Volume 15 Number 10 1987 Nucleic Acids Research

A novel GC-rich dispersed repeat sequence in *Drosophila melanogaster*

Andrew J.Flavell[1]*, Julian Dyson[2] and David Ish-Horowicz[3]

[1]Department of Biochemistry, The University, Dundee DD1 4HN, [2]Department of Genetics, Cambridge University, Downing Street, Cambridge and [3]Imperial Cancer Research Fund, Developmental Biology Unit, Department of Zoology, Oxford University, South Parks Road, Oxford OX1 3PS, UK

Received February 23, 1987; Revised and Accepted April 23, 1987 Accession number Y00321

ABSTRACT

A novel family of dispersed repeat sequences from *Drosophila melanogaster* is described. Sequence analysis of two members of this family show them to contain greater than 75% GC bases. These are comprised of multiple repeats of GGX triplets interspersed occasionally with CGPy and TTPy. Southern blotting shows that these repeats are not transposable elements. Twenty four homologous recombinants have been localised by *in situ* hybridization to seven sites in the *Drosophila* genome. Polyadenylated RNAs homologous to this repeat family are expressed in a complex pattern which is developmentally regulated. We suggest that this family encodes a set of glycine-rich domains in *Drosophila* proteins.

INTRODUCTION

The genomes of all multicellular eukaryotes examined to date contain a wide variety of repeated DNA sequences. Repetitious DNA can be present either in tandem arrays, as in satellite DNA (1) or interspersed within other sequences. *Drosophila* contains many different classes of interspersed repeats (2,3). The majority of these sequences are of discrete size (usually several kilobases) and are usually transposable in the fly genome . These sequences are perhaps best considered as parasitic or selfish DNA (4,5) as, to date, no function encoded by them has been found to be essential to the host. Shorter middle repetitious sequences are also known. One of these, the homoeo box, is non mobile, well conserved between distantly related eukaryotes and forms part of several genes crucial to the development of *Drosophila* spatial organisation or pattern (6-8). Another immobile sequence, the *opa* repeat, is contained in several genes specifying the development of spatial pattern and the nervous system of *Drosophila* (9,10). The *opa* repeat is comprised of repeats of the triplets CAPu and encodes a glutamine-rich region in the *Notch* gene which is involved in *Drosophila* neurogenesis (9) . Both two types of repeat therefore form part of genes which are essential for correct development of the *Drosophila*. We describe here the isolation of a GC rich repeat (GCR repeat) from *Drosophila* which we suggest encodes a glycine-rich region in proteins which may also be important in *Drosophila* development.

MATERIALS AND METHODS

Isolation of recombinants containing GCR repeats

DNA was isolated from 200-500 Oregon R strain flies by homogenisation in 2 ml 10 mM Tris-HCl; pH 7.5, 60 mM NaCl, 10 mM EDTA, 0.15 mM spermine, 0.15 mM spermidine at 0°C using a Dounce homogeniser ('A' pestle). 0.7 ml lysis buffer was added (0.5M Tris-HCl; pH 8, 30 mM EDTA, 2% SDS, 0.2 mg/ml proteinase K (Merck)), the lysate was mixed gently then incubated at 37°C for 1h. The lysate was extracted twice with phenol mixture (redistilled phenol containing 0.2% 8-hydroxyquinoline and equilibrated with 1M Tris-HCl pH 7.5) and DNA spooled after addition of two volumes of ethanol. The DNA was redissolved in TE (10 mM Tris-HCl, 1 mM EDTA), respooled and redissolved in TE. Between 2 and 5 µg of Oregon R DNA was partially digested with *Mbo* 1 to an average length of 10-20 kb, treated with calf intestinal phosphatase and ligated to 2 µg *Bam* H1-digested EMBL4 bacteriophage DNA (11). Ligation mixes were packaged in extracts as described by Scalenghe *et al* (12). Nitrocellulose lifts of phage libraries were as described (13). Screening of libraries under reduced stringency conditions was described by Shilo and Weinberg (14). Phage DNA preparations were as described by Frischauf *et al* (11).

DNA analysis

Sequence determination of the 0.26 kb *Eco* R1-*Bam* H1 fragment of λGCR1 containing the GCR element was by the method of Maxam and Gilbert (15). The GCR element in λGCR6 was subcloned into pEMBL8(-) (16) as a 2 kb *Bam* H1-*Xho* 1 fragment. Sequential deletion of DNA flanking the *Bam* H1 site was performed by the method of Frischauf *et al* (17). The deletion clones were sequenced by the dideoxy method (18) using the M13 reverse primer (New England Biolabs) and by Maxam-Gilbert sequencing (15). Southern blotting was by the method of Southern (19) with modifications described (14). The GCR1 probe fragment was a 0.26 kb *Bam H1 -EcoRI* fragment. and the GCR6 probe a 0.6Kb MspI fragment. Both fragments are shown in Figure 1. *In situ* hybridization was as described by Pardue (20)

RNA analysis

Fly cages were as described by Elgin (21). RNA isolation from *w; b adh^{n4} Drosophila melanogaster* eggs, larvae, pupae and adults was by the method of Chia *et al* (22). Northern blotting to Biodyne A membranes used the methods provided by the manufacturers with UV crosslinking of RNA to the filter (23) additional to baking at 80° for 1h. Probing of the filters with DNA fragments was as described (24), using the same GCR fragment probes as for Southern blotting. Single stranded probing used an RNA probe containing the GCR6 MspI fragment described above synthesised from subclones of this fragment in pGem2 (Promega Biotec). The manufacturers conditions for synthesis and hybridization were used.

