ* hybridization to sectioned tissues[32–35], that there is specific expression of *Drosophila* regulatory genes in particular regions of the embryo. The *in situ* hybridization experiments demonstrate that expression is controlled at the level of RNA synthesis or stability, which leads to the question: how is spatial regulation of RNA accumulation attained? A second question is how the regulatory genes direct the development of particular structures in lieu of the alternative structures that develop when the genes are mutated. The structure of the protein encoded by *ftz*, and of structurally related proteins including products of *Antp* and *Ubx*, indicate that DNA binding may be an important aspect of the function of genes that control segmentation.

The data presented above suggest that the protein structure common to *ftz*, *Antp* and *Ubx* products includes a DNA-binding region of the type observed in some bacterial proteins. The primary (but not the only) part of the bacterial proteins involved in DNA sequence recognition is the helix 3 domain. Note that the amino acid sequences of *Antp*, *Ubx* and *ftz* products are identical in the region of the homoeotic homologue of helix 3, and in flanking regions. The one exception is a serine residue in *ftz*, just N-terminal to helix 3, in the position occupied by threonine in *Antp*, *Ubx*, *MATa1* and *Fnr* (Fig. 5). The near identity of the homoeotic gene products in this critical region suggests that they may bind to similar or identical DNA sequences, and may in fact compete with each other during early embryogenesis. Such competition suggests a mechanism behind mutations such as ftz^{Rp1}, in which a BX-C recessive loss of function phenotype (*pbx*) results from a dominant *ftz* mutation. The truncated ftz^{Rp1} protein product may compete with or otherwise influence the DNA binding of BX-C products. The binding affinities could also be dependent on cooperative or competitive interactions with other proteins. It is also possible that *ftz* regulates expression of ANT-C or BX-C genes.

There are many other genes that are not obviously homoeotic in which mutations alter the number or pattern of segments that develop[2,5–7]. DNA hybridization experiments suggest that the region of homology common to *ftz*, *Antp* and *Ubx* cannot be present at most of the other segmentation loci (ref. 15 and unpublished results). However, it is possible that a similar

