Discordant expression and variable numbers of neighboring GGA- and GAA-rich triplet repeats in the 3' untranslated regions of two groups of messenger RNAs encoded by the rat polymeric immunoglobulin receptor gene

Katherine S. Koch¹, Anatoly S. Gleiberman^{1,3}, Takashi Aoki^{1,4}, Hyam L. Leffert^{1,2,*}, Anna Feren⁵, Albert L. Jones⁵ and Eric J. Fodor⁵

¹Department of Pharmacology and ²Center For Molecular Genetics, School of Medicine, University of California, San Diego, La Jolla, CA 92036-0636, USA, ³Laboratory of Immunochemistry, Institute of Carcinogenesis, Russian Cancer Research Center, 24 Kashirskoye sh., Moscow 115478, Russia, ⁴Department of Biochemistry, Faculty of Pharmaceutical Sciences, Health Sciences University of Hokkaido, Ishikari-Tobetsu, Hokkaido 061-02, Japan and ⁵Cell Biology and Aging Section, Department of Veterans Affairs Medical Center, Departments of Medicine, Anatomy and the Liver Center, University of California, San Francisco, San Francisco, CA 94121, USA

Received February 22, 1995; Accepted February 22, 1995

GenBank accession nos[†]

ABSTRACT

An unusual S1-nuclease sensitive microsatellite (STMS) has been found in the single copy, rat polymeric immunoglobulin receptor gene (PIGR) terminal exon. In Fisher rats, elements within or beyond the STMS are expressed variably in the 3' untranslated regions (3'UTRs) of two 'Groups' of PIGR-encoded hepatic mRNAs (plg-R) during liver regeneration. STMS elements include neighboring constant regions (a 60-bp d[GA]-rich tract with a *chi*-like octamer, followed by 15 tandem d[GGA] repeats) that merge directly with 36 or 39 tandem d[GAA] repeats (Fisher or Wistar strains, respectively) interrupted by d[AA] between their 5th–6th repeat units. The Wistar STMS is flanked upstream by two regions of nearly contiguous d[CA] or d[CT] repeats in the 3' end of intron 8; and downstream, by a 283 bp 'unit' containing several inversions at its 5' end, and two polyadenylation signals at its 3' end. The 283 nt unit is expressed in *Group 1* plg-R mRNAs; but it is absent in the *Group 2* family so that their GAA repeats merge with their poly_A tails. In contrast to genomic sequence, GGA triplet repeats are amplified ($n \geq 24$ –26), whereas GAA triplet repeats are truncated variably ($n \leq 9$ –37) and expressed uninterruptedly in both mRNA *Groups*. These results suggest that 3' end processing of the rat PIGR gene may involve misalignment, slippage and premature termination of RNA polymerase II. The function of this unusual processing and possible roles of *chi*-like octamers in quiescent or extrahepatic tissues are discussed.

INTRODUCTION

Animal cell genomes contain repetitive sequences that regulate gene expression (1). Apart from pseudogenes (2), these sequences include populations of major and minisatellites and poly-dispersed simple tandemly repeated microsatellites (STMS; 3,4). STMSs are polymorphic (with repeat unit sizes of 2–5 nucleotides [nt]), abundant (with $\geq 2 \times 10^5$ copies per genome), implicated in recombination (5) and associated with unusual DNA structures including cruciforms, palindromes, 'knots' and left-handed helices (6).

STMSs have been linked etiologically with human neurological disorders (7) and with gastrointestinal malignancies associated with familial colorectal cancer (8). In several cases, genetic etiologies correlate with unstable transmission of amplified stretches of constitutive tandem triplet repeats, and with subsequent expression of aberrant repeats in the 5' UTRs, coding regions, or 3' UTRs of disease-associated mRNAs (9–12). To our knowledge, no animal models are available for investigating mechanisms generating trinucleotide repeat diversity in such genes, or in the STMSs of genes unrelated to the inherited disorders identified thus far.

Our attention was drawn to these problems when, during studies of normal rat liver regeneration (13), we observed variable arrays of tandem GAA repeats in the 3' untranslated regions (3' UTRs) of several hepatic cDNAs putatively associated with 'plg-R' mRNAs (14) encoded by the polymeric immunoglobulin receptor gene (PIGR; 15,16). The PIGR gene is a member of the immunoglobulin (Ig) superfamily (17); it encodes a 120 kDa membrane glycoprotein (18) that transports dimeric plasma IgA and IgM transcytotically to mucosal spaces bordered by epithelial cells including rat hepatocytes (19,20). Such

* To whom correspondence should be addressed

[†]U07886, U02506, L22660, U01145, U00762, U00763, U08271, U08272, U08273, U17366 and U17367

transport is essential for animal survival as it mediates mucosal protection against extracellular and intracellular pathogens (21).

Variable STMSs in rat pIg-R mRNAs are unexpected and puzzling since rat hepatocytes produce only one 4 kb pIg-R mRNA (22). In contrast, rabbits produce two STMS-free pIg-R mRNAs by alternative splicing of an apparently single gene with a 'repetitive element' in its 3' UTR (23) and, although the human PIGR 3' UTR sequence is defined incompletely (16), an STMS occurrence at its 3' terminus may well be inferred from known unstable rearrangements, duplications, insertions, deletions, oncogene-associated breakpoints and gastrointestinal malignancies (24) at or near its chromosomal locus (1q31-q32; 25).

To further investigate provocative aspects of PIGR expression, we present evidence here of an S1-nuclease sensitive STMS in a single-copy rat PIGR gene that encodes GGA- and GAA-triplet repeats from its 3' UTR. Remarkable differences between genomic and cDNA sequences spanning the STMS repeat regions suggest that unusual processing of the PIGR terminus or its transcripts generates 3' UTR diversity among pIg-R mRNAs found in regenerating hepatic tissue.

MATERIALS AND METHODS

Animal surgery and RNA isolation

Male Fisher 344, Wistar or Sprague Dawley rats (180–220 g; Simonson, Gilroy, CA) were used for surgical procedures and isolation of hepatic RNAs according to standard procedures (13).

Construction and screening of Fisher rat liver cDNA libraries

Two cDNA libraries were constructed from different animals. Size-selected polyA⁺ RNAs were obtained from pooled regenerating liver lobes of each rat 24 h after 67% hepatectomy (13,14). The cDNAs of these RNAs were ligated into λ -Uni-ZapXR vectors (cat. #200400; lot #UC116) following the protocols of the manufacturer (Stratagene, La Jolla, CA). Both libraries were amplified (1×10^{10} p.f.u./ml; ≈ 1.5 complexities) without subsequent background change. Libraries were screened with a 1245 bp *NotI* random primer ³²P-labelled double stranded (ds)-cDNA probe containing the entire human cyclin D1 coding sequence (13). Hybridizations were performed at 60°C (0.5 $\times$ SSC, 3.5% SDS; 26). After four cycles of plaque-purification, plasmids carrying 0.3–4.0 kb inserts were recovered from 21 of 48 positive clones.

Isolation of genomic clones encoding the Wistar rat PIGR gene

To obtain genomic clones encoding the rat PIGR gene, a pIg-R 5' cDNA probe was prepared by reverse transcriptase-polymerase chain reaction (RT-PCR) amplification of Sprague Dawley rat liver RNA. Five pmol of a single-stranded (ss) antisense pIg-R oligodeoxynucleotide primer, 3'-d[*cggcaccagtatttcg*]-5' (spanning pIg-R ss-mRNA nt positions 227–243), were annealed to 5 μ g liver RNA (in 20 μ l containing 0.5 mM dNTPs, 50 mM Tris-HCl, pH 8.5, 8 mM MgCl₂, 30 mM KCl, 1 mM DTT); extension reactions with 25 U AMV RT (Boehringer Mannheim) were performed at 42°C for 1 h. The nt numbering scheme is based upon the published rat pIg-R cDNA sequence (15). Although nucleotides 1–20 of this sequence are pUEx cloning vector, we have not

changed published notation (GenBank accession number X15741). After inactivation at 95°C for 5 min, 20% of the cDNA product was combined with additional antisense and sense primer, 5'-d[*gccgcaacctcttggaagcc*]-3' (spanning ss-mRNA nucleotide positions 46–66) at final concentrations of 50 μ M. PCR was then performed in 50 μ l containing 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1 mM MgCl₂, 200 μ M dNTPs and 2 U of *Taq* polymerase (Promega) for 35 cycles, each consisting of: 95°C, 40 s; 42°C, 45 s and 70°C, 40 s using a thermal cycler (MJ Research). Reaction products were separated on a 7% polyacrylamide gel in 1 $\times$ TBE and the 198 bp PCR cDNA band excised and subcloned in pGEM-T (Promega). The identity of the 198 bp insert was confirmed by DNA sequencing, isolated preparatively and ≈ 50 ng radiolabelled by random priming (Amersham) using [α -³²P]dCTP (≥ 3000 Ci/mmol, NEN) (26).

An adult male Wistar rat liver genomic library (Clontech; supplied by J. Kriger, Stanford University) was screened in duplicate by standard methods (26) with the 198 bp 5' cDNA PCR probe or with a random primer labelled 2.3 kb *Bam*H1 pIg-R cDNA probe (nucleotides 958–3269). Positive plaques (λ vector EMBL3) were carried subsequently through two additional rounds of plating and selection. Candidate clones were purified from a clear lysate as described (26). Fragments of genomic regions, 17 kb λ 10IgR and 14.4 kb λ 2IgR, were subcloned into pGEM-7Z[-] (pGEM, Promega).

Manual sequencing and analyses of cDNA and genomic clones

Plasmid cDNAs were purified by two cycles of centrifugation through CsCl-EtBr gradients. Supercoiled cDNAs (40 μ g) were linearized by digestion with restriction enzymes at sites far removed from regions to be sequenced, and repurified on agarose minigels (26). The [+] and [-] strands of ds-DNA inserts from six plasmids (pBluescriptIIISK[-]; 'pBS': clones c.1, 2 and 3 [library #1]; c.15, 9, and 7 [library #2]) and from their truncated subclones (c.15.5 *Bg*III/*Pvu*II, c.15.16 5'/*Bg*III and c.7.3 5'/*Pvu*II fragments) were sequenced entirely using Stratagene 'reverse' and 'forward' M13 -20 primers (5'-d[*ggaaacagctatgaccatg*]-3' and 3'-d[*tgaccggcagcaaatg*]-5', respectively) and [α -³⁵S]dATP, as outlined in *SEQUenase* or *TAQUenase* kits (USB). Resolution and gel compression problems were solved by varying the conditions of labelling and termination with respect to time and temperature (5–10 min and 70–78°C), employing appropriately substituted dGTP analogs (dITP or 7-deazaGTP), and varying gel running times (2–14 h). Sequence data were plotted, aligned, analyzed and downloaded using DNAsis (Hitachi), MACAW (27), BLAST and RETRIEVE software (NCBI), respectively. Plus or [-] strands of ds-DNA genomic inserts from appropriate λ 2IgR subclones (pBS-ESPS5 and pBS-ESPS3; or, pGEM-ESBB and pGEM-A) were purified, linearized and analyzed as described above.