Table 1. Chromosomal location of GCR repeat sequences

Recombinant	Chromosomal location
GCR-1,-3,-4,-5,-7 -8,-9,-12,-22,-25 -28,-36,-42,-44,-47	100D/E
GCR-6	59C5
GCR-11,-14,-17	15E/F
GCR-16	68A
GCR-20	50D/E
GCR-21	79B
GCR-27,33	57B/C

Chromosomal location of *v myc*-homologous recombinants by *in situ* hybridisation to the polytene salivary gland chromosomes of Oregon R *Drosophila melanogaster* larvae.

RESULTS

Isolation of the GC-rich sequence repeat family

We isolated the GCR family fortuitously during a search for a *Drosophila* homologue to the avian *myc* oncogene. v *myc* hybridizes to a family of *Drosophila* sequences under low stringency conditions (14). We used these conditions to isolate recombinant phage sharing homology with v *myc* from *Drosophila melanogaster* genomic phage libraries. Twenty four v *myc* -homologous clones were isolated. These clones were grouped into seven classes on the basis of shared restriction fragments (data not shown). The location of each of the twenty four inserts on the *Drosophila* polytene chromosomes was determined by *in situ* hybridisation. Different members of each class hybridized to the same chromosomal location and each class hybridized to a different site (Table 1).

Two classes of recombinants (those hybridizing to 100 D/E and 59C5) which showed the strongest hybridization to radioactive v *myc* probe were selected for further study. Restriction maps of representative recombinants from these two classes are shown in Figure 1. There is no similarity between the two maps. The region homologous to v *myc* in the two phage DNAs was determined by Southern blot hybridization (19). The v *myc* homologous region of the 100 D/E clone, λGCR1 is entirely contained within a 0.26 kb *Eco* R1-*Bam* H1 restriction fragment and the corresponding region in the 59C5-homologous clone λGCR6 within 0.35 kb to the right of the *Bam* H1 site in Figure 1 (data not shown).

Sequence analysis of two GCR elements

To determine the basis of the homology between v *myc* and the λGCR phage DNAs, the sequence of the v *myc* -homologous regions in two of these phages, λGCR1 and λGCR6, was

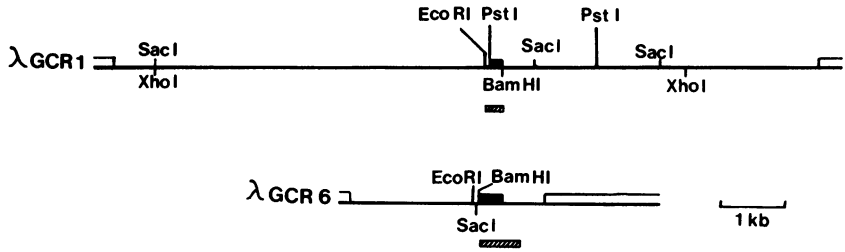

Figure 1 Restriction maps of GCR repeat-containing clones. Regions which share homology with v-*myc* are boxed (▬). Double lines represent vector sequences. GCR1 and GCR6 repeat-containing fragments used for radioactive probes are marked (▨). λGCR6 contains additional PstI sites outside the two shown which flank the GCR repeat. The λGCR6 insert does not contain any PstI sites.

determined (Figure 2). The two sequences share regions with unusual repetitious features. These comprise tandemly repeated GGX triplets where X is usually A, T or C, interspersed with different triplets, usually either CGPy or TTPy. The two regions are both approximately 76% GC-rich, an unusually high value for *Drosophila melanogaster* whose overall GC content is about 40%.

The nucleotide sequences of the v *myc* - homologous regions in GCR1 and GCR6 were compared with v *myc* (25), using a computer program written by D M J Lilley. Diagonal matrix comparison of these sequences shows no extended region of homology (greater than 20 bases with an allowed mismatch of 1 base per 10 nucleotide block; data not shown) Instead, short stretches of homology were seen. On comparison of the matrix printout with the nucleotide sequences, these patches correspond entirely to GGX oligomeric regions in GCR1, GCR6 and v *myc*. A comparison between GCR1 and GCR6 sequences gives a similar result. The sequence of v *myc* contains regions of highly GC-rich sequence (25) which are responsible for the hybridisation to GCR repeats. The GC-rich regions of *v myc* which give rise to this homology are not contiguous on v *myc* and are not all in the same reading frame (25). Furthermore, the predicted translation products of both inserts in all three reading frames show no detectable extended homology with *v myc* (data not shown). We therefore conclude that the GCR family does not encode *v myc* -related polypeptides.

Do GCR repeats encode proteins?