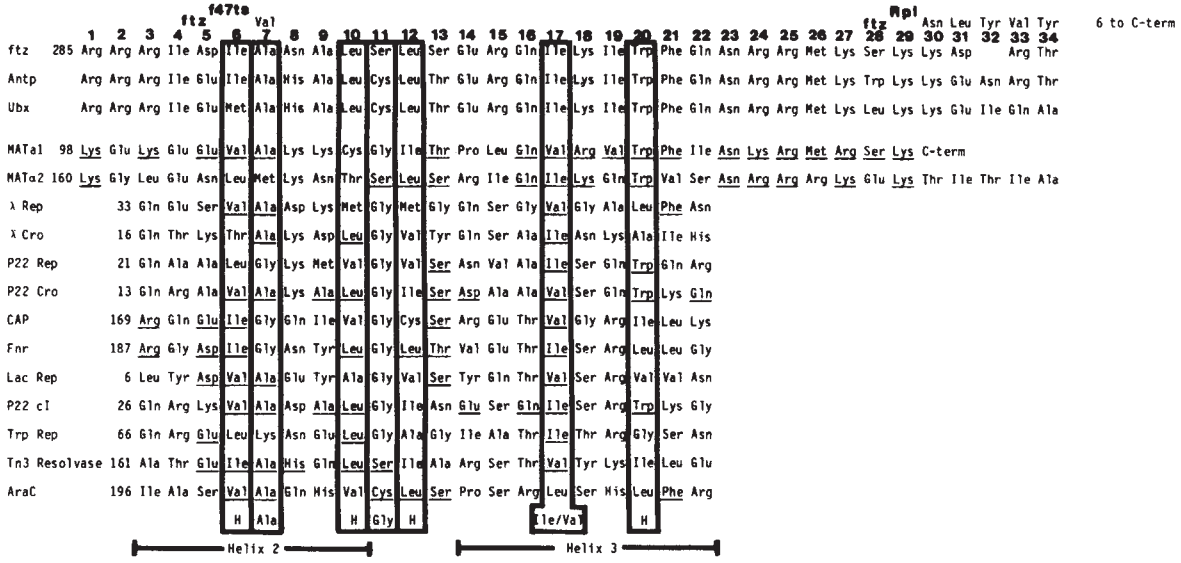

Fig. 5 Putative DNA-binding domain within the homoeotic homology region. *ftz*, *Antp* and *Ubx* sequences at the 3′ end of their shared homology are aligned with prokaryotic and potential yeast DNA-binding domains[24]. The sequence of the yeast *MATa1* protein past the tryptophan at position 118 is different from that reported previously due to previously undiscovered *MATa1* introns (K. Nasmyth, personal communication). Sequence changes due to *ftz* mutations are shown at the top. The positions of the two α-helices of the DNA-binding domain (numbered 2 and 3 according to convention for the λ cro protein) are shown at the bottom. Conserved residues within the domain, important for close alignment of the two helices, are boxed. Amino acids that are structurally homologous to *ftz*, *Antp* or *Ubx* residues are underlined. The number to the left of each sequence is the position of the first amino acid shown within the complete protein. The positions of the amino acids are arbitrarily numbered at the top (1–34). The collection of bacterial sequences shown is from a review by Pabo and Sauer (ref. 24 and references therein). Yeast sequences are from ref. 43 and K. Nasmyth (personal communication); *Drosophila* sequences are from ref. 16.

domain is encoded by some of the other loci but is sufficiently divergent in sequence to be recognizable only when the gene, or its protein, has been sequenced.

If part of the homology region of the homoeotic genes encodes a DNA-binding domain, what could be the function of the rest of the homology region? The most striking aspect of the remaining sequence is its basicity: arginine and lysine comprise about 30% of the amino acids in the homology domain[16] Very basic sequences are also present in the C-terminal parts of *Antp* and *Ubx*, just downstream from the homology region (M.P.S., unpublished data). Such a basic protein sequence is suggestive of nonspecific DNA or RNA binding; domains known to be involved in specific DNA binding are not especially basic.

There are important limitations to the correlations we have made between the structure of the homoeotic gene products and the bacterial DNA-binding proteins: (1) Only three of the bacterial proteins have been crystallized[18–20], so the full range of possible variations in the structure of the DNA-binding helix-turn-helix structure is unknown. (2) It is unknown whether the parts of the two yeast mating-type protein products that are homologous with the bacterial proteins are in fact used for DNA binding. (3) Some known DNA-binding proteins apparently lack this type of binding sequence[36,37]. However, the similarity of the *Drosophila* sequences to the bacterial sequences is striking, and the type of helix-turn-helix structure found in these DNA-binding proteins has not been observed in proteins which are not DNA-binding proteins. It has also been shown that the *MATα2* product is a sequence-specific DNA-binding protein (A. Johnson and I. Herskowitz, personal communication). Thus, we believe that the structural homology between the homoeotic gene products and known DNA-binding proteins is strong enough to suggest that the homoeotic genes may, at least in part, function through DNA-binding mechanisms.

We thank Ian Duncan for the opportunity to analyse ftz^{Rp1} and for discussions; Thomas Kaufman for ideas about *ftz* and ftz^{Rp1}; and Nello Bossi for conversations on sequencing strategies. M. Goldschmidt-Clermont, R. Saint, P. A. Beachy and D. S. Hogness provided the cDNA library and unpublished information about BX-C. Robert Laymon helped with subcloning of DNA. Sean Carroll, Bill Wood and Dan Stinchcomb provided useful criticisms of the manuscript and Cathy Inouye helped to prepare it. A.L. is the recipient of an American Cancer Society postdoctoral fellowship, and M.P.S. is supported in part by an American Cancer Society Junior Faculty Research Award. The research support was provided by a grant from the NIH (HD 18163-01).

Received 25 May; accepted 7 June 1984.