Northern and RT-PCR analyses of Fisher rat liver mRNAs

For Northern analyses, polyA⁺ or total RNA samples from individual rats were subjected to HCHO-agarose gel electrophoresis as described elsewhere (13). Antisense ss-oligodeoxynucleotide probes, designed with OLIGO software (28), were constructed and purified by the UCSD CMG Oligodeoxynucleotide Synthesis Facility, and 5' end-labelled with [γ -³²P]ATP (6000 Ci/mmol, Amersham) using T4 polynucleotide kinase (26) or 'Ready-to-Go' DNA labelling kits (Pharmacia).

The sequences of these probes were (5'→3') d[gattccaggctg-gccttacactcactaggt] (common to both mRNA 'Groups'), d[gggtgaagtacagagctcagaggaacag] (specific for Group 1 mRNAs) and d[tcc]_{n=11} (specific for GAA repeats), respectively. The 2.3 kb *Bam*HI pIg-R cDNA hybridization probe (specific for downstream coding and upstream 3' UTR regions in both Groups of mRNAs) was labelled with [α -³²P]dCTP by random priming. Prehybridization, hybridization, washing, Ambis scanning and autoradiography conditions were described previously (13); *T_m* and *A*₂₆₀ values were calculated with OLIGO software.

For RT-PCR analyses, ss-DNA primers designed with OLIGO software were constructed by the UCSD Cancer Center Oligodeoxynucleotide Facility (UCSDCC). The sequences of primers UCSDCC#1, #2, #3 and #5 were (5'→3') d[ttcagtcagcacaat-agc], d[tttttttgggtggcaatc], d[ttttttttcttctct] and d[ggaagg-gaaaggtgagac], respectively. RT-PCR was performed in two stages. First, 1st strand cDNA synthesis was performed with 10 µg RNA (from a regenerating liver 2.5 h post-67% hepatectomy) and 10 pmol ss-antisense primers #2 or #3 (specific for Group 1 or Group 2 pIg-R mRNA families) using 200 U SuperScript RNaseH⁻ RT according to the manufacturer's protocol (Gibco-BRL). Secondly, PCR products were produced in 50 µl, under 20 µl oil (Chill-out 7, MJ Research), containing 4 µl Stage 1 reaction products, 2 U *Taq* polymerase (Boehringer Mannheim), 50 pmol ss-primers #[1+5] or #[1+3] (specific for Group 1 or Group 2 pIg-R mRNA families, respectively), 1.5 mM MgCl₂ (Boehringer Mannheim 10 × PCR buffer), 50 mM KCl, 200 µM ultrapure dNTPs (Pharmacia) and 10 mM Tris-HCl, pH 8.3. Reactions were performed with a thermal cycler (MJ Research) for 35 cycles as follows: 40 s at 95°C, 60 s at 50°C and 60 s at 70°C. PCR reaction products were recovered and analyzed by electrophoresis through 1.2% agarose minigels (26) using a 1 kb DNA ladder (Gibco-BRL). Southern blots from minigels (10 µl PCR reaction products/lane) were performed by PosiBlot (Stratagene), using [γ -³²P]ATP 5' end-labelled 5'-[tcc]_{n=11}-3' and 5'-[tcc]_{n=11}-3' probes, under conditions similar to Northern analyses (13).

Southern analyses

DNA samples were transferred to nylon membranes (Gene Screen) at 21°C using a vacuum slot blot (BRL) or PosiBlot apparatus (Stratagene). Before transfer, samples destined for slot blots (10 ng DNA of subcloned pGEM *Sst*I inserts A-E, pBS-ESPS5, pBS-ESPS3 or a 1.6 kb *Pst*I insert in pBS-ESPS' mid'; or, 500 ng Wistar λ 2IgR DNA) were heated 5 min at 100°C in alkaline solutions (0.4 M NaOH) containing salmon sperm carrier DNA (1 µg/slot) and shifted to 4°C. DNA samples destined for positive pressure blots were obtained following electrophoresis through 1.2% agarose minigels (1 µg DNA/track). Before transfer, gels were stained with EtBr, photographed, then incubated for 30 min at 21°C in solutions of 0.5 M NaOH and 1.5 M NaCl, followed by 1.5 M NaCl and 1 M Tris-HCl, pH 7.5. After transfer (at 21°C in 10 × SSC for 45 min), all membrane strips were incubated three times in 2 × SSC for 5–10 min at 21°C and DNAs crosslinked by UV irradiation (Stratalinker®). The strips were dried and subjected to prehybridization (4 ml solutions in sealed plastic bags containing 50% formamide, 0.025 M K₃PO₄, pH 7.4, 5 × SSC, 5 × Denhardt's, 0.1% SDS and 50 µg/ml Salmon sperm carrier DNA) overnight at 45°C in a shaking water bath. Hybridizations were performed

under identical conditions in solutions supplemented with 10% dextran sulfate and 5'-d[gcagggggaggagaaaaag]-3' (UCSDCC4), 5'-d[tcc]_{n=11}-3' (GAA), or 5'-[tcc]_{n=11}-3' (UCSDCC6) ss-probes (3–8 × 10⁶ c.p.m./blot). OLIGO software designed probes (≈2–3 × 10⁸ c.p.m./µg DNA) were constructed, purified and 5' end-labelled with [γ -³²P]ATP. Blots were washed at 21°C (in 2 × SSC for 5 min; and, twice in 1 × SSC, 0.1% SDS and once in 0.1 × SSC, 0.1% SDS for 30 min), drained briefly and subjected to autoradiography or radioisotopic scanning (13).

PIGR copy number determinations

DNA samples were applied to nylon membranes by vacuum slot or positive pressure blots, and subjected to Southern analyses, autoradiography and radioisotopic scanning. Slot blot samples, denatured in 100 µl solutions of 0.5 M NaOH and 1.5 M NaCl, consisted of: no DNA or 100 ng salmon sperm carrier DNA (controls); or, 5 µg DNA purified from Fisher rat liver nuclei (experimental; 13); or, 100 ng carrier DNA plus 22.5, 45, 90 or 180 pg of Wistar λ 2IgR DNA (standards). The amount of each DNA standard corresponded to theoretical numbers of ~0.7, 1.4, 2.8 and 5.6 genes per haploid equivalent, given a 45 kb size for the EMBL3 standard and the assumption that 10 pg of experimental DNA (1 haploid rat genome) (29) contained ~2 × 10⁻⁵ pg of a 14.4 kb fragment of the 21 kb PIGR gene. Following transfer, membrane strips were neutralized in two changes of 2 × SSC, irradiated with UV and subjected to hybridization (5 × 10⁶ c.p.m./blot) with either *Sst*I 0.24 kb cDNA inserts derived from pGEM-A or *Bam*HI pIg-R cDNA probes (1–1.1 × 10⁹ c.p.m./µg) labelled by random priming with [α -³²P]dCTP (≥3000 Ci/mmol, ICN). Prehybridization incubations with 75 µg carrier DNA were 4 h, washing for 15 min. DNA samples (Wistar λ 2IgR or total hepatic DNA) were digested with either *Eco*RI, *Bam*HI, *Kpn*I, *Nco*I, *Hind*III or combinations of them (26). Following 1.2% agarose minigel electrophoresis (1 µg DNA/track) and PosiBlot transfer, hybridizations were performed with the labelled *Bam*HI cDNA probe.

Formation and analyses of PCR products from Wistar λ 2IgR and Fisher rat DNA targets

Ss-DNA primers were designed and constructed by the UCSD Cancer Center Oligodeoxynucleotide Facility. Primer sequences were: 5'-d[gcagggggaggagaaaaag]-3' (UCSDCC4, upper); and UCSDCC5 (lower), as above. Unless noted, unlabelled PCR products were produced in 50 µl, under 20 µl oil (Chill-out 7, MJ Research), containing linearized ds-DNA target (40 ng, pBS-c.15; 50 ng Wistar λ 2IgR; or, 200 ng Fisher nuclear DNA [from a single rat liver]), 2 U *Taq* polymerase (Boehringer Mannheim), 50 pmol ss-primers, 1.5 mM MgCl₂ (Boehringer Mannheim 10 × PCR buffer), 50 mM KCl, 200 µM ultrapure dNTPs (Pharmacia) and 10 mM Tris-HCl, pH 8.3. PCR was performed (as above) with an eight-step protocol: (i) 1 min at 94°C (initial cycle), (ii) 40 s at 92°C, (iii) 40 s at 70°C (attenuate 0.5°C/cycle), (iv) go to Step 2 (19×), (v) 40 s at 92°C, (vi) 40 s at 60°C (plus 1 s/cycle), (vii) go to Step 5 (19×) and (viii) hold at 4°C. Unlabelled PCR reaction products were recovered, analyzed by electrophoresis through 1.2% agarose minigels, and subjected to Southern analyses by PosiBlot and hybridization with 5' end-labelled 5'-d[tcc]_{n=11}-3' or 5'-d[tcc]_{n=11}-3' ss-probes. Labelled PCR products were generated by standard procedures (11,26) using a

[25]:[1] molar ratio of [unlabelled UCSDCC4]:[5' [γ - 32 P]ATP end-labelled UCSDCC4].

In some cases, single bands from gel separated PCR products, derived from Wistar λ 2IgR and Fisher DNA, were isolated, repurified (with GeneClean II; BIO 101) and reanalyzed by agarose minigel electrophoresis and Southern transfer. In other cases, DNA targets were subjected to 1st and 2nd rounds of PCR, then analyzed by 3% agarose or 8% non-denaturing polyacrylamide minigels (26); or 32 P-labelled PCR products were recovered, heated 2 min at 100°C in a solution of 95% ultrapure formamide and 0.1% xylene cyanol and bromophenol blue, electrophoresed through 8 M urea/6% polyacrylamide minigels and, after drying 1.5 h, visualized by autoradiography (13); or, further analyzed by digestion with *Cvi*II (30) and analyzed by non-denaturing 3.5% agarose minigel electrophoresis.

Subcloning and sequencing of Fisher rat genomic PCR products

Genomic Fisher 250 bp PCR products were prepared with primers phosphorylated with T4-polynucleotide kinase (26). PCR reaction products were gel purified, blunted with T4 DNA polymerase, and ligated into *Sma*I-blunted pBSIISK[−] plasmids treated subsequently with calf intestinal alkaline phosphatase (26). Blue–white color selection of SURE cells (Stratagene) transformed with this plasmid mixture resulted in two colonies containing inserts, from a total of three transformants. The plasmids from these two clones containing inserts (F1[PCR] and F2[PCR]) were amplified and purified by ultracentrifugation through CsCl followed by Qiagen Q20 column chromatography (Qiagen, Chatsworth, CA). The sizes of the inserts (including plasmid backbones flanking each insert) were determined by digestion at proximal multicloning sites with *Xho*I and *Xba*I and electrophoresis through 1.2% agarose minigels.

