Transcript heterogeneity of the human reduced folate carrier results from the use of multiple promoters and variable splicing of alternative upstream exons

Long ZHANG, So C. WONG and Larry H. MATHERLY[1]

Experimental and Clinical Therapeutics Program, Karmanos Cancer Institute, 110 East Warren Avenue, Detroit, MI 48201, U.S.A., and the Department of Pharmacology, School of Medicine, Wayne State University, Detroit, MI 48201, U.S.A.

We previously identified three separate cDNAs (KS6, KS32 and KS43) for the human reduced folate carrier (RFC) with unique 5′ untranslated regions (5′ UTRs) [Wong, Proefke, Bhushan and Matherly (1995) J. Biol. Chem. **270**, 17468–17475]. Multiple RFC transcripts were confirmed in CCRF-CEM cells and transport-up-regulated K562.4CF cells by 5′ rapid amplification of cDNA ends (5′ RACE) and/or primer extension analysis. Two groups of 5′ RACE clones were identified, one containing a variable length sequence identical with the KS43 cDNA 5′ UTR, and another consisting of variants of the KS32 5′ UTR, apparently generated by alternative splicing. The 5′ UTR for the KS6 cDNA was not detected. A single band was detected on Southern blots of CCRF-CEM genomic DNA probed with a 326 bp genomic fragment common to all three cDNA species. The unique 5′ UTRs for the KS43 and KS32 transcripts were localized to separate non-coding exons (exons 1 and 2 respectively), upstream from a large (approx. 3.42 kb) intron; the KS6 5′UTR also mapped to exon 1. Exons 1 and 2 were contiguous with 996 and 342 bp GC-rich 5′ flanking regions (designated Pro43 and Pro32 respectively) that contained multiple SP1 and AP2 but no TATA or CAAT boxes. Both Pro43 and Pro32 exhibited strong promoter activities when cloned in front of a luciferase reporter gene and transfected into HT1080 and K562 cells. By an analysis of promoter deletion mutants we identified two 89 bp tandem repeats that seemed to increase Pro32 activity, and a 240 bp distal sequence that repressed Pro43 activity. Taken together, our results show that multiple human RFC transcripts are encoded by a single gene locus and that the heterogeneous 5′ UTRs result from multiple transcriptional starts and variable splicing of alternative non-coding exons transcribed from separate promoters.

INTRODUCTION

Methotrexate (MTX) is actively transported into mammalian cells by the same reduced folate carrier (RFC)-mediated process normally used by 5-methyl tetrahydrofolate and other reduced folate cofactors [1,2]. Membrane transport of MTX and related antifolates by RFC is critical to their antitumour effectiveness [1,2]. Furthermore, defects in RFC expression result in decreased MTX accumulations within tumour cells and are crucial to the development of MTX resistance [3–7].

In recent years, a number of laboratories have focused attention on identifying and cloning putative components of the RFC system. Hence, mouse [8], hamster [9] and human [10–13] cDNA species for RFC were isolated that on transfection restored MTX sensitivities and MTX uptake in transport-impaired cultured cells. For both the murine [6] and human [5] RFC cDNAs, the restored transport exhibited an extensive array of properties typical of the classic RFC system, including characteristic transport kinetics, inhibition by RFC transport substrates and inhibitors, and trans-stimulation by leucovorin. The full-length human clone, designated KS43, encodes an approx. 92 kDa glycoprotein, detectable by photoaffinity labelling with APA-^{125}I-ASA-Lys[N^{α}-(4-amino-4-deoxy-10-methylpteroyl)-N^{ϵ}-(4-azido-5-^{125}I-salicylyl)-L-lysine] [5,10]. On enzymic deglycosylation the molecular mass of the affinity-labelled species decreased to 65 kDa, the size predicted from the full-length human RFC cDNA [5].