1. Raff, R. A. & Kaufman, T. C. *Embryos, Genes and Evolution* (Macmillan, London, 1983).
2. Nusslein-Volhard, L. & Wieschaus, E. *Nature* **287**, 795–801 (1979).
3. Wakimoto, B. T. & Kaufman, T. C. *Devl Biol.* **81**, 51–64 (1981).
4. Wakimoto, B. T., Turner, F. R. & Kaufman, T. C. *Devl Biol.* **102**, 147–172 (1984).
5. Nusslein-Volhard, C., Wieschaus, E. & Kluding, H. *Wilhelm Roux Arch. dev. Biol.* (in the press).
6. Jurgens, G., Wieschaus, E., Nusslein-Volhard, C. & Kluding, H. *Wilhelm Roux Arch. dev. Biol.* (in the press).
7. Wieschaus, E., Nusslein-Volhard, C. & Jurgens, G. *Wilhelm Roux Arch. dev. Biol.* (in the press).
8. Lewis, E. B. *Cold Spring Harb. Symp. quant. Biol.* **16**, 159–174 (1951).
9. Lewis, E. B. *Nature* **276**, 565–570 (1978).
10. Bender, W. *et al. Science* **221**, 23–29 (1983).
11. Kaufman, T. C., Lewis, R. & Wakimoto, B. *Genetics* **94**, 115–133 (1980).
12. Wiener, A. J., Scott, M. P. & Kaufman, T. C. *Cell* (in the press).
13. Scott, M. P. *et al. Cell* **35**, 736–776 (1983).
14. Garber, R. L., Kuroiwa, A. & Gehring, W. J. *EMBO J.* **2**, 2027–2036 (1983).
15. McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. *Nature* **308**, 428–433 (1984).
16. Scott, M. P. & Weiner, A. J. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
17. Kozak, M. *Microbiol. Rev.* **47**, 1–45 (1983).
18. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. *Nature* **290**, 754–758 (1981).
19. McKay, D. B. & Steitz, T. A. *Nature* **290**, 744–749 (1981).
20. Pabo, C. O. & Lewis, M. *Nature* **298**, 443–447 (1982).
21. Matthews, B. W., Ohlendorf, D. H., Anderson, W. F. & Takeda, Y. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1428–1432 (1982).
22. Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Pabo, C. O. *Nature* **298**, 447–451 (1982).
23. Weber, I. T., McKay, D. B. & Steitz, T. A. *Nucleic Acids Res.* **10**, 5086–5102 (1982).
24. Pabo, C. O. & Sauer, R. T. *A. Rev. Biochem.* (in the press).
25. Nelson, H. C. M., Hecht, M. H. & Sauer, R. T. *Cold Spring Harb. Symp. quant. Biol.* **47**, 441–449 (1983).
26. Hecht, M. H., Nelson, H. C. M. & Sauer, R. T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2676–2680 (1983).
27. Matthews, B. W., Ohlendorf, O. H., Anderson, W. F., Fisher, R. G. & Takeda, Y. *Cold Spring Harb. Symp. quant. Biol.* **47**, 427–433 (1983).
28. Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., De Robertis, E. M. & Gehring, W. J. *Nature* **310**, 70–71 (1984).

29. Hochschild, A., Irwin, N. & Ptashne, M. *Cell* **32**, 319–325 (1983).
30. Miller, J. H. in *The Operon* (eds Miller, J. H. & Reznikoff, W.) 31–88 (Cold Spring Harbor Laboratory, New York, 1978).
31. Akam, M. *Trends biochem. Sci.* **8**, 173–177 (1983).
32. Akam, M. *EMBO J.* **2**, 2075–2084 (1983).
33. Levine, M., Hafen, E., Garber, R. L. & Gehring, W. J. *EMBO J.* **2**, 2037–2046 (1983).
34. Hafen, E., Levine, M. & Gehring, W. J. *Nature* **307**, 287–289 (1984).
35. Hafen, E., Kuroiwa, A. & Gehring, W. J. *Cell* (in the press).
36. Battey, J. *et al. Cell* **34**, 779–787 (1983).
37. Fiers, W. *et al. Nature* **273**, 113–120 (1978).
38. Dente, L., Cesareni, G. & Cortese, R. *Nucleic Acids Res.* **11**, 1645–1655 (1983).
39. Frischauf, A. M., Garoff, H. & Lehrach, H. *Nucleic Acids Res.* **8**, 5541–5549 (1980).
40. Maniates, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1982).
41. Heffron, F., So, M. & McCarthy, B. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6012–6016 (1978).
42. Biggin, M. D., Gibson, T. J. & Hong, G. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963–3965 (1983).
43. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C. & Smith, M. *Cold Spring Harb. Symp. quant. Biol.* **45**, 961–981 (1980).

LETTERS TO NATURE

Nature of blue galaxies in the cluster Cl1447+2619

Harvey R. Butcher

Kapteyn Astronomical Institute, Postbus 800, 9700 AV Groningen, The Netherlands

Augustus Oemler Jr

Yale University Observatory, PO Box 6666, New Haven, Connecticut 06511, USA

Photometric studies of distant clusters of galaxies provide evidence for significant evolution of the member galaxies during the relatively recent cosmological past[1–5]. Here, spectroscopy is presented that confirms a previous photometric result that the distant ($z = 0.38$) cluster Cl1447+2619 contains the largest excess of blue galaxies yet encountered. The galaxy surface density profile of the cluster shows it to be open and irregular, making it the first such cluster to be studied at high redshift. The spectra indicate that the blue colours are predominantly the result of recent star formation. We speculate that vigorous star formation in galaxies, independently of their environments, was much more common several Gyr ago than it is in the Universe today.