DNA sequences of both strands of the F1[PCR] and F2[PCR] inserts were determined by *Taq* Dye Deoxy-Termination Cycle sequencing reactions (Applied Biosystems), using either Stratagene primers (as above), or UCSDCC4 and UCSDCC5 primers. Sequences were aligned and tested for statistical significance using Sequencher (MacIntosh) and MACAW software (27).

Restriction site mapping and nuclease-S1 digestion

The 0.24 and 0.62 kb genomic Wistar DNA inserts in pGEM-A (control) and pGEM-ESBB, respectively, were analyzed by digestion with restriction enzymes (Gibco-BRL) and electrophoresis through 0.8% agarose minigels. Restriction mapping of larger genomic fragments from total Wistar rat liver DNA was performed similarly with blotting to nitrocellulose. PIGR-specific fragments were detected by Southern analyses as described above with a random primer labelled *Bam*HI pIg-R cDNA probe. After hybridization in 6 × SSC for 24 h, the filters were washed at 52°C in 1 × and then 0.2 × SSC (26), dried and subjected to autoradiography for 20 h at −70°C.

For S1-analyses of Fisher cDNA inserts, 8 µg of purified supercoiled DNA plasmids (control 2.9 kb pBluescriptIISK[−] = pBS; or, experimental 3.3 kb subclone pBS+[c.3]) were incubated without ('Nothing') or with 5 U of nuclease-S1 (Pharmacia) or 10 U of *Sca*I (Gibco-BRL) for 30 min or 3 h at 37°C, respectively (26). For sequential treatment, digestion products remaining after incubation with nuclease-S1 (~½ vol of a reaction buffer

containing 0.2 M NaCl, 0.05 M sodium acetate, 1 mM ZnSO₄, pH 4.5) were cleaned first with phenol and then with CHCl₃, precipitated with EtOH, redissolved in TE buffer and treated with 10 U of *Sca*I. Final reaction products were analyzed by electrophoresis through 1.2% agarose minigels using a 1 kb DNA ladder. Tests for S1-sensitivity of genomic subclones from Wistar rats were performed identically.

RESULTS

Regenerating Fisher rat liver cDNAs contain a novel STMS harboring two tandem triplet repeat regions

STMS-containing cDNAs were initially isolated serendipitously from rat cyclin D1 cloning attempts, using two directed hepatic cDNA libraries (14). In this screening, partial sequencing of several cDNA inserts revealed four 'families' exhibiting extensive internal homologies although none harbored significant homologies to published sequences of human or murine cyclin D1. Two families (~1 and 4 kb [2 and 5 clones, respectively]) showed no homologies to current nucleic acid sequence database entries. Another family (~0.5 and 1.3 kb [1 and 2 clones, respectively]) showed perfect homologies to murine and human complement C'9 mRNAs. Partial sequencing of 11 clones in a 4th family revealed a novel STMS consisting of two regions of tandem trinucleotide repeats, d[GGA]_n, GAG or AGG.. d[GAA]_n, AAG or AGA, encoded into the 3' ends of several cDNAs. Figure 1A shows the complete sequence of one such clone ('c.3'). Its core STMS structure, 5'..[60 nt GA-rich region][GGA]_{n=26}[GAA]_{n=35}.3', was composed almost entirely of dG and dA residues; poly_A signals were noticeably absent.

STMSs are found in two groups of polyadenylated Fisher rat hepatic pIg-R cDNAs and mRNAs

To investigate STMS properties further, five inserts in the 4th cDNA family (c.15, c.7, c.9, c.1 and c.2) were sequenced. Their properties are summarized and alignments shown in Figures 1B and 2, respectively. The 5' ends of the six inserts were related perfectly ($P(N) \leq 1 \times 10^{-22}$) to a 669 nt non-repetitive 3' UTR spanning positions 2601–3269 of the rat pIg-R mRNA (Fig. 2). The complete sequences of five remaining 4th family clones (c.20, c.34, c.25, c.26 [Group 1-like] and c.23 [Group 2-like]) have not been determined.

All six 3' UTRs contained STMS elements at their 3' ends with a *chi*-like octamer (d[GCAGGAGG], nucleotides 3279–3286; ref. 31) in each GA-rich constant region; no STMS-free 3' UTRs were observed (Fig. 2). GGA and GAA tracts in each STMS displayed unusual sequence diversity, manifested by variable ($n = 24, 25$ or 26) or hypervariable numbers of triplet repeat units ($n = 9, 24, 29, 35$ or 37), respectively. In all cases, GGA variable regions were 'altered' at the 1st nt positions of their 8th triplet repeat by a dA residue (instead of dG); in one case (c.1), a similar alteration leading to a d[AGA] triplet occurred immediately 5' to the '1st' GAA repeat unit. In contrast, GAA repeat regions were unaltered; and, an inverse linear relationship between their repeat numbers and poly_A tail lengths was seen ($r = -0.94$; Fig. 1C). Although this correlation coefficient is significant statistically ($P < 0.01$), the possibility that poly_A tail length varies randomly, depending upon hybridization of oligo_dT at the initial cloning stage, was not eliminated.

Figure 1. Structural features and homologies of an STMS among 3' UTRs of two groups of pIg-R cDNAs in regenerating Fisher rat livers. (A) Manual sequencing gel autoradiogram of the [-] strand of c.3. *Taq* polymerase and a reverse primer were employed. The gel was run 6.5 h at 60°C and the film was exposed 60 h at -70°C. (B) Diagram of relationships among the 3' UTRs of *Group 1* and *Group 2* cDNAs, and a truncated rat pIg-R mRNA containing the full length coding region (rectangle with diagonal bars). The length of each cloned cDNA insert is given in parentheses. A unique 283 nt 'unit' in the *Group 1* family (see Fig. 2) is indicated by solid black rectangles downstream of the STMS (shaded rectangles) and upstream of poly_A tails (saw-tooth lines). (C) Relationship between GAA repeat size and poly_A tail length (▲ [*Group 1*], ■ [*Group 2*]). Dotted lines show 95% confidence limits. (D) Sequence alignments of significant regions of homology among human cyclin D1 cDNA screening probe D1 and pIg-R cDNA clones from panel B. Dark highlighting, capital letters or wide rectangles (block diagram at left) indicate regions of statistically significant alignment scores. Absolute nucleotide positions in the cDNA inserts or probe are given in parentheses.

We next divided the STMS-containing inserts into ‘*Group 1*’ (c.15, c.7) and ‘*Group 2*’ (c.3, c.9, c.1, c.2) pIg-R cDNA families respectively distinguished by the presence or absence of putative poly_A signals (Figs 1B and 2; see ref. 32). A prominent difference was the presence of a 283 nt 3’ end ‘unit’ in *Group 1* cDNAs absent in *Group 2* cDNAs (Fig. 2). This unit (spanning nucleotides 3525–3807) contained two AATTAA sites (nucleotide positions 3701–3706 and 3783–3788) flanked upstream by GAA repeats and downstream by poly_A tails; and, a complex 2° structure (33) that included a histone 3’ UTR-like stem–loop (34) and GAA- and CT-rich inversions. In contrast, the hypervariable GAA tracts in *Group 2* mRNAs merged uninterruptedly with their poly_A tails, resulting in a configuration with minimal 2° structure and designated as a poly(A)-proximal tandem triplet repeat (PAPTTR).

The fortuitous isolation of truncated pIg-R cDNA clones in both libraries might be attributed to relatively stable hybrids formed between the probe and both *Groups* of cDNAs (the T_m for dG·dC base pairing in this GGA-rich region = 84°C [%GC method]) during screening. This is supported by sequence alignments (Fig. 1D) showing a region of significant GGA sequence homology between the cyclin D1 screening probe and GGA repeat regions of the sequenced inserts.

To investigate the possibility that STMSs in rat pIg-R cDNAs were generated by cloning artifacts or produced specifically during liver regeneration, Northern and RT-PCR analyses were made with RNAs isolated directly from quiescent and regenerating liver. When Northern blots of total or polyA⁺ hepatic RNAs were probed with the 2.3 kb *Bam*HI pIg-R cDNA, a broad band of 3.5–3.9 kb was observed in RNA samples obtained from quiescent or regenerating tissue (data not shown). Based upon cDNA sequencing results, and the assumption that size distribu-

tions of native hepatic mRNAs are similar to those present during library construction, the predicted size range of composite pIg-R mRNAs would be 3.5–3.9 kb (3520, 3528, 3533, 3548, 3817 and 3823 nt for **c.1**, **c.2**, **c.3**, **c.9**, **c.7** and **c.15**, respectively). Steady-state levels of pIg-R 3.5–3.9 kb mRNAs, detected on Northern blots, were invariant between 0–12 h after sham or 67% hepatectomy. However, 24–192 h post-operatively, the levels of these mRNAs rose 3–10-fold in regenerating liver compared to sham operated controls, concurrently with hepatocyte membrane levels of mature pIg-R receptors detected by immunofluorescence *in situ* and by Western blotting (L. Saucan, G. E. Palade, K. S. Koch, A. S. Gleiberman and H. L. Leffert, unpublished results). Additional measurements suggested that regeneration-associated levels of *Group 2* pIg-R mRNAs exceed *Group 1* levels 24 h post-operatively. Though gel loading and RNA transfer artifacts did not appear to account for these differences, judged by EtBr-staining and control blots using a 1.2 kb G3PDH cDNA probe, further work is needed to confirm these observations. When stripped (26) or parallel blots were analyzed with probes for the pIg-R 3' UTR regions common to both *Group 1* and *Group 2* (nucleotides 3227–3256 and GAA repeats), or specific to *Group 1* 283 nt units (nucleotides 3709–3738; see Fig. 2), each probe hybridized to mRNAs in an identical size range. Discrete bands of higher M_r were not observed (35).