An intriguing feature for both the rodent and human RFCs involves their transcript heterogeneity: multiple cDNA isoforms were described for human [10], mouse [14,15] and hamster [16] RFCs differing in their 5′ untranslated regions (UTRs) and exhibiting variable deletions of coding sequence, as well. A human cDNA (KS32), containing a 5′ UTR distinct from KS43 and a 625 bp deletion in the open reading frame and 3′ UTR, encoded a functional RFC [10]. A range of predicted molecular masses were also reported for the cDNA-encoded RFCs from L1210 cells [6,14,15]. This diversity of RFC forms could result from the alternative splicing of transcripts from a single promoter, the transcription of alternative exons from more than one promoter, and/or the expression of multiple, polymorphic RFC genes. Although initial studies suggested that 5′ UTR and coding sequence heterogeneity for the hamster and mouse RFCs might result at least in part from the combined effects of multiple transcriptional start sites and alternative splicing of 'redundant' coding and non-coding exons [14–16], the extent to which these or other mechanisms contribute to human RFC transcript heterogeneity has not been determined.

In this study we sought to establish the molecular basis for the

Abbreviations used: MTX, methotrexate; 5′ RACE, 5′ rapid amplification of cDNA ends; RFC, reduced folate carrier; 5′ UTR, 5′ untranslated region.

[1] To whom correspondence should be addressed (e-mail matherly@kci.wayne.edu).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number AF046920.

heterogeneity of human RFC transcripts suggested by our earlier isolation of three distinct RFC cDNAs (KS6, KS32 and KS43) with different 5′ UTRs [10]. Our results show that the multiple RFC cDNAs arise from distinct human RFC transcripts encoded by a single RFC gene locus, and that the heterogeneous 5′ UTRs are the result of multiple transcriptional starts and variable splicing of alternative non-coding exons transcribed from separate promoters. Although the heterogeneous 5′ sequence does not alter RFC protein structure, these differences might modulate the expression of RFC at a post-transcriptional level, including effects on transcript processing, stabilities or translational efficiencies.

MATERIALS AND METHODS

Cell culture

The HT1080 human fibrosarcoma cell line was obtained from American Type Culture Collection (Rockville, MD, U.S.A.). The CCRF-CEM lymphoblastic leukaemia line was a gift from Dr. Andre Rosowsky. Both lines were maintained in RPMI 1640 medium containing 10% (v/v) heat-inactivated iron-supplemented calf serum (Hyclone Laboratories), 2 mM L-glutamine, 100 i.u./ml penicillin and 100 μg/ml streptomycin, in a humidified atmosphere at 37 °C in the presence of air/CO_2 (19:1). The characteristics and maintenance of the RFC-up-regulated K562.4CF subline were as described previously [17,18].

Analysis of RFC transcripts by 5′ rapid amplification of cDNA ends (5′ RACE) assay

Double-stranded cDNA with a 5′ Marathon cDNA adaptor was synthesized from CCRF-CEM poly(A)$^+$ RNA with a Marathon cDNA Amplification Kit (Clontech) as recommended by the supplier. Primary PCR and nested PCR reactions were performed with adaptor-ligated cDNA as template, and gene-specific RFCo-1/RFCn-1 primers and AP1/AP2 adaptor primers. RFCo-1 and RFCn-1 primers were synthesized (Genosys Biotechnologies) on the basis of the conserved nucleotide sequences of the human KS6/KS32/KS43 RFC cDNA species [10]. The sequence for RFCn-1 (5′-AGCTCCGGAGGGGACGAAGGTGACACTG-TG-3′) is complementary to a region from −19 to −48 in the cDNAs (where A of the ATG translation initiation site is +1) (Figure 1); RFCo-1 (5′-GCCATGAAGCCGTAGAAGCAA-AGGTAGCACAC-3′) is complementary to +84 to +115 in the RFC cDNAs [10]. Primary PCR was performed for 7 cycles (94 °C for 2 s, 72 °C for 3 min) and 36 cycles (94 °C for 2 s, 67 °C for 3 min) with the gene-specific RFCo-1 primer and the AP1 adaptor primer with a 9600 DNA Thermal Cycler (Perkin Elmer-Cetus). Nested PCR on the primary PCR products was performed for 5 cycles (94 °C for 2 s, 72 °C for 3 min) and 20 cycles (94 °C for 2 s, 67 °C for 3 min) with the nested gene-specific RFCn-1 and AP2 adaptor primers. 5′ RACE products were subcloned and sequenced with Sequenase 2.0 (Amersham/U.S. Biochemical Corp.) and adenosine 5′-[α-[^{35}S]thio]triphosphate (1400 Ci/mmol; Dupont/New England Nuclear).