We have discussed elsewhere[3] broad-band photometric data on 33 nearby and distant (to $z = 0.54$) clusters of galaxies and the nearby field. That sample was chosen from the literature to include as many large redshift clusters as were known to be very rich and to have at least one redshift measurement. Analysis of the radial profiles of galaxy surface density showed that most clusters were of the compact, regular type, indicative of advanced dynamical age. After taking into account selection effects due to k-corrections, the colour–magnitude relation for E and S0 galaxies, and accidental errors, and after careful treatment for foreground and background contamination, a trend of increasing fraction of blue galaxies with redshift was found for these compact clusters. It was suggested that this is evidence for a general depletion of gas in the Universe during recent epochs, and that exceptions to the trend indicate superimposed phenomena related to the particular evolution of each cluster.

Only one distant, rich cluster of the open or irregular type has been identified and studied so far, namely Cl1447+2619 (incorrectly called Cl1446+2619 in our previous work) at $z = 0.38$. Such systems are thought to be unrelaxed, and therefore to be of intermediate or young dynamical age. Nearby clusters of this type contain a galaxy-population mix intermediate between those of compact clusters and the field. Cl1447+2619 and the three nearby open systems were found to exhibit higher blue galaxy contents than the compact clusters, suggesting that characteristic population differences between cluster types obtain at all redshifts (at least to $z = 0.5$). In addition, the distribution of galaxy colours in Cl1447+2619 appears to be skewed more to the blue than is the case in any nearby population, even including the field sample. One is tempted, by continuity arguments, to speculate that the field population at high redshift will also be found to show signs of strong evolution.

Nearby open clusters contain elliptical, S0, and spiral galaxies roughly in the proportion E/S0/Sp = 1:2:3 (ref. 6). The fractional blue galaxy content, f_B, as defined in ref. 3, for nearby open systems is $f_B \simeq 0.15$. Thus, nearby open clusters have a ratio of spiral galaxy fraction, f_{Sp}, to blue galaxy fraction of $f_{Sp}/f_B \simeq 3.3$. For Cl1447+2619, $f_B = 0.36 \pm 0.05$; if the ratio to spiral fraction is similar to that nearby, one would predict that all galaxies in the cluster must be spirals! The cluster is likely, therefore, to be an ideal testing ground for theories of morphological evolution between galaxy (Hubble) types, as well as an excellent site for the study of the photometric evolution of galaxies such as our own.

The photometric technique used to determine f_B relies on the statistical subtraction of foreground and background contamination across the field of the cluster. The fact that galaxies are not distributed randomly across the sky, but tend to clump in groups, clusters, clouds and superclusters, causes some uncertainty as to whether the decontamination has been correctly accomplished in any individual case and an accurate value for f_B determined. The only recourse is to obtain spectroscopy for as many galaxies as possible in the most interesting clusters identified photometrically. Cl1447+2619 is clearly one such cluster, and so we have tried to verify by spectroscopy its exceedingly large blue galaxy content.

The instrumental apparatus used for this work has been described in detail previously[7]. Briefly, it consists of a large format, low-noise charge-coupled device (CCD) detector in a low resolution, optically efficient spectrometer on the 4-m Mayall telescope at Kitt Peak. Masks having patterns of 2 arc s diameter apertures, configured to correspond to the locations of objects on the sky, are inserted at the entrance to the spectrometer, enabling simultaneous observation of many objects across a 5 arc min field of view. Round apertures have been used in preference to short slits because they permit work on larger numbers of galaxies in the rather crowded cluster field; they have the disadvantage of an increased susceptibility to the effects of atmospheric dispersion. For the present study, a resolution of 17 Å and a wavelength coverage of roughly 4,500–7,500 A were used.

Two aperture plate masks were prepared for the Cl1447+2619 field, yielding spectra of 21 galaxies in the field of the cluster. None of the observations could be made near meridian passage, and no atmospheric dispersion compensation prisms were available. Therefore, the total integration was divided into 1-h segments, and these were distributed between the two aperture plates and over two nights of observation. Even so, the effects of atmospheric dispersion are evident among the data, and some data for outlying objects in some integration segments have been discarded. The resulting integration times for the 21 galaxies range from 2 to 5 h. Seeing during both nights was 2 arc s FWHM. Examples of the final spectra are shown in Fig. 1.

The results for the 21 galaxies are given in Table 1. The first 16 objects are from ref. 5, and include all blue galaxy candidates brighter than red magnitude 20.2, except one (no. 15). An additional five objects were observed, which lie outside the field having photometric data, but which seemed to be in the right magnitude range to be cluster members.

Of the total sample, two galaxies are seen clearly to be non-cluster members (nos 1 and a), and five (including no. 9) have spectra of too low a quality to yield a reliable redshift. Of the remaining 11 objects with photometry, seven are blue (according to the definition in ref. 3) and four are red. If we suppose that the uncertain objects among the first 16 are all unrelated to the