When RT-PCR was performed with RNAs from regenerating liver, 1st strand cDNA products were formed initially using 3' end UCSDCC2 and UCSDCC3 primers, downstream of GAA repeats, acting as selective anchors for the postulated *Group 1* or *Group 2* mRNA regions (**c.7** [nucleotides 3790–3808] and **c.15** [nucleotides 3796–3814], or **c.2** [nucleotides 3431–3449], **c.1** [nucleotides 3476–3494], **c.9** [nucleotides 3488–3506] and **c.3** [nucleotides 3509–3527], respectively; see Fig. 2). PCR products

Figure 2. Sequence alignments of the 3' ends of a truncated rat pIg-R mRNA and STMS-containing cDNAs from regenerating Fisher rat liver libraries. Manual sequencing methods were used. STMS core elements (60 nt GA-rich regions including a *chi*-like octamer at relative nt positions 3279–3286, variable GGA repeats and hypervariable GAA repeats); restriction enzyme, G→A conversion ('G→A'), cryptic- or histone 3' UTR-like stem-loop polyadenylation sites; 283 nt *Group 1* units; and canonical *Group 1* poly_A tracts are annotated. Dark highlighting and capital letters indicate regions of statistically significant alignment scores. GenBank accession numbers for subcloned inserts of the *Group 1* {c.15, c.7} and *Group 2* {c.3, c.9, c.1 and c.2} mRNA families are {U07886, U02506} and {L22660, U01145, U00762 and U00763}, respectively. Absolute nucleotide positions are given in parentheses.

were then produced from ds-cDNAs predicted from both *Groups* of endogenous message, using a sense primer (UCSDCC1 [nucleotides 2331–2349]) spanning the known terminal portion of the pIg-R coding region (amino acids 735–741 of the mature protein) and lower primers, downstream of GAA repeats, specific to the postulated *Group 1*-like CT-rich region (UCSDCC5 [nucleotides 3566–3584]) or targeted to the *Group 2*-like anchor (UCSDCC3). PCR primer UCSDCC1 was specific for pIg-R mRNAs as NCBI nucleic acid database searches revealed no unrelated homologies. As expected from the sizes of the predicted

full-length pIg-R-derived PCR products determined from Figure 2 (1248 and 1254-bp [*Group 1* (c.7 and c.15)] or 1119, 1164, 1176 and 1197 bp [*Group 2* (c.2, c.1, c.9 and c.3)], respectively), the sizes of the major PCR products observed were ~1.2 (*Group 1*-like) or ~1.1 kb (*Group 2*-like), respectively. Major PCR products were specific since neither were formed from linearized insert-free plasmid DNA targets or H₂O control samples; and, both displayed positive hybridization signals on Southern blots for GAA and GGA repeats using labelled [ttc]_{n=11} and [tcc]_{n=11} probes (data not shown). Both experiments produced STMS-con-

Figure 3. Additional genes with significant sequence homologies to the Fisher rat pIg-R STMS core structure. Statistically significant DNA sequences obtained from NCBI nucleic acid databases were aligned with MACAW and analyzed for [+]strand homologies with the pIg-R STMS in c.3. Dark highlighting and capital letters indicate regions of statistically significant alignment scores. GenBank accession numbers (top → bottom [β -globin → ... → Lens 3' UTR], respectively) are X14061, M17284, X03541, M81784, X17562, V01144, X17093, M64933, M35983 and L10113. Alignments are displayed according to relative bp positions. Absolute nucleotide positions are given in parentheses.

taining PCR products of smaller size: ~740, 880 and 1080 bp (*Group 1*); and, 290, 480, 740 and 880 bp (*Group 2*). Further work is needed to determine the sequence and origin of these unexpected bands.

Rat pIg-R STMS homologies to related and unrelated sequences

Upon inspection of reported pIg-R cDNA sequences of four other mammalian species (36–39), no STMS-like homologies were found. However, *chi*-like octamers were found in human, bovine and rabbit cDNAs (and again, upstream in the rat open reading frame [nucleotides 1976–1983]). Upon inclusion of the mouse cDNA sequence, further evidence of eight conserved homologous tracts in the pIg-R 3' UTRs of these mammals was revealed. These data can be supplied to interested investigators by sending a reprint or E-mail request to hleffert@ucsd.edu.

When a search was made for homologies to the rat pIg-R STMS (nucleotides 66–311 in c.3), significant homologies ($P(N) = 2.1 \times 10^{-52}$ – 1.8×10^{-11}) were identified among 10 GenBank entries. These alignments are shown in Figure 3. The sequences represented in this GGA and GAA-rich [+]strand family include the 3' and 5' UTRs of two members of the Ig superfamily (*mus* IgG2a and human HLA-F5P, respectively [17]) both of which contain a *chi*-like octamer; the 5' UTR of the human *TRK* oncogene (40) which contains 12 *chi*-like octamers; a mouse enhancer (41); a B1-repetitive element in mouse β -globin (42); upstream regions of the 3' UTRs of cephalopod Ω -crystallin (43) and a putative 4.6 kb rat hepatic K^+ channel (rK+K13; ref. 35) which also contains a *chi*-like octamer; spacer regions between sea urchin histone (44) and mouse CYP₄₅₀ genes (Interger) (45); and a rat ID element (46). Minus-strand HSPs are not listed.

Organization of STMS-like elements at the 3' end of the Wistar rat PIGR gene

Four independent and overlapping genomic isolates were obtained spanning ~25 kb of chromosomal sequence including 5' and 3' flanking DNA. One of them, 'λ2IgR', was divided into six *Sst*I fragments ($A \approx 241$; $B \approx 840$; $C \approx 1800$; $D \approx 3000$; $E \approx 3500$; and,

Figure 4. Schematic diagram, restriction map and sequencing strategy for portions of λ2IgR, a genomic clone from a Wistar rat, and λ2IgR subclones. The nt numbering schemes for E2, E3, E8 and the Terminal E are based upon rat pIg-R cDNA and composite full-length *Group 1* pIg-R[c.15] sequences, respectively (see Fig. 2). Exonic, intronic, STMS ('Microsatellite') and intergenic regions are diagrammed to scale (see legend at right). Sequenced regions in the four subclones are designated by rightward (→, [+]strand) or leftward arrows (←, [–]strand). GenBank accession numbers are U08271 (a 374 nt tract at the 5' end of pBS-ESPS5), U08272 (a 312 nt tract at the 3' end of pBS-ESPS5) and U08273 (a continuous 999 nt tract from overlapping contigs in pBS-ESPS3 and pGEM-ESBB, and the insert in pGEM-A).

$E \approx 5000$ bp) and subcloned into pGEM-7Z. Partial sequencing revealed portions of I2 and exon 3 (E3) at the 5' boundary of fragment D (diagrammed in Fig. 4); and, a stretch of 87 nt in fragment A perfectly homologous to nucleotide positions 1121–1207 and 546–632 of the poly A signal-containing proximal regions of *Group 1* Fisher rat pIg-R c.15 and c.7 cDNAs (Fig. 2). These observations suggested that λ2IgR contained a complete 3' terminal region and, assuming a PIGR genomic size of ~21 kb (16,25), λ2IgR spanned $\geq 70\%$ of the length of the PIGR gene.

To identify exonic fragments in λ2IgR containing the 3' UTR, Southern blot analyses were performed using the pIg-R 2.3 kb *Bam*HI cDNA probe (Figs 4 and 5) and indicated that fragments E and F contained sequences homologous to this probe (data not shown). Fragment F was not analyzed further, since restriction site mapping of fragment E (Fig. 4) revealed co-linearity between a 3'

Figure 5. Regional localization of the STMS in the Wistar rat genomic clone λ 2IgR by Southern analyses. Purified ds-DNA fragments from pGEM *Sst*I inserts A–E, untreated λ 2IgR, subclones of pGEM-E (see Fig. 4) and control salmon sperm DNA were subjected to slot (top and middle panels) or PosiBlot transfers (bottom panel) and analyzed by hybridization to either GAA-specific (GAA), 60 nt GAA-rich constant region-specific (UCSDCC4) or GGA-specific labelled ss-probes (UCSDCC6). Blot autoradiograms were made after film exposures of 18 (GAA and UCSDCC4 probes) or 48 h (UCSDCC6 probe) at -70°C .

*Pst*I–*Sst*I sequence and restriction cleavage sites flanking the STMS in the 3' UTRs of the *Group 1* Fisher rat pIg-R mRNAs (Fig. 2).

These observations suggested that λ 2IgR-derived fragment E contained STMS core structure. To confirm this prediction, Southern blot analyses (Fig. 5) were performed with all λ 2IgR *Sst*I fragments and three subclones of fragment E (pBS-ESPS5; pBS-ESPSmid [housing an intervening *Pst*I–*Pst*I 1.6 kb insert]; and, pBS-ESPS3 [housing a 3' end 758 bp *Pst*I–*Sst*I insert]) using ss-probes specific for the three STMS core elements. These analyses showed that STMS-like sequences were present only in pBS-ESPS3.

The presence of *Sst*I sites at the 5' boundary of fragment E and at nucleotide 1993 in the pIg-R cDNA, suggested that fragment E contained portions of E8 and/or I8. To investigate this possibility, the subclone pBS-ESPS5, containing the 0.9 kb *Sst*I–*Pst*I from the 5' end of fragment E was sequenced partially. The results are shown in Figure 6 (top). Nucleotides 1–104 were identical to pIg-R mRNA nucleotides 1993–2096. This stretch also matched the terminal portion of human E8; and, its predicted translation product was identical to the predicted sequence of the mature rat protein along this region (amino acids 623–656; see Fig. 6 [top]). A conventional 5' splice donor site (d[GT..]) was observed at nucleotides 105–106 in this insert (nucleotide 104 was the 1st nt of codon 657 [asp]); this site and 32 nt of adjoining I8 are also shown in Figure 6 (top).

The results of additional sequencing of the 0.9 kb insert in pBS-ESPS5 (Fig. 4) between nucleotides 139–374 and 589–900 at its 5' and 3' ends, respectively, are given in GenBank accession numbers U08271 and U08272, respectively. These results indicated that two additional STMSs were present in the 3.3 kb terminal region of the rat PIGR gene. Both were composed of dinucleotide tandem repeats of d[CA] ($n = 18$) or d[CT] residues ($n = 19$, with the 4th and 5th repeats of this tract interrupted by d[TT]). Purine-rich stretches ($\approx 70\%$) of 105 nt immediately flanking these STMSs were also observed.

The combined results indicated that λ 2IgR *Sst*I fragments D, E, E and A were oriented tandemly (5'→3') and contiguously, and were composed of exons and introns expected from published

Fisher rat pIg-R mRNA and human PIGR sequences, respectively. The precise orientations of fragments B and C were not determined. Based upon sequencing results with additional genomic clones, if either λ 2IgR-derived *Sst*I fragments B or C were situated 5' to fragment D, then portions of E2 and I2 are likely to be components of these fragments (E. J. Fodor *et al.*, manuscript submitted for publication).

Structure of triplet repeat STMSs in the 3' UTR of PIGR genomic clone λ 2IgR

To localize triplet repeat STMS-rich regions and flanking sequences precisely, overlapping inserts in pBS-ESPS3, pGEM-ESBB (a 622 bp *Bgl*III–*Sst*I subclone of pBS-ESPS3) and pGEM-A, as diagrammed in Figure 4, were sequenced. A composite sequence (999 nt) was analyzed for similarities to rat pIg-R mRNA and *Group 1* cDNA clone (c.15). This analysis is shown in Figure 6 (bottom). The results indicated that the STMS-rich triplet repeat region was embedded in an 845 bp 'terminal exon', bounded by 3' nucleotide 3807 and 5' nucleotide 2941 (Figs 2 and 6 [bottom]). Matches along stretches of 413, 104 and 287 nt were evident spanning nt positions 2941–3353 (housing the 60 nt d[GAA]-rich constant region including seven tandem d[GGA] triplet repeats), 3415–3518 (housing 34 tandem d[GAA] triplet repeats), and 3521–3807 (housing the 283 nt *Group 1* unit), respectively.