Southern blot analysis of genomic DNA

Genomic DNA from CCRF-CEM cells were isolated by using the Puregene System (Gentra System); aliquots (10 μg) were digested with restriction enzymes. After electrophoresis on a 0.7% agarose gel, resolved DNAs were transferred to a nylon membrane (Genescreen Plus; Dupont). The blot was hybridized with an [α-^{32}P]dCTP-labelled 326 bp human RFC genomic fragment amplified from CCRF-CEM genomic DNA by PCR with two primers (P1, 5′-CGCAGCCTCTTCTTCAACCGC-3′, nt 622–643 as described in [10]; P5, 5′-ATCAGCGTGGAG-GCAGCATCT-3′, nt 927–948) that prime within the same downstream exon (L. Zhang and L. H. Matherly, unpublished work). The blot was washed under conditions of low (0.5 × SSC/0.1% SDS at 50 °C) or high (0.1 × SSC/0.1% SDS at 68 °C) stringency. The blot was autoradiographed after each stringency wash.

Primer extension analysis

The 5′ ends of the RFC mRNA were mapped by primer extension analysis with the protocol of Triezenberg [19]. Total RNA was isolated from CCRF-CEM and K562.4CF cells with TRIzol Reagent (Life Technologies); poly(A)$^+$ RNA was prepared from total RNA using the PolyATract mRNA Isolation Systems (Promega). The RFCn-1 oligonucleotide complementary to the 5′ UTR from positions −19 to −48 (Figure 1) was end-labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase. After purification on a Sephadex G25 spin column, the end-labelled primer was hybridized at 65 °C for 2 h in hybridization buffer to 3 or 5 μg of poly(A)$^+$ RNA from CCRF-CEM cells, or 50 or 100 μg of total RNA from RFC-up-regulated K562.4CF cells. The annealed primer was extended with avian myeloblastosis virus reverse transcriptase (Promega) for 2 h at 42 °C in reverse transcriptase buffer. Primer extension products were separated on a 6% (w/v) polyacrylamide sequencing gel together with a ^{35}S-labelled sequencing reaction as molecular size markers.

Isolation of upstream genomic clones

A human genomic walking ('Promoter Finder') kit was purchased from Clontech Laboratories and used for upstream genomic walking PCR with two nested synthetic oligonucleotides (RFCo-1 and RFCn-1). PCR conditions were identical with those used for 5′ RACE analysis with this primer set. PCR products were subcloned into the pCR II plasmid by using the TA cloning kit (Invitrogen) and sequenced by using both universal and gene-specific primers as described above. For continued upstream genomic walking, additional gene-specific primers were synthesized on the basis of new sequence data from the previous PCR products.

A human peripheral blood leucocyte genomic library constructed in EMBL3SP6/T7 (Clontech) was screened with [α-^{32}P]dCTP-labelled human RFC cDNA (KS32), which was prepared by random priming (Boehringer-Mannheim). Confirmation that the positive clones obtained from the primary screen contained an upstream RFC genomic sequence was by PCR with primers to the unique 5′ UTR of KS32 cDNA [10]. Positive clones containing KS32 sequence were subjected to two additional screens to obtain pure phage clones. Altogether, three positive clones were obtained by screening 1.2×10^6 plaque-forming units. Phage DNAs from these clones were purified and characterized by restriction mapping and Southern hybridization with the unique 5′ UTRs from KS43, KS6 and KS32 cDNAs [10] as probes. One fragment (RFCg1-3d, 2.6 kb) that hybridized with all three unique 5′ UTRs was subcloned into the pGL3-Basic vector (Promega). A series of deletion constructs containing inserts of different sizes was generated by the Erase-A-Base system (Promega) and sequenced in both directions by using RVprimer3 and GLprimer2 vector universal primers.