The strongly maintained triplet periodicity of GCR repeats prompts the question of whether these sequences encode any polypeptides. We approached this question in two ways. First, we inspected the nucleotide sequence of both repeats. If GCR1 and GCR6 repeats encode proteins they presumably use the same reading frame. We therefore, looked for open reading frames in both sequences. The only open frame through both GCR repeats is the GGX frame on the strands shown in Figure 2 (see Figure 3). This places the variable base for all three triplets

A

GCR I

ATTGATGAGATCACCAGTTTGATATGGAATTTGCACTAACCTCCACTACCTA

CTTTCAGTTCTTAGTGTTCGCGCTGGCCGCTCTTGCTGCAGCAGAGCCACC

ATCCGGATATAACTATCCC CGC GGC GGA GGT GGT GGT GGC GGC GGA

TTT GGA GGT GGC TTC GGC GGC GGA TTC GGA GGC GGA CTA GGA GGT

GGC GGC GGT GGT GGC GGC GGC GGC TAC CAG GCT GTG AGC GGC

GGC TTC CAG ACG TCC GAG GGC CAG AAC GAA

GCR 6

AACCGAAATTCAAATGCGGATTTTTTCTATTGCAGGATTCAGTCA CGC GGT

GGT GGC GGC GGC GGC GGA GGA GGA GGA GGT GGA TTC CGT GGG

CGT GGT GGC GGC GGC GGC GGA GGA GGC GGC GGA TTT GGA GGC

GGC CGT GGA CGC GGT GGC GGC GGC GAT CGC GGT GGA CGT GGA

GGA TTC GGT GGT GGC CGT GGA GGA GGT GGT CGC GGT GGT GGT

GGC GGC GGC GGC CGT GGT GCC TTC GGA GGA CGT GGC GGC GGC

GGT GGT CGC GGT GGC GGC GGC CGT GGA GGC GGT GGT CGC GGC

GGA GGA GGA CGT GGT GGT GGT GCT GGC GGC TTC AAG GGC

GGCAAGACCGTCACTATCGAGCCGCATCGTCACGAGGGAGTGTTCATTGCC

CGCGGAAAGGAGGACGCTCTGGTCACCAGGAACTTTGTACCTGGATCC

B

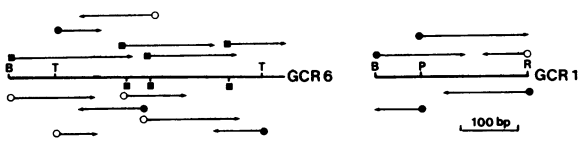

Figure 2A; Nucleotide sequence of GCR repeat sequences. The GGX repeats are boxed. Interspersed TTPy and CGPy triplets are underlined.
2B; Sequencing strategy. Arrows refer to regions sequenced from (●) 5' end labelled or (o) 3' end-labelled fragments by Maxam-Gilbert or (■) the dideoxy method.
B; Bam H1. T; Taq 1. P; Pst 1. R; Eco R1. (■); Deletion site.

(GGX, TTPy and CGPy) as the third base. The third base in the triplet code, the wobble base, is often irrelevant in determining the encoded amino acid (26). This is true in all three cases here, the triplets GGX, TTPy and CGPy encode glycine, phenylalanine and arginine respectively. Thus, the variability in sequence of the triplets is not reflected in any heterogeneity at the amino acid level. As a further test for whether the GGX frame is a plausible message, we compared the codon usage in the open reading frames containing GCR1 and GCR6 with a compiled codon usage table for *Drosophila melanogaster* genes (provided by Dr Michael Ashburner). We only considered the three amino acids glycine, phenylalanine and arginine because these are the only ones represented in sufficient numbers to give a reasonable statistical significance to the data.

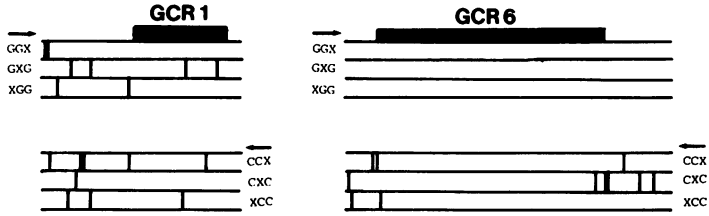

Figure 3 Open reading frames in GCR1 and GCR6 repeats. The GCR repeat is marked (▬). Stop codons are marked by vertical lines.

Table 2 shows that the codon usage of GCR1 and GCR6 for these three amino acids is broadly similar to that of their average usage in *Drosophila*. When we compared the codon usage profiles for the other five possible reading frames with the compiled usage, none gave a reasonable correlation (data not shown). By sequence criteria, therefore, the GGX reading frames of both GCR repeats appear to be plausible messages.

GCR repeats are transcribed into polyadenylated RNAs

Our second approach to the question of whether GCR repeats encode proteins was to ask whether they are represented in polyadenylated RNA in *Drosophila melanogaster*.

Table 2 Codon usage of GCR repeats.

		PERCENTAGE USE OF CODON		
		Drosophila	GCR1	GCR6
Phe	UUU	26 (146)	38 (3)	28 (2)
	UUC	74 (413)	62 (5)	72 (5)
Arg	CGU	19 (168)	0 (0)	55 (11)
	CGC	45 (380)	100 (1)	32 (6)
	CGA	10 (92)	0 (0)	5 (1)
	CGG	12 (107)	0 (0)	5 (1)
	AGA	5 (49)	0 (0)	0 (0)
	AGG	9 (76)	0 (0)	5 (1)
Gly	GGU	24 (337)	23 (7)	28 (23)
	GGC	42 (602)	52 (16)	38 (31)
	GGA	29 (407)	26 (8)	32 (26)
	GGG	5 (71)	0 (0)	1 (1)

Numbers of times each codon appears in the respective sequences are bracketed.