Thus, a region of $\sim 95\%$ homology was represented (804 of 845 nt) between the terminal exon in λ 2IgR and the transcribed 3' UTR of Fisher rat pIg-R *Group 1* mRNAs. As predicted from cDNA sequencing, cryptic (nucleotides 3211–3217 [ref. 34]) and canonical poly_A signals were observed (nucleotides 3701–3706 and 3783–3788), a *chi*-like octamer was seen at nucleotides 3279–3286, and several inversions were present at nucleotides 3529–3546 (d[GAA]-rich), 3557–3579 (d[CT]-rich) and 3643–3652 (d[GAGTT:TTGAG]), respectively.

One PIGR copy per haploid genome

To quantify rat PIGR gene copy number, DNA slot blots were performed with high M_r Fisher rat liver nuclear DNA ('F'), using ^{32}P -labelled λ 2IgR-derived fragment A or the *Bam*HI fragment of pIg-R cDNA as probes flanking the STMS. The 14.4 kb λ 2IgR genomic clone was employed as a quantitative standard of haploid 'equivalents'. These standards were based upon estimates of 20 ± 2 pg DNA/diploid equivalent, determined from measurements of total nuclear DNA content of primary adult Fisher rat hepatocytes (29). The results of these studies, obtained from quantitative scanning of individual tracks, suggested that Fisher rats contain one PIGR gene copy per haploid DNA equivalent (Fig. 7A).

Single high M_r *Bam*HI and *Kpn*I fragments were also observed on genomic Southern blots, using the 2.3 kb *Bam*HI pIg-R cDNA probe, following treatment of Wistar rat liver DNA with these and several combinations of additional restriction enzymes (Fig. 7B). These findings suggest a single copy PIGR gene and are consistent with prior analyses of the single copy rabbit PIGR genomic sequence (23).

Figure 6. Partial DNA sequence of the terminal 3.3 kb region of the PIGR gene from Wistar rat genomic clone λ 2IgR. The illustrated region ('GENE') spans the 3' and 5' ends of E8 and I8, respectively, and the 3' end of a putative penultimate exon through intergenic spacer flanking the remaining and complete portions of the 3' terminal exon. The nt numbering schemes (5'→3') are based upon pig-R cDNA ('mRNA') and full-length composite Fisher rat *Group 1 pig-R* ('c.15') sequences. Dark highlighting or capital letters indicate regions of statistically significant alignment scores. Top panel, terminal portion of E8, the E8-I8 junction and the flanking 34 nt 5' end of I8 (GenBank accession number U08271). The amino acid sequence ('Protein') specified by the reading frame (→) is given in single letter code. Bottom panel, terminal exon and intergenic spacer (GenBank accession number U08273). Annotated regions include: STMS core elements (a 60 nt GA-rich tract with a *chi*-like octamer at its 5' end [nucleotides 3279–3286], a tract of 15 GGA tandem repeats, and a tract of 39 GAA tandem repeats interrupted by d[AA] between its 5th and 6th repeat units); restriction, dG→A conversion, 'mismatch junction', and cryptic- and canonical poly_A signal sites; and a *Group 1*-specific 283 nt unit containing far upstream inversions including d[GAA]- and d[CT]-rich tracts (flanked by poly_A tails in c.15 and c.7). Arbitrary placement of the 22 nt genomic 'deletion' (nucleotides 3394–3414) by the MACAW alignment program is reformatted in Figure 9.

Rat PIGR genomic STMSs are homologous but not identical to STMSs in Fisher rat pIg-R cDNAs

With the exception of the 60 nt constant GA-rich regions in Fisher cDNA and Wistar genomic sequences, 5% of the sequenced portion of the terminal exon in $\lambda 2\text{IgR}$ (bounded by nucleotides 2941 and 3807) did not match Fisher rat 3' UTR cDNA sequences. Except for a single base change at nucleotide 3124 (dG→A), mismatches occurred exclusively along homologous STMS stretches extending from nucleotides 3333–3520; specifi-

cally, in regions of GGA and GAA core structure and, with respect to *Group 1* cDNAs, at a 2 nt terminus (d[GGA]₁₅[GAA]₅AA[GAA]₃₄AT; see Fig. 6 [bottom]).

Three principle differences were notable. Firstly, in contrast to the numbers ($n = 24-26$) and discontinuities (an Δ GA trinucleotide between the 7th-9th repeat units) of GGA repeats in *Group 1* and *Group 2* Fisher pIg-R cDNAs, the Wistar PIGR genomic sequence displayed insufficient numbers ($n = 15$) along a continuous stretch of tandem GGA repeat units. Upon inspection of one kind of alignment format of comparisons with *Group 1*

Figure 7. Determination of PIGR copy number in Fisher and Wistar rats. Purified hepatic DNA samples were obtained from purified liver nuclei (Fisher) or total liver tissue (Wistar). Experimental and control DNA samples (salmon sperm) were untreated (Fisher) or subjected to digestion with several restriction enzymes (Wistar), transferred to solid support filters and analyzed by hybridization using random primer labelled cDNA probes. (A) Inbred Fisher rat. Copy number was determined by visual comparisons among PIGR genomic copy number standards and a theoretically predicted mass equivalent of nuclear DNA (copy number = 1) using hybridization probes flanking the known STMS upstream (the *Bam*HI 2.3 kb pIg-R cDNA) or downstream (insert 'A' from the λ 2IgR genomic subclone, pGEM-A). The autoradiogram of the resulting slot blot was made after a film exposure of 19 h at -70°C . (B) Outbred Wistar rat. Copy number was estimated by the analyses of restriction fragment sizes obtained from total DNA samples and from information obtained from additional restriction digestion analyses of known PIGR genomic clones λ 2IgR (Fig. 4) and λ 10IgR. 'R1' and 'Bam' are abbreviations for *Eco*RI and *Bam*HI, respectively; the positions of standard 3.3 and 18 kb ds-DNA size markers are given at the left. The autoradiogram of the resulting electrophoretic blot was made after a film exposure of 20 h at -70°C .

c.15 cDNA, these differences created a stretch of five intermittent dA→G conversions (at nucleotides 3379, 3382, 3385, 3388 and 3391) and a 'deletion' between nucleotides 3393–3414 (Fig. 6 [bottom]). Secondly, although λ 2IgR GAA repeat units ($n = 39$) accounted sufficiently for the maximal numbers of GAA repeat units observed in cDNAs ($n = 37$), GAA repeat units in cDNAs were observed without discontinuities whereas genomic d[GAA] repeats in λ 2IgR were interrupted by 2 nt (...d[AA]...) between the 5th and 6th repeat units. Thirdly, two genomic nucleotides at positions 3519–3520 were changed from AT to GA at coordinate sites in **c.15**.

To determine if strain differences accounted for these discordant differences between STMS regions in genomic and *Group 1* cDNAs, PCR analyses and PCR-product sequence determinations were made with genomic Fisher DNA ('F'), and λ 2IgR (' λ ') and **c.15** ds-DNAs as control DNA targets, using upper (nucleotides 3315–3333) and lower (nucleotides 3566–3584) PCR primers flanking variable regions of STMS structure (Fig. 6 [bottom]).

PCR products are shown in Figure 8. In the initial studies (panel A), a predicted 270 bp product, obtained from the Fisher cDNA target, was clearly distinguished from the observed and predicted 248 bp product of the Wistar λ genomic target; no PCR products were detected when linearized insert-free pBS or H₂O controls were used as targets. When F genomic targets were used, major

and infrequent minor products of ~250' and ~340–400 bp, respectively, were obtained. Southern blots to individually analyzed gel purified products, analyzed with labelled d[*tcc*]_{n=11} or d[*tcc*]_{n=11} probes, revealed hybridization signals for GAA and GGA repeats only within PCR products of 270 and ~250 bp (data not shown). Additional PCR analyses using slightly altered conditions consistently demonstrated identically sized Wistar and Fisher genomic products (~250 bp) and failed to reproduce the ~340–400 bp products from either λ or F targets (panels B–E). These observations suggest that the ~340–400 bp F products were PCR artifacts; further analyses were focussed, therefore, exclusively on the 250 bp forms.

The 250 bp Fisher PCR product (from panel A) was gel purified and subcloned. Two clones with inserts were recovered ('pBS-F1' and 'pBS-F2') and characterized by *Xho*I and *Xba*I digestion (flanking polylinker restriction sites) to yield chimeric fragments of ~330 and 300 bp (designated 'F1' and 'F2', respectively; data not shown). Both strands of each insert (pBS-F1: [reverse sequencing primer binding site ('RSPBS')..XhoI...{3'/5'[-]strand insert}..XbaI.. forward sequencing primer binding site ('FSPB')]; pBS-F2: [RSPBS..XhoI...{5'/3'[-]strand insert}..XbaI..FSPB]) were subjected to automated ds-DNA sequencing. The results are shown in Figure 9.

As inferred from the different sizes of their chimeric fragments, the sequences of the STMSs in the two PCR inserts derived from F1 and F2 were not identical. However, inspection of the alignments of the 239 bp insert ('F1[PCR]') and the *Hae*III–*Taq*I region overlapping the 248 bp STMS-containing sequence of λ 2IgR ('Wistar Gene') revealed that, except for a 'deletion' of three GAA triplet repeats in F1[PCR], the STMS sequences were identical in both strains, defining a consensus STMS (d[GGA]₁₅[GAA]₅AA[GAA]₃₁ATAGAA).

In contrast, the 207 bp insert ('F2[PCR]') did not show this consensus. Like its two companions (F1[PCR] and the Wistar Gene), it contained perfectly identical 5' and 3' ends (27 and 60 nt, respectively), with each 3' end housing a complete portion of the far upstream end of the *Group 1* 283 nt unit (showing identical *Cvi*II sites 8 nt downstream from perfectly conserved 18 bp GAA-inversions at nucleotides 264–281 [Fig. 9]). But, unlike its companion inserts, it contained a continuous tract of 40 GAA repeats and it was 'missing' three key consensus features: 12 triplet repeats in its d[GGA] region, and d[AA] and d[ATAGAA] sites interrupting and flanking its continuous GAA repeat tract, respectively (Fig. 9).

Sequencing artifacts did not account for the observed differences between F1[PCR] and F2[PCR], or for the three fewer GAA triplets in F1[PCR] compared to the Wistar STMS, since identical results were obtained with two different sets of primers (Fig. 9); and, no *Taq* polymerase artifacts of flanking d[A/T] residues were seen. In addition, in both cases of automated sequencing of F PCR products, the flanking vector sequences obtained (composed of ~70–100 bp as expected from the sizes of the chimeric fragments in digests of pBS-F1 and pBS-F2) were perfectly identical (data not shown) as predicted from the known primer binding and polylinker sequences flanking the *Sma*I site of the subcloning vehicle.