RFC–luciferase fusion gene constructs and transient expression assays

RFCg1-3d was digested with *Kpn*I to generate a fragment (−1955

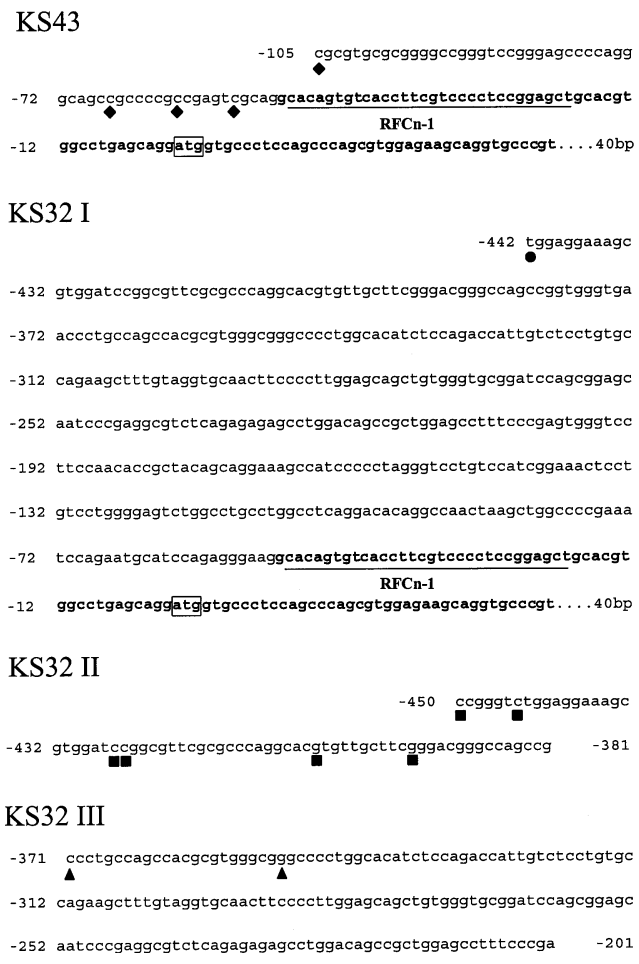

Figure 1 5′ Heterogeneity of human RFC transcripts characterized by 5′ RACE

5′ RACE analysis was performed as described in the Materials and methods section. Sequence data are shown for the four separate groups of 5′ RACE products, including KS43 and KS32I-III, as described in the text. The numbering is relative to the ATG translational start site where the first base (A) is +1. The multiple transcriptional start sites deduced from the 5′ RACE clones are noted by the symbols ◆, ●, ■ and ▲. For the KS43 and KS32I groups both the unique upstream sequence and common sequence from 40 to −49 (the latter designated by bold letters) are indicated. For KS32II and KS32III clones the unique upstream sequence data only are shown. The location of the gene-specific nested primer RFCn-1 is underlined.

to −896) containing 61 bp from the 5′ end of exon 1 (−956 to −896) and a 996 bp 5′ flanking sequence (designated Pro43; −1955 to −957; see Figure 4 and Figure 5, left panel). Another *Kpn*I fragment (−901 to −277) contained the exon 2 sequence (−450 to −277), fused to a 342 bp 5′ flanking region (designated Pro32; −451 to −792) and a 3′ portion of exon 1 (−901 to −793; see Figures 4 and 5). Both fragments were subcloned into the pGL3-Basic vector (Promega) at the *Kpn*I site in both the sense and the anti-sense orientations. For the sense clones, 5′ → 3′ unidirectional deletion mutants were generated by the Erase-A-Base system. All constructs (including sense and anti-sense orientations) were confirmed with appropriate restriction digests and DNA sequencing. Plasmid constructs were isolated with Wizard Midi Prep plasmid isolation kits (Promega) for transient transfections.

RFC–luciferase fusion gene constructs or pGL3-Promoter control vectors (5 μg) were co-transfected with 0.1 μg of pRL-SV40 plasmid (Promega) into approx. 50% confluent HT1080 and K562 cells by using Lipofectin reagent (Life Technologies) in accordance with the manufacturer's protocols. Lipofectin treatments were for 20 h and, after an additional 48 h of incubation in complete RPMI 1640 growth medium containing 10% (v/v) heat-inactivated iron-supplemented calf serum, cells were harvested and lysates were prepared. Firefly luciferase activities were assayed with a Dual-Luciferase Reporter Assay System (Promega) in a Turner TD2420 luminometer and normalized to *Renilla* luciferase activity. RFC promoter activity was expressed as a percentage of the pGL3-Promoter control plasmid, whose value was arbitrarily set at 100.