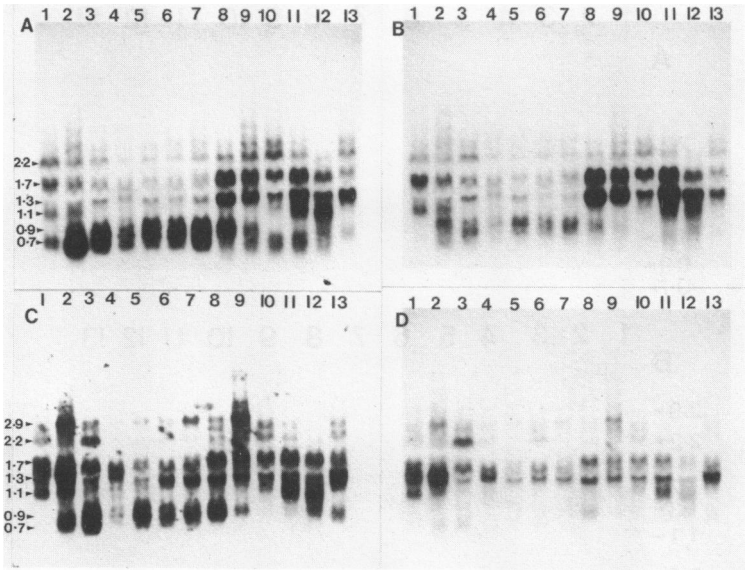

Figure 4 GCR repeats are transcribed during *Drosophila* development. 10μg samples of polyadenylated RNA, isolated from different stages in development of *Drosophila melanogaster* were probed with GCR repeat sequences. 4A; GCR1 probe. The blot was washed in 0.5 X SSC at 50°. Lane 1; Embryos. Lane 2 ; First instar larvae. Lane 3; First/second Instar larvae. Lane 4; Second Instar larvae. Lane 5; Second/third instar larvae. Lane 6; Third instar larvae. Lane 7; Climbing third instar larvae. Lane 8; Prepupae. Lane 9; Young Pupae. Lane 10; Mid Pupae. Lane 11; Mid-late pupae. Lane 12; Late pupae. Lane 13; Adults. Figure 4B;. The blot in Figure 4A was washed in 0.03 X SSC at 50°. Figure 4C; The blot in Figure 4B was stripped of radioactive probe, reprobed with the GCR6 repeat then washed in 0.5 X SSC at 50° . Figure 4D; The blot in 4C was washed in 0.03 X SSC at 50°. All exposures are 16 hrs with intensifying screen. The sizes in kb of prominent RNAs are marked.

Polyadenylated RNA from different developmental stages of *D. melanogaster* was subjected to Northern blot analysis using GCR1 and GCR6 repeat probes (Figure 4; reference 24). Under low stringency washing conditions both probes hybridize to a complex pattern of transcripts (Figure 4A and 4C). There are at least 6 bands which are homologous to GCR1 (Figure 4A). Most of these are only prevalent at certain stages of development. Two small RNAs 0.9 kb and 0.7 kb long are found mainly in larvae while 1.7 kb and 1.3 kb RNAs are expressed mainly in pupae, a 1.3 kb species is also prevalent in adults (Figure 4A). The 0.9 kb and 0.7 kb RNAs are not strongly homologous to GCR1 as high stringency washing of the filter largely removes the GCR probe hybridization to these RNAs (Figure 4B). GCR6 similarly hybridizes to many RNAs, most of which are of similar size to GCR1-specific RNAs. An exception is a 2.9 kb GCR6-specific RNA which is not homologous to GCR1 (Figure 4A and 4C).

The data in Figure 4 do not distinguish which of these RNAs are encoded by GCR1 and