To investigate the possibility that the 207 bp F2[PCR] arose from a PCR artifact, PCR was repeated with F and λ targets in parallel. The results, similar regardless of ds-DNA origin, are shown in Figure 8B–E. When PCR products were formed under standard conditions ('1st PCR') but analyzed on 3% agarose gels, EtBr-stained bands <250 bp were not observed; if present,

Figure 8. Analyses of PCR-generated products from a Fisher rat pIg-R-derived cloned cDNA, Wistar rat genomic clone λ 2IgR and Fisher rat liver nuclear DNA. In all cases, unless noted, PCR was performed using unlabelled 60 nt GA-rich region-specific UCSDCC4 ('Primer 4') and 283 nt unit-specific UCSDCC5 ('Primer 5') probes as upper and lower primers, respectively. Vertical lanes in each panel are referred to in the text with consecutive numbers from left to right. (A) Size differences between Fisher cDNA and genomic STMS-containing PCR products. In each case (lanes 2–5), a single round of PCR was performed. PCR products were revealed by EtBr-staining following electrophoresis through 1.2% agarose minigels. A scaled diagram (upper left) of a composite Fisher rat 3823 nt *Group 1* pIg-R mRNA [c.15] is shown with its coding region (nucleotides 74–2383) and 3' UTR STMS (hatched rectangle); the nucleotide positions of the PCR primers and the *Bgl*III site (shown in the diagram) are taken from Figure 2. The predicted 270 bp PCR product generated from known STMS sequence in the ds-cDNA insert (c.15, nucleotides 2601–3823; double solid line diagram at lower left) is shown in the gel photograph (at right) in lane 2 (a 1 kb standard ladder in lane 1). The predicted 248 bp PCR product generated from the known STMS sequence in the Wistar rat λ 2IgR genomic clone (' λ ') is shown in lane 3. Individual gel purified genomic PCR products generated from Fisher rat nuclear DNA ('F') are shown as discrete major and (infrequently observed) diffuse minor bands in lanes 4 (~250 bp) and 5 (~340–400 bp), respectively. (B) Analyses of genomic products of 1st and 2nd rounds of PCR on 3% agarose minigels. First round PCR was performed using 0.1 ng Wistar rat λ 2IgR genomic clone (' λ ') and 0.5 ng Fisher rat nuclear DNA targets ('F') essentially as described in panel A, except for changes in the total reaction volume (20 μ l) and *Taq* polymerase (2.5 U [BRL]). Two μ l portions of each 1st round PCR sample were used as targets for a 2nd round of PCR, using an altered 6-step thermal cycling protocol as follows: (i) 3 min at 94°C; (ii) 40 s at 93°C; (iii) 55 s at 58°C; (iv) 3 min at 70°C; (v) repeat 3 cycles of steps 2–4; and (vi) 5 min at 70°C. Nine μ l of the resulting 1st (λ , lane 4; F, lane 3) and 2nd round PCR reaction products (λ , lane 2; F, lane 1) in each sample were analyzed by gel electrophoresis; no diffuse ~340–400 bp bands were observed under these conditions. Numbers shown along the left vertical axis designate relative positions of 1 kb ladder size markers (lane 5). (C) Analyses of genomic products of 1st and 2nd rounds of PCR on 8% polyacrylamide minigels. Equal portion aliquots (9 μ l) were taken directly from the remaining PCR samples described above in panel B: 1st round PCR products from λ and F targets (lanes 4 and 3, respectively); 2nd round PCR products from λ and F 1st round PCR product targets (lanes 2 and 1, respectively). PCR products were identified by EtBr-staining following gel electrophoresis (BioRad MiniProteinII, 60 $\times$ 80 $\times$ 1 mm). Numbers and connected indicator lines shown along the right vertical axis designate relative positions of 1 kb ladder size markers (lane 5). (D) Analyses of 1st round 5' [γ -32P]-end-labelled genomic PCR products on 8M urea/6% polyacrylamide minigels. 1st round radiolabelled PCR products from λ and F targets (lanes 2 and 3, respectively) were prepared (20 μ l, total volume), diluted 3-fold into deionized H₂O, and analyzed (3 μ l samples) by denaturing gel electrophoresis (1.5 h at 200 V) and autoradiography. The autoradiogram of the resulting fixed (30 min in 5% MeOH/5% acetic acid) and dried gel was made after a film exposure of 72 h at -70°C. Indicator lines connected to the numbers shown near the left vertical axis designate relative positions of 5' [γ -32P]-end-labelled 1 kb ladder size markers (lane 1). (E) Analyses of genomic PCR products following digestion with *Cvi*II on 3.5% agarose minigels. First round PCR products from λ and F DNA target samples (as in panel B) were subjected to electrophoresis through 1% agarose gels. The ~250 bp main bands from each reaction product were gel purified. Two μ l of the purified samples were then subjected to mock- (lanes 4 and 3) or *Cvi*II-digestion (lanes 2 and 1), respectively, at 37°C for 1.5 h, in 20 μ l aqueous reaction volumes containing 2 and 1 μ l of 10 $\times$ enzyme buffer and *Cvi*II (0.5 U, as indicated), respectively. Under the electrophoresis conditions employed (1.5 h at 100 V), the expected 29 bp fragment 'common' to each target would have migrated off the gel and would not be visualized by EtBr-staining. Numbers shown along the right vertical axis designate relative positions of 1 kb ladder markers (lane 5).

however, stained bands of 200–210 bp should have been seen (panel B). Although some ‘diffusion’ of stained products above and below the edges of the major bands was detected (lanes 3 and 4), this heterogeneity probably arose from incomplete 1st PCR cycles as a 2nd round of PCR (‘2nd PCR’) attenuated diffusion at these edges considerably (lanes 1 and 2). Similar results were obtained when these same PCR products were analyzed further on a higher resolving non-denaturing polyacrylamide gel (panel C), although these results revealed that the homogeneously sized PCR products in the mixture were clearly capable of forming topologically heterogeneous subpopulations (detected as several bands >250 bp [lanes 3 and 4]) that, following 2nd PCR, became more homogeneously sized in the neighborhood of 250 bp (lanes 1 and 2). When PCR products were formed with an end-labelled primer and analyzed on a denaturing high resolution polyacrylamide gel (panel D), ~95% of the products observed showed a size of ~250 nt; no higher M_r products were seen, but ~5% of the products from both targets showed smaller sizes of ~150–200 nt (panels 2 and 3). Lastly, if F targets that generated F2[PCR] were truly endogenous to nuclear DNA, and were consistently

amplified by PCR, then digestion of this PCR mixture with *Cvi*I would be expected to produce 178 and 29 bp fragments; in contrast, an endogenous 250 bp F1[PCR]-like or λ substrate would be expected to produce fragments of 219 and 29 bp after digestion (see Fig. 9). The results of *Cvi*I digestion showed evidence of only 219 bp fragments; no 178 bp fragments were seen (panel E).

All of these observations suggest that F2[PCR] was not consistently derived from the major 250 bp PCR product of F DNA. It is possible that F2[PCR] arose during PCR as a deletion product of F1[PCR] or of a >250 bp topological conformer (see for example, lane 3 or 4 in Fig. 8C). If F2[PCR] derives from the minor ~200 bp species present in a mixture of 1st round PCR products (see, for example, lane 2 or 3 in Fig. 8D), and was included in the agarose gel purified band used as target material (lane 4 in Fig. 8A), it is unlikely to be faithfully genomic in origin since both species were apparently produced by the λ target as well. If these two species from the single λ target prove to be identical to F1 and F2, then it is also unlikely that F2[PCR] arose from an artifact during the construction and/or bacterial ampli-

Figure 9. Alignment of STMS sequences of Wistar and Fisher rat Form-1 and Form-2 PIGR genomic DNAs. The genomic 369 nt Wistar *HaeIII/TaqI* sequence is taken from Figure 6 (relative nucleotide positions 3244–3624). The two Fisher sequences were obtained from subcloned inserts derived from an ~250 bp gel-purified PCR product (upper and lower primers UCSDCC4 and UCSDCC5, respectively) using purified rat liver nuclear DNA as the target. Three STMS core elements, a *chi*-like octamer, the 5' region of the 283 nt *Group I* unit, a d[AA] repeat (underlined bars at relative nucleotide positions 150–151) and a *CviII*-site are annotated. The GAA-inversion downstream of the 'mismatch junction' (landmarks annotated in Fig. 6) is shown at relative nucleotide positions 264–281. Dark highlighting and capital letters indicate regions of statistically significant alignment scores. GenBank accession numbers of the 369 nt Wistar genomic, 239 nt F1[PCR] and 207 nt F2[PCR] sequences are U08723, U17366 and U17367, respectively.

fication of pBS-F2 (6). The possibility that the truncated GAA region in F1[PCR] also arose artifactually or, instead, that it reflects a true allelic difference between the two rat strains remains to be determined.

Rat pIg-R cDNA and PIGR genomic STMSs are sensitive to digestion with nuclease-S1

To determine if rat pIg-R STMS-containing cDNA sequences manifested non-B DNA structure (41), supercoiled pBS plasmids lacking (3.3 kb) or carrying the 326 bp *c.3* insert were subjected to digestion with *ScaI* (unique to nt 2526 in pBSISK[–], but absent from the *c.3* insert), nuclease-S1, or nuclease-S1 followed sequentially by *ScaI*. Specific S1-sensitivity was observed only in pBS+[*c.3*] following sequential treatment, as indicated by the formation of two digestion products with sizes of 1.2 and 2.0 kb (Fig. 10A). A preliminary map of S1-sensitive sites (Fig. 10B) placed their positions within core STMS structure between pBS+[*c.3*] nucleotides 725–825 (nucleotides 3379–3478 of *c.3* cDNA sequence in Fig. 2). When similar studies were performed with pBS+[*c.15*, *c.7*, *c.9*, *c.1* or *c.2*], or with *BamHI* instead of *ScaI*, core STMS regions in all of the inserts were nuclease-S1 sensitive at similar sites.

Figure 10. Nuclease-S1 sensitivity of plasmids carrying STMSs of pIg-R cDNA or PIGR genomic origin. Purified supercoiled plasmids ('sc') were subjected to mock-digestion ('Nothing') or to digestion with one or two restriction enzymes, and analyzed by electrophoresis through 1.2% agarose minigels. Treatment conditions are summarized below each photograph of its stained gel. For both gels, standard ds-DNA ladders in the flanking tracks are indicated along the vertical margins (at right); positions of unique *ScaI* vector sites are marked in the associated diagrams. (A) pBS (control, no insert) and pBS+[*c.3*] (STMS-containing Fisher cDNA insert, *c.3*). EtBr-stained minigel. (B) Deduced location of the S1-sensitive region in pBS+[*c.3*]. Insert orientation is from *EcoRI*(5')→*XhoI*(3'). (C) pGEM-A (λ 2IgR Wistar genomic fragment control) and pGEM-ESBB (λ 2IgR Wistar STMS-containing genomic fragment). EtBr-stained minigel. (D) Deduced location of the S1-sensitive region in pGEM-ESBB. Insert orientation is from *BglII*(5')→*SstI*(3').

Similar experiments were performed with supercoiled 3.0 kb plasmids carrying λ 2IgR-derived STMS-free (241 bp, pGEM-A) or GGA- and GAA-rich (622 bp, pGEM-ESBB) inserts. The ds-plasmids were subjected to digestion with *ScaI* (a unique pGEM restriction site at nt 2500 absent from inserts ESBB or A), nuclease-S1, or nuclease-S1 followed sequentially by *ScaI*. Specific S1-sensitivity was observed only in ESBB following sequential treatment, as indicated by the formation of two specific digestion products with sizes of ~1.4 and 2.0 kb (Fig. 10C). When similar experiments were performed with *XbaI* instead of *ScaI*, sensitivity to nuclease-S1 was observed again at similar sites, as indicated by the formation of two specific digestion products with sizes of ~3.2 and 0.4 kb (Fig. 10C). A preliminary regional map of S1-sensitive sites (Fig. 10D) placed their positions within core STMS structure between ESBB nucleotides 320–480 (nucleotides 3424–3584 of composite genomic sequence in Fig. 6). Plasmid pBS-ESPS5 carrying d[CA] and d[CT] repeats was not tested.

DISCUSSION

Four STMSs are present in the terminal 3.3 kb region of the Wistar rat PIGR gene. Two distal clusters are composed of rare tandem triplet repeats (47) found in both Wistar (d[GGA]₁₅[GAA]₅AA[GAA]₃₄) and Fisher (d[GGA]₁₅[GAA]₅AA[GAA]₃₁) rats. Both distal STMSs are flanked upstream by a *chi*-like octamer (d[GCAGGAGG]) at the 5' end of a 60 nt

d[GAA]-rich constant region and downstream by a 283 nt unit containing several inverted repeats and two poly_A signals. Surprisingly, both STMSs are expressed discordantly and variably in the 3' UTRs of two *Groups* of Fisher hepatic pIg-R mRNAs.

Although the rat PIGR gene might be polymorphic or allelic (48) in the distal region of its STMS, or harbor neighboring tandemly duplicated STMSs (an unlikely possibility from available PCR results), we have obtained strong evidence of a single copy gene and perfect matches in a 155 bp region of the STMSs from two strains. Furthermore, no markers of processed pseudogenes (2) are revealed among widely spread regions of two full-length Wistar genomic PIGR clones. Our results are also consistent with human PIGR mapping studies (25) and rabbit PIGR copy number measurements from Southern blots suggestive of a single gene (23). We conclude, therefore, that the rat genome encodes a previously unrecognized STMS in a single functional PIGR gene and this exhibits unusual processing at its 3' end.

Additional evidence supports such conclusions. Firstly, at least six cloned Fisher rat hepatic cDNAs, devoid of intronic repeats, have been found to contain similar but non-identical patterns of tandem GGA and GAA repeat tracts in their 3' UTRs. These cDNAs also show striking dissimilarity unrelated to their STMSs, based upon the presence or absence of the 283 nt unit, leading to their operational classifications into *Group 1* and *Group 2* cDNAs. Lacking identifiable polyadenylation signals, and owing to the missing 283 nt unit, the continuous GAA-repeats of *Group 2* mRNAs merge uninterruptedly with poly_A tracts to form unusual mRNA 'ends' (PAPTTs) lacking canonical poly_A signals.

Secondly, variable STMSs in these cDNAs are likely to represent portions of the truncated ends of polyadenylated pIg-R mRNAs, as exact homologies exist among the different *Groups* of cDNAs, 3' exonic regions of PIGR genomic sequences, and coding and UTRs of the rat pIg-R cDNA (15). Cloning artifacts of 1st strand cDNA synthesis involving priming at internal oligo_A sites, or DNA slippage replication artifacts of manual or automated dideoxy-DNA sequencing involving T7 or *Taq* polymerases and different pairs of sequencing primers, are unlikely to have produced these STMSs or their variable structure: none of these STMSs contain unusual internal oligo_dA tracts; each STMS is defined distinctly by a stable and unique GGA and GAA repeat unit composition; no evidence of slippage replication is observed on sequencing gels; and both *Groups* of molecules containing variable STMS core structure are present in directed cDNA libraries from two separate animals. Furthermore, Northern and RT-PCR analyses of *Group 1* and *Group 2* mRNAs from regenerating and quiescent livers, reveal full-length endogenous pIg-R mRNAs under both conditions (3.5–3.9 kb) and predictably-sized pIg-R mRNA-derived products during regeneration (1.1 kb *Group 1*-like and 1.2 kb *Group 2*-like PCR products, respectively); STMSs containing constant GA-rich regions and GGA and GAA repeats are also observed by hybridization analyses in both cases. Although inspection of the available PIGR 3' UTR sequences among human (~447 nt of an incomplete 3' UTR) and non-human sources indicate many conserved regions of far upstream and far downstream 3' UTR homology, the genomic organization and the expression patterns of these or other STMSs in the pIg-R mRNAs of quiescent and extrahepatic tissues have not yet been analyzed rigorously in rats or other pIg-R-expressing species.

Thirdly, regardless of PIGR polymorphisms within or between Wistar or Fisher rats, predictions of mRNA sequences based upon identical 155 bp genomic STMSs in both strains are not borne out by cognate STMS sequences in either *Group* of Fisher pIg-R mRNAs. Thus, apart from noticeable changes at two sites (one of two dG→A conversions creating an upstream cryptic poly_A signal at nucleotide 3124, and, in the neighborhood of a 'mismatch junction' at nucleotides 3519–3524, where novel [...dG→A...] or conventional forms of RNA editing at nt 3519 [...dA→A→G...] and 3520 [...dT→U→A...] might account for d[AT]→GA), neither of the conserved purine-rich triplet repeat STMSs (d[GGA]₁₅[GAA]₅AA[GAA]₃₁) map precisely, as would be expected, into the triplet repeat regions of Fisher rat pIg-R mRNAs. Instead, genomic d[GGA]-repeats are amplified in both *Groups* of mRNAs (≥9–11 triplets longer than the genomic tract); whereas, despite genomic d[GAA]-tracts and a 3' flanking d[GAA]-rich inversion in both strains which might account for the longest of GAA repeats observed (*n* = 37), genomic d[GAA]-repeats are rearranged into truncated continuous tracts selectively in *Group 2* mRNAs.

A better understanding of mechanisms which generate variability in PIGR 3' UTR triplet repeats and the unusual processing at the 3' end of the rat PIGR gene might illuminate the pathophysiology of triplet repeat disorders (7,9,10). For example, in myotonic dystrophy (DM), misalignment and slippage during DNA replication are proposed to underlie abnormal germline (9–11) and somatic expansion (49) of d[CTG] STMSs in the 3' UTRs of this polymorphic gene. Although the details of these aberrant DNA replication processes are unknown, studies with chromatin systems (50) and chimeric tissue culture cell lines (12) implicate defective transcription and/or splicing of aberrant 3' UTR STMSs as contributing causes of attenuated MT-PK mRNA and protein levels in afflicted patients (11).

There are as yet no known inherited disorders related to the rat PIGR gene. Whether the limited discrepancies in the PIGR genomic STMS sequences reported here are shown to be strain specific allelism or polymorphisms between individual rats, it seems unlikely that DNA slippage replication play a major role in generating STMS or *Group*-specific diversity at the mRNA level; nevertheless, genomic sequencing from 67% hepatectomized individuals will further address this problem. Cryptic splicing or unusual editing post- or co-transcriptionally, for example, by the use of ...AG as a 3' internal acceptor, or GA... as an internal 5' donor site (51), would also be less probable determinants of unusual processing at the PIGR 3' end, but neither possibility can yet be eliminated. Alternatively, two other hypotheses might be considered: one, that unusual processing involve DNA rearrangement and recombination, coupled with normal RNA processing; and, the second, RNA polymerase II misalignment, slippage and premature transcription termination.

The first of these hypotheses may be consistent with the presence in the 5' end of the rat PIGR STMS of a transcribed *chi*-like octamer, a sequence strongly associated with chromosome translocation breakpoints and rearrangements of several oncogenes downstream of the octamer and with aberrant VDJ recombinase activity (52). This octamer is upstream of additional PIGR STMS sequence homologous to polypurine-polypyrimidine-rich repeats in other Ig superfamily members including introns of murine genes encoding the μ - δ Ig-heavy chain (53), the IgG3 constant region (54) and the DR2-C α -switch region of the

mouse gene encoding IgA (55). Although consistent with variability generated in lymphoid members of this Ig superfamily, perhaps too many stages of DNA rearrangement and recombination would be required to generate the variable configurations of the *Group 1* and *Group 2* pIg-R mRNAs observed in this non-lymphoid tissue. Despite the considerably different STMS sequence of F2[PCR] (d[GGA]₃[GAA]₄₀[283 nt unit..]) from that of F1[PCR], there is a possibility that F2[PCR] was generated artifactually; although polymorphisms of the rat PIGR STMS have not been detected, the generation of variability by DNA rearrangement and recombination as by DNA slippage, can only be examined by further individual genomic sequencing.

In contrast, RNA polymerase II misalignment and slippage replication might play a major role in processing, along with specifically conserved upstream and downstream 'diversity generating' sequences flanking the terminal d[GAA] repeat. A biochemical precedent for generating amplified tandem poly AAG repeat ss-RNAs from ss-d[ttc]₃ templates is well documented in cell free systems (56). But, to our knowledge, RNA misalignment and slippage replication have not been observed or suggested previously as *in vivo* generators of amplified or truncated numbers of triplet repeats in eukaryotic 3' UTRs. Thus, we would consider a process of this kind to be distinct from 'stuttering' phenomena reported elsewhere (57); and, were there genetic signals underlying such hypothetical processing, they are clearly unknown. Interestingly, the 13.7 kb sequence of a normal human DM gene (58) reveals 12 *chi*-like octamers (~14 more than expected by chance); one of them, located in the terminal intron (nucleotide 12 623) of the DM gene, is identical to the rat PIGR 3' UTR *chi*-like octamer and is ~600 nt upstream of the unstable 3' UTR d[CTG] STMS of the DM locus. Given these parallels between the 3' ends of the unstable human DM and the variable rat PIGR STMSs, one might speculate that a second role of *chi*-like octamers might be found at the level of RNA replication and that either one or both of the pIg-R *chi*-like octamers might serve as upstream destabilizing signals for promoting RNA polymerase II misalignment and slippage. Downstream of the PIGR STMS, a conserved hexamer (d[ATA-GAA]) is found at the site of the mismatch junction separating the STMS terminal d[GAA] repeat and the 5' end of the 283 nt unit, where most of the CT-rich inversions reside. These inversions house sequences reminiscent of CT-rich tracts regulatory of transcription termination (59). If operative, along with the d[ATAGAA] hexamer, then these signals might collectively determine when and where premature transcription termination might occur beyond the terminal triplet of variable pIg-R GAA-tracts.