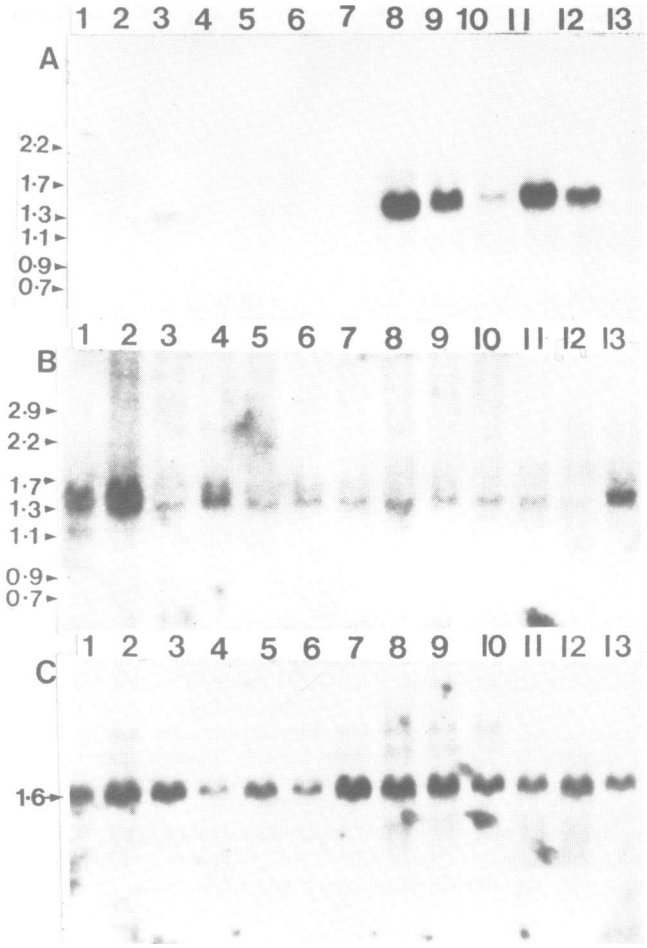

Figure 5 λGCR and λGCR6 encode 1.3kb RNAs with different developmental specificities. The Northern blot in Figure 4 was stripped of bound probe and reprobed with radioactive fragments adjacent to the GCR repeats in λGCR1 and λGCR6. 5A; the probe was a 1.6kb Pst 1 fragment to the immediate left of the Pst1 fragment containing the GCR1 repeat. (Figure 1). The blot was washed in 0. 03 X SSC at 55°; 16hr exposure. 5B; the same Northern blot was stripped and reprobed with a 0.65 kb Sau 3A fragment to the immediate right of the GCR6 repeat in Figure 1. Washing conditions were the same as for Figure 5A; 4 days exposure. Figure 5C; The same Northern blot was stripped and reprobed with radioactive pDm ras64 (27); 4 day exposure. The mobilities of the GCR repeat containing RNAs in Figure 4 are marked.

GCR6 as even after high stringency washing both probes hybridise to several RNAs (Figure 4B and 4D). Under these conditions GCR1 and GCR6 cross hybridize only very weakly (data not shown), suggesting that these transcripts are more closely related to the particular repeats which detect them than the GCR1 and GCR6 repeats are to each other. To determine which RNAs are

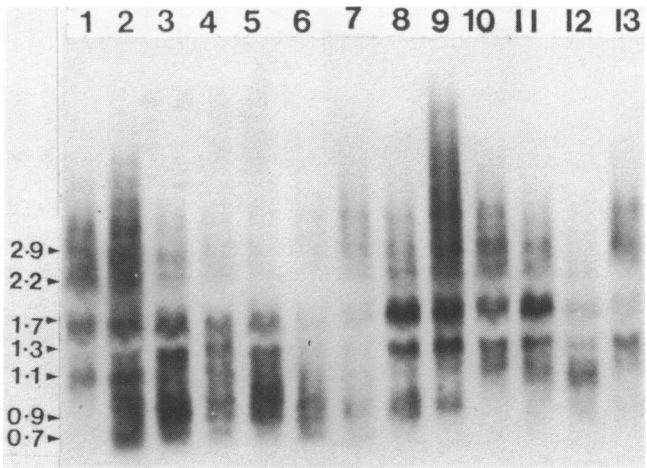

Figure 6. GCR-repeat-homologous RNAs contain the GGX strand. The Northern blot in Figure 5 was stripped of bound probe and reprobed with a single stranded RNA probe containing the GCR6 repeat. The blot was washed in 0.03 X SSC + 0.05% SDS at 60°. 16 hr. exposure. The mobility of the GCR repeat-containing RNAs in Figure 4 are marked.

encoded by GCR1 and GCR6, the Northern blot in Figure 4 was stripped and reprobed with flanking fragments which do not contain GC-rich repeats. A 0.85 Kb Pst1 fragment lying to the left of the GCR1 repeat in Figure 1 hybridizes specifically to the 1.3kb RNA which is expressed at high levels in pupae and at low levels in larvae (Figure 5A). A 0.65 kb Sau3A fragment to the immediate right of the GCR6 repeat in Figure 1 hybridizes to two different RNAs. One is a 1.3kb RNA which is expressed throughout development but is present in larger amounts in adults, embryos and first instar larvae (Figure 5B). The other is slightly larger and is again prevalent in adults, embryos and young larvae. Probing the same blot with a *Drosophila ras* probe whose transcription does not vary during development (27,28) shows that the apparent fluctuations of GCR6 and GCR1 transcription during larval (lanes 2-7) and pupal (lanes 8-12) development respectively are due to variable sample loading (Figure 5C).

To determine which strand of GCR repeats is expressed in RNA we reprobed the Northern blot used in Figure 5 with single stranded RNA probes containing the λGCR6 repeat (Figure 6). A single stranded RNA probe complementary to the GGX strand gives a very similar result to that obtained with the double stranded probe (Figure 6 and Figure 4). This suggests that the majority if not all of the RNAs hybridizing to the GCR6 repeat contain the GGX strand. Probing with the complementary strand gave no signal following a similar exposure.