It is currently unclear if rat 3' UTR PIGR STMSs are biologically functional or regulatory of gene expression, although various observations indirectly support such roles. Firstly, during liver regeneration when hepatocytes proliferate, differential expression of pIg-R mRNAs may occur. However, the processing and translational competence of either one of the natural hepatic pIg-R mRNA populations remains to be determined. Secondly, in preliminary transfection studies, the Wistar *HaeIII*-*TaqI* STMS-containing fragment attenuates reporter gene expression when placed into the 3' UTRs of a family of luciferase-encoding chimeric pSP72 shuttle vectors (T. Aoki, K. S. Koch and H. L. Leffert, unpublished results). Physiological interpretations of *in vitro* studies of this kind are complicated further by the likelihood that triplet repeat and poly_A tail length variability occur between

individual animals and, possibly, individual cells *in vivo*. Thirdly, STMSs in supercoiled genomic and cDNA plasmids are topologically curved and suggestive of non-B structure (60) shown by their S1-nuclease sensitivity and electrophoretic heterogeneity. STMSs like these are implicated in the regulation of transcription (41) and the attenuation of DNA polymerization in cell free systems (6,61). Finally, findings in unrelated systems further indicate novel and unconventional roles for 3' UTRs in gene expression (62) but the relevance of these observations to PIGR expression is unclear.

ACKNOWLEDGEMENTS

This work was supported by grants from the USPHS (DK28215 [HLL], DK15878 [ALJ]), the American Cancer Society (IN93R [KSK]), the University of California Tobacco-Related Disease Research Program (4RT-0076 [HLL]), the University of California Academic Senate (RT96-M and RU148-M [HLL]) and an ACS Eleanor Roosevelt Fellowship (IUAC-WHO [ASG]). We thank Jim Kadonaga, Peter Geiduschek, Bob Wells and Bert Mullhaupt for helpful comments; L. Bibbs for help with automated sequencing; and Pam Alford for assistance. One of us dedicates this paper to the memory of his father, Lionel Seth Leffert.

REFERENCES

- Davidson, E.H. and Britten, R.J. (1979) *Science*, **204**, 1052-1059.
- Vanin, E.F. (1985) *Ann. Rev. Genet.*, **19**, 253-272.
- Jeffreys, A.J., Royle, N.J., Wilson, V. and Wong, Z. (1988) *Nature*, **332**, 278-281.
- Serikawa, T., Kuramoto, T., Hilbert, P., Mori, M., Yamada, J., Dubay, C.J., Lindpainter, K., Ganten, D., Guenet, J.L., Lathrop, G.M. and Beckmann, J.S. (1992) *Genetics*, **131**, 701-721.
- Wells, R.D. (1988) *J. Biol. Chem.*, **263**, 1095-1098.
- Sinden, R.R. and Wells, R.D. (1992) *Current Opin. Biotech.*, **3**, 612-622.
- Richards, R.I. and Sutherland, G.R. (1992) *Nature Genet.*, **1**, 7-9.
- Aaltonen, L.A., Peltomaki, P., Leach, F.S., Sistonen, P., Pylkanen, L., Mecklin, J.P., Jarvinen, H., Powell, S.M., Jen, J., Hamilton, S.R., Petersen, G.M., Kinzler, K.W., Vogelstein, B. and de la Chapelle, A. (1993) *Science*, **260**, 812-816.
- Caskey, C.T., Pizzuti, A., Fu, Y.H., Fenwick, R.G. and Nelson, D.L. (1992) *Science*, **256**, 784-789.
- Kunkel, T.A. (1990) *Biochemistry*, **35**, 8003-8011.
- Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T., Sohn, R., Zemelman, B., Snell, R.G., Rundle, S.A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P.S., Shaw, D.J. and Housman, D. (1992) *Cell*, **68**, 799-808.
- Carango, P., Noble, J.E., Marks, H.G. and Funanage, V.L. (1993) *Genomics*, **18**, 340-348.
- Lu, X.P., Koch, K.S., Lew, D.J., Dulic, V., Pines, J., Reed, S.I., Hunter, T. and Leffert, H.L. (1992) *J. Biol. Chem.*, **267**, 2841-2844.
- Koch, K.S., Fletcher, R.G., Grond, M.P. and Leffert, H.L. (1993) *Hepatology*, **18**, 226-228.
- Banting, G., Brake, B., Braghetta, P., Luzio, J.P. and Stanley, K.K. (1989) *FEBS Lett.*, **254**, 177-183.
- Krajci, P., Kvale, D., Tasken, K. and Brandtzaeg, P. (1992) *Eur. J. Immunol.*, **22**, 2309-2315.
- Williams, A.F. and Barclay, A.N. (1988) *Ann. Rev. Immunol.*, **6**, 381-405.
- Sztul, E.S., Howell, K.E. and Palade, G.E. (1985) *J. Cell. Biol.*, **100**, 1248-1262.
- Brown, W.R. and Kloppel, T.M. (1989) *Immunol. Investig.*, **18**, 269-285.
- Aroeti, B., Casanova, J., Okamoto, C., Cardone, M., Pollack, A., Tang, K. and Mostov, K. (1992) *Int. Rev. Cytol.*, **137B**, 157-168.
- Mazanec, M.B., Kaetzel, C.S., Lamm, M.E., Fletcher, D. and Nedrud, J.G. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 6901-6905.
- Huling, S., Fournier, G.R., Feren, A., Chuntharapai, A. and Jones, A.L. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4260-4264.
- Deitcher, D.L. and Mostov, K.E. (1986) *Mol. Cell. Biol.*, **6**, 2712-2715.

- 24 Brito-Babapulle, V. and Atkin, N.B. (1981) *Cancer Genet. Cytogenet.*, **4**, 215–225.
- 25 Krajci, P., Grzeschik, K.H., Geurts, A.H.M., Olaisen, B. and Brandtzaeg, P. (1992) *Hum. Genet.*, **90**, 215–219.
- 26 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual* 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., V.1–3.
- 27 Schuler, G.D., Altschul, S.F. and Lipman, D.J. (1991) *Proteins Struct. Funct. Genet.*, **9**, 180–190.
- 28 Rychlik, W. and Rhoads, R.E. (1989) *Nucleic Acids Res.*, **17**, 8543–8551.
- 29 Leffert, H.L., Koch, K.S., Sell, S., Skelly, H. and Shier, T. (1983) In *Application of Biological Markers to Carcinogen Testing* (eds Milman, H.A. & Sell, S.). *USEPA Environ. Sci. Res.*, **29**, 271–293.
- 30 Fitzgerald, M.C., Skowron, P., Van Eppen, J.L., Smith, L.M. and Mead, D.A. (1992) *Nucleic Acids Res.*, **20**, 3753–3762.
- 31 Smith, G.R. (1994) *Experientia*, **50**, 234–241.
- 32 Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature*, **263**, 211–214.
- 33 Zuker, M. and Stiegler, P. (1981) *Nucleic Acids Res.*, **9**, 133–148.
- 34 Pandey, N.B. and Marzluff, W.F. (1987) *Mol. Cell. Biol.*, **7**, 4557–4559.
- 35 Drewe, J.A., Verma, S., Frech, G. and Joho, R.H. (1992) *J. Neurosci.*, **12**, 538–548.
- 36 Mostov, K.E., Friedlander, M. and Blobel, F. (1984) *Nature*, **308**, 37–43.
- 37 Krajci, P., Solberg, R., Sandberg, M., Dyen, D., Jahnsen, T. and Brandtzaeg, P. (1989) *Biochem. Biophys. Res. Commun.*, **158**, 783–789.
- 38 Kulseth, M.A., Krajci, P., Myklebost, O. and Rogne, S. (unpublished) L04797.
- 39 Piskurich, J.F., Hsieh, M., Youngman, K.R. and Kaetzel, C.S. (unpublished) U06431.
- 40 Martin-Zanca, D., Hughes, S.H. and Barbacid, M. (1986) *Nature*, **319**, 743–748.
- 41 Drescher, U., Chowdhury, K. and Gruss, P. (1987) *DNA*, **6**, 307–316.
- 42 Shehee, W.R. (unpublished) X14061.
- 43 Tomarev, S.I., Zinovieva, R.D. and Piatigorsky, J. (1993) *Biochem. Biophys. Acta.*, **1216**, 245–254.
- 44 Birnsteil, M., Portmann, R., Busslinger, M., Schaffner, W., Probost, E. and Kressman, A. (1979) *Alfred Benson Symp.*, **13**, 117–132.
- 45 Zepf, N.E. and Ogata, R.T. (1991) *J. Immunol.*, **147**, 2756–2763.
- 46 Usui, J., Kuwano, R., Maeda, T., Araki, K., Sakimura, K., Kushiya, E. and Takahashi, Y. (1987) *Biochem. Int.*, **15**, 809–816.
- 47 Siedlaczek, I., Epplen, C., Rieb, O. and Epplen, J. (1993) *Electrophoresis*, **14**, 973–977.
- 48 Lin, B., Nasir, J., MacDonald, H., Hutchinson, G., Graham, R.K., Rommens, J.M. and Hayden, M.R. (1994) *Hum. Mol. Genet.*, **3**, 85–92.
- 49 Asizawa, T., Dubel, J.R. and Harati, Y. (1993) *Neurology*, **43**, 2674–2678 (1993).
- 50 Wang, Y.H., Amirhaeri, S., Kang, S., Wells, R.D. and Griffith, J.D. (1994) *Science* **265**, 669–671.
- 51 Burfeind, P., Belgardt, B., Szpirer, C. and Hoyer-Fender, S. (1993) *Eur. J. Biochem.*, **216**, 497–505.
- 52 Krowczynska, A.M., Rudders, R.A. and Krontiris, T.G. (1990) *Nucleic Acids Res.*, **18**, 1121–1127.
- 53 Richards, J.E., Gilliam, A.C., Shan, A., Tucker, P.W. and Blattner, F.R. (1983) *Nature*, **306**, 483–487.
- 54 Wels, J.A., Word, C.J., Rimm, D., Der-Balan, G.P., Martinez, H.M., Tucker, P.W. and Blattner, F.R. (1984) *EMBO J.*, **3**, 2041–2046.
- 55 Collier, D.A., Griffin, J.A. and Wells, R.D. (1988) *J. Biol. Chem.*, **263**, 7397–7405.
- 56 Nishimura, S., Jacob, T.M. and Khorana, H.G. (1964) *Proc. Natl. Acad. Sci. USA*, **52**, 1494–1501.
- 57 Cunningham, P.R., Weitzmann, C.J. and Ofengand, J. (1991) *Nucleic Acids Res.*, **19**, 4669–4673.
- 58 Mahadevan, M.S., Amemiya, C.T., Jansen, G., Sabourin, L., Baird, S., Neville, C.E., Wormskamp, N., Segers, B., Lamerdin, J., de Jong, P., Wieringa, B. and Korneluk, R.G. (1993) *Hum. Mol. Genet.*, **2**, 299–304.
- 59 Kerppola, T.K. and Kane, C.M. (1990) *Biochemistry*, **29**, 269–278.
- 60 Bates, A.D. and Maxwell, A. (1993) *DNA Topology* (1st Edit., Oxford University Press, Oxford), pp.1–114.
- 61 Dayn, A., Samadashwily, G.M. and Mirkin, S.M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11406–11410.
- 62 Rastinejad, F. and Blau, H.M. (1993) *Cell*, **72**, 903–917.