Two GCR repeats are not transposable elements

To determine whether GCR1 and GCR6 repeats are transposable, Southern blots of restriction digested DNA from two *Drosophila melanogaster* strains (Oregon R and Canton S)

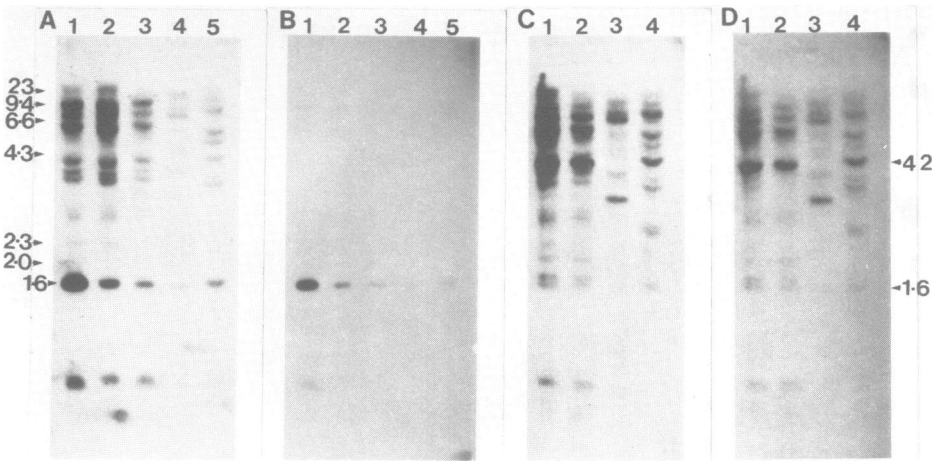

Figure 7 GCR sequences are not transposable. *Pst* 1- digested *Drosophila* DNAs (5μg) were subjected to Southern blot analysis. Figure 7A; GCR1 probe. The blot was washed in 2xSSC at 70°. Lane 1; Oregon R DNA plus 1ng *Pst* 1-digested λGCR1 DNA. Lane 2; Oregon R DNA. Lane 3; Canton S DNA. Lane 4; *Drosophila simulans* DNA. Lane 5, *Drosophila teissieri* DNA. The mobilities of size markers (in kb) and the1.6kb *Pst* 1 fragment containing the GCR1 repeat is marked. Figure 7B; The blot in Figure 7A was washed in 0.2 xSSC,70°. Figure C; GCR6 probe; Lane 1; Oregon R DNA. Lane 2; Canton S DNA. Lane 3; *Drosphila simulans DNA*. Lane 4; *Drosophila teissieri* DNA. Figure 7D; the blot in Figure 7C was washed in 0.2 xSSC 70°. The 4.2 kb GCR6-homologous band is mentioned in the text.

were probed with GCR1 and GCR6 fragments (Figure 7). Low stringency probing shows a complex set of sequences homologous to GCR1 and GCR6 (Figure 7A and 7C, lanes 1 and 2). The pattern of bands detected by the two probes are related but not identical. Several bands are shared, many are considerably more intense with one probe than the other. There is little or no difference between the Oregon R and Canton S patterns using either probe. These strains of flies were independently isolated from different regions of the USA over forty years ago (29) and share few common locations for *copia* -like transposable elements (30). Therefore GCR1, GCR6 and the related sequences revealed by them in Figure 7 are not transposable.

To further assess the integrity of GCR repeats on a considerably greater time scale, digests of DNAs isolated from other *Drosphila* species were probed in a similar manner. The closely related but distinct species *D. simulans* and *D. teissieri* also contain GCR repeat-homologous sequences (Figure 7A and 7C). Here, however, there is a greater divergence in the restriction pattern. It is not easy to determine from Figures 7A and 7C whether any bands are shared between these species and *D. melanogaster*. However, by increasing the stringency of the blot washing, the conservation of the *Pst* 1 restriction fragment containing GCR1 can be addressed. A higher stringency wash (0.5xSSC, 70°) of the blot in Figure 7A reveals two GCR1-homologous bands only (Figure 7B). The larger of these bands comigrates with the 1.6kb *Pst* 1

fragment containing the GCR repeat in λGCR1 (Figure 7B, lane 1). This band is also present in a *Pst* 1 digest of *D, teissieri* and *D. simulans*. Therefore this fragment has probably not suffered any major internal sequence rearrangement since the divergence of these different sub species of *Drosophila* at least one million years ago (31).

A high stringency wash of the GCR6-probed blot produced a less clear-cut result (Figure 7D). We do not know the size of the genomic *Pst* 1 fragment containing the GCR6 repeat as there is no site for this enzyme in the 3kb insert of λGCR6 (Figure 1). It may be the 4.2 kb band marked in Figure 7D as this has the strongest homology under high stringency washing. This band is shared by both *D. melanogaster* strains and *D. teissieri* but is absent from *D. simulans*. It is therefore possible that GCR6 is transposable at an extremely low frequency. However, we think it more likely that a mutation unrelated to the presence of the GCR6 sequence caused this change.

DISCUSSION

This paper describes the isolation of a novel class of interspersed sequence repeat from *Drosphila*. The method of isolation used low stringency hybridization to the vertebrate *myc* oncogene but the two clones which were sequenced are not true *myc* homologues. What function if any do these repeats have? Both GCR repeats have open reading frames which could encode oligo glycine-rich polypeptides. The triplet periodicity of these open reading frames is rigidly maintained in both repeats, even in the regions separating the GGX triplets. The triplets show strong conservation in the first and second bases (when read in the GGX frame) but considerably less in the third base. This is true not only for the GGX triplets but also for the other triplets separating them. Almost all of the latter are either CGPy or TTPy (the underlined bases in Figure 1). This lack of conservation is consistent with selection at the amino acid level since both CGC and CGT encode arginine and both TTT and TTC encode phenylalanine. Furthermore, the codon usage for these three amino acids (glycine, arginine and phenylalanine) broadly matches the corresponding values for a compilation of *Drosophila* genes. Lastly, the strand containing the GGX frame is expressed as polyadenylated RNA. It is therefore likely that GCR1 and GCR6 repeats are expressed as oligo glycine tracts in proteins.

Are these repeats a new class of selfish or parasitic DNA? We think that this is unlikely. GCR repeats are not transposable, in contrast to the majority of interspersed repeats in this organism which are mobile and do not encode any known essential function for the host. Another possibility is that they are contained in genes which are involved in the correct functioning of the fly. We have two reasons to believe this to be the case. First, they are expressed in polyadenylated RNAs. A variety of polyadenylated transcripts which are homologous to GCR repeats are found in the developing *Drosophila* organism. These transcripts are not confined to any particular stage of development. Some are present exclusively in larval stages, others appear at different times and yet others display little or no developmental

programming. Some of the RNAs are quite abundant. The signal levels for the most abundant GCR1-homologous RNAs is much stronger than that obtained when using a *Drosophila ras* gene probe which constitutes very roughly 0.02% of the polyA$^+$ RNA throughout development (references 27,28; Figure 5C).

Our second reason for confidence that GCR repeats encode functions important to *Drosophila* derives from the existence of the GGX motif in the open reading frames of several known genes. The most extreme example of this is the *grp-1* gene of petunia (32). The predicted amino acid product of *grp-1* contains 67% glycine with the central part comprised of seven repeats containing 77% glycine interspersed occasionally with Phe, Ala, Leu or His. This bizarre protein is believed to be a cell wall structural protein in the plant. Four other examples of oligo GGX sequences encoding glycine in *Drosophila* genes are known. Three of these genes, *Ultra bithorax* (*Ubx*) *Deformed* (*Dfd*) and *female sterile (1) homeotic* (*fshl*) are involved in the development of the fly's three dimensional structure (33-37). The fourth, p9, is of unknown genetic function but the predicted amino acid sequence bears strong similarities to rat helix destabilizing protein (35), a protein which is involved in hnRNP particles. Another GGX region in Ubx encodes oligo alanine (37).

These examples raise the exciting possibility that GCR repeats encode oligo glycine, and possibly also oligo alanine, domains which are involved for the function of a variety of developmentally important *Drosophila* genes. In support of this hypothesis is the observation that genes which are crucial to *Drosophila* development have been found to contain other repeated sequence motifs. Each repeat is present in some but not all of these genes. The homoeo box is present in the homoeotic genes *Ultrabithorax* and *Antennapaedia* (33,34) and in segmentation genes such as *fushi tarazu* and *paired* (34,38). Another repeat, *opa*, resembles the GCR repeat more closely in that it is internally repetitious, unlike the homoeo box and consists primarily of CAPu triplets encoding oligoglutamine (9). *opa* is found in the homoeotic *Antennapaedia* and *Deformed* genes (34), in the *fushi tarazu* segmentation gene (34) and in *Notch*, a gene important in *Drosophila* neurogenesis (9). It is not known whether all *opa* -containing genes are involved in the development of the fly but it is tempting to speculate that both *opa* and the GCR repeat fulfil important roles in this process. Clearly, further study of the genes containing these sequences will be necessary to establish whether such speculation is justified. If this should turn out to be the case, such repeats should prove to be extremely useful in the isolation of previously uncloned developmentally important genes, as has already been achieved with the homoeo box (38). In this context it is of interest to note that the location of GCR6 in the *Drosophila* genome (59C5) corresponds closely to the position of *twisted*, a gene affecting dorsal-ventral pattern in the developing embryo (Table 1, reference 39). Additionally, GCR21 is positioned within the limits of the defined chromosomal region containing another pattern gene, *Polycomb* (40). The other localizations in Table 1 do not correspond to any known genes involved in pattern formation.

It is interesting that the preferred amino acid interspersed between the oligo glycine repeats varies between these genes. Phenylalanine is usually present, the exceptions are *Deformed* and *fsh(1)* which contain tyrosine and serine. In GCR6 and helix destabilizing protein, arginine is also found and *grp 1* possesses histidine. The function of these glycine rich regions remains a matter for speculation. It has been variously proposed that such a sequence defines a region of little or no secondary structure, constituting a 'hinge' region (33) or alternatively a rigid pleated configuration (32). Clearly, further experiments are necessary to resolve this question.

ACKNOWLEDGEMENTS

During the preparation of this manuscript, we learned that Haynes *et al* have independently isolated a GGX repeat (which they originally termed the *pen* repeat) in the *fsh(1)* gene (35). Robert Weinzierl and Mike Akam have also used a GGX repeat from the Ubx gene to isolate cDNA clones containing oligo glycine -encoding regions (in preparation). We are grateful to both groups for communicating these results prior to publication. We also thank Sheena Pinchin for help with *in situ* hybridizations, Susan Parkhurst for advice on developmental Northern blots, David Lilley and Mike Waterfield for computer sequence analysis and Kathy Howe, Alan Sneddon and Mike Smith for technical help.

*To whom correspondence should be addressed

REFERENCES

1. Brutlag, D. L. (1980) Ann. Rev. Genet. 14 121-144.
2. Spradling, A. C. and Rubin, G .M.(1981) Ann. Rev. Genet. 15 219-264.
3. Finnegan, D. J .(1985) Int. Rev. Cytol. 93 281-326.
4. Doolittle, W.F. and Sapienza, C. (1980) Nature 284, 601.
5. Orgel, L.E. and Crick, F.H.C. (1980) Nature 284, 604.
6. McGinnis, W. Levine, M .S., Hafen, E., Kuriowa, A. and Gehring, W.J.(1984) Nature 308 428-433.
7. Laughon, A. and Scott, M.P. (1984) Nature 310 25-31.
8. McGinnis, W., Hart, C. P., Gehring, W. J. and Ruddle, F. H.(1984) Cell 38 675-680.
9. Wharton, K.A., Yedvobnick, B., Finnerty, V. G. and Artavanis-Tsakonas, S. (1985) Cell 40 55-62.
10. Kidd, S., Lockett, T. H. and Young, M. W. (1983) Cell 34 421-433.
11. Frischauf, A. M., Lehrach, H., Poutska, A.and Murray, N .(1983) J. Mol. Biol. 170 827-842.
12. Scalenghe, F., Turco., E, Edstrom., Pirrotta, V. and Melli, M. (1981) Chromosoma (Berlin) 82 205-216.
13. In 'Molecular Cloning ' (1982) Eds. Maniatis, T. Fritsch, E. and Sambrook, J. Cold Spring Harbor Laboratory.
14. Shilo, B-Z. and Weinberg, R.A.(1981) Proc. Natl. Acad. Sci. USA 78 6789-6792.
15. Maxam, A. M.and Gilbert, W. (1980) Meth. Enzymol. 65 499-560.
16. Dente, L. Cesarini, G. and Cortese, R.(1983) Nucleic Acids Res. 11 1645-1655.
17. Frischauf, A. M. Garoff, H. and Lehrach, H .(1980) Nucleic Acids Res. 8 5541-5549.
18. Sanger, F., Nicklen, S. and Coulson, A. R.(1977) Proc. Natl. Acad. Sci. USA 74 5463-5467.
19. Southern, E.M. (1975) J. Mol. Biol. 98 503-517.

20. Pardue, M-L., (1985) In Nucleic Acid Hybridisation, a practical approach, Eds. B. D. Hames and S. J .Higgins, IRL Press, Oxford and Washington D. C.
21. Elgin, S. R. (1978) In 'The Genetics and Biology of *Drosophila* Vol. 2a Eds. M. Ashburner and T. R. F. Wright, Academic Press.
22. Chia, W., Savakis, C. , Karp, R., Pelham H. and Ashburner, M. (1985) J. Mol. Biol. 186 679-688
23. Church, G. M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81 1991-1995.
24. Wahl, G.M,. Stern, M. and Stark, G.R. (1979) Proc. Natl. Acad. Sci. USA 76 3863-3687..
25. Alitalo, K.Bishop, J. M., Smith, D. H., Cher, E.Y., Colby, W. W. and Levinson, A. D. (1983) Proc. Natl. Acad. Sci. USA 80 100-104.
26. Crick, F. H .C. .(1966) J. Mol. Biol. 19 548-555.
27. Mozer, B. Marlor, R., Parkhurst, S. and Corces, V. (10985) Mol. Cell Biol 5 885-889.
28. Lev, Z., Kinchie, Z., Hessel, R. and Seger, O. (1985) Mol. Cell Biol. 5 1540-1542.
29. Bridges, C .B. and Brehme, K. F. (1944) Carnegie Institution, Washington Publication 552.
30. Strobel, E., Dunsmuir, P. and Rubin, G. M. (1979) Cell17 429-439.
31. Lemeunier, F., David, J.R., Tsacas, L. and Ashburner, M. (1986) In 'The Genetics and Biology of DrosophilaVol3e, 146-256. Academic Press (London).
32. Condit, C. M. and Meagher, R .B (1986) Nature 323, 178-179.
33. Beachy, P. A., Helfland, S. L. and Hogness, D .S .(1985) Nature 313 545-551.
34. Laughan, A., Carroll, S. B., Storfer, F. A., Riley, P. D. and Scott, M.P. (1985) Cold Spring Harbor Group Quart. Biol. 50 253-262.
35. Haynes, S. R., Rebbert, M. L., Mozer, B.A,. Forquignon, F. and Dawid, I.B. (1987) Proc. Natl. Acad. Sci. USA (In Press).
36. Digan, M.E., Haynes, S. R., Mozer, B. A., Darvid, I.B., Forquignon, F. and Gans, M. (1986) Developmental Biol. 114 161-169
37. Weinzierl , R. and Akam, M. (1987) Genes and Development (In Press).
38. Bopp, D., Burri, M., Baumgartner, S., Frigeris, G. and Noll, M. (1986) Cell 47, 1033-1040.
39. Nusslein-Volhard, C., Wieschaus, E. and Kluding, H. (1984) Roux's Arch. Dev. Biol. 193 267-282.
40. Jurgens, G., Wieschaus, E., Nusslein-Volhard, C. and Kluding, H. (1984) Wilhelm Roux's Arch. Dev. Biol. 193 283-295.