RESULTS

Analysis of 5′ UTR transcript heterogeneity by 5′ RACE

5′ RACE PCR was used to characterize the extent of 5′ UTR heterogeneity in human RFC transcripts, suggested by our previous finding of three distinct RFC cDNA isoforms (KS6, KS32 and KS43) with differing 5′ termini [10]. Double-stranded cDNA templates were prepared from CCRF-CEM poly(A)$^+$ RNA and ligated to anchor adaptors. After PCR with nested anchor primers and two gene-specific primers, the 5′ RACE

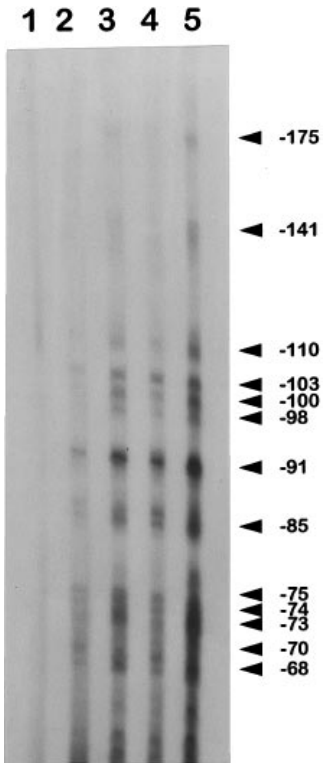

Figure 2 Primer extension analysis of RFC transcription start sites

A 30 bp end-labelled primer (RFCn-1) extending from positions −19 to −48 was annealed to 3 or 5 μg of poly(A)$^+$ RNA from CCRF-CEM cells (lanes 2 and 4 respectively) or to 50 or 100 μg of total RNA from K562.4CF cells (lanes 3 and 5 respectively) and extended upstream with reverse transcriptase. The extension products were analysed on a 6% (w/v) polyacrylamide gel and the sizes (in bp) of the extension products (indicated with arrowheads) were calculated from a sequencing ladder run in parallel. A reaction containing no RNA was included as a control (lane 1).

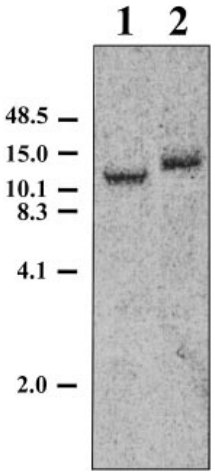

Figure 3 Identification of a single RFC gene locus in CCRF-CEM cells

Genomic DNA (10 μg) from CCRF-CEM cells digested with *Hin*dIII (lane 1) or *Bam*HI (lane 2) was analysed by Southern blotting. The membrane was hybridized with a 326 bp ^{32}P-labelled human RFC genomic fragment generated by PCR, then autoradiographed. The positions of DNA standards (in kb) are indicated at the left.

products (sizes ranging from 36 to 425 bp) were ligated into the pCR II plasmid for transformation and sequencing.

By sequencing 17 clones from both strands we were able to identify four separate groups of RFC transcripts characterized by distinct upstream sequences fused to a common region (starting at position −50, where 1 is the translational start). The DNA sequence data for the 5′ RACE products are shown in Figure 1. (1) Five clones (36–86 bp; designated KS43 and labelled ◆ in Figure 1) contained the 5′ untranslated sequence previously reported for the KS43 cDNA (positions −19 to −95) [10], and one of them extended for an additional 10 bases upstream from the sequence previously described. (2) Two identical clones (KS32 I, marked ● in Figure 1) contained 400 bp of the 5′ non-coding sequence of the published KS32 cDNA (positions −19 to −418) [10] along with an additional 25 bp on the 5′ end. (3) Seven related clones (46–101 bp; KS32 II, marked ■ in Figure 1) contained only the most 5′ non-coding sequence of the KS32 cDNA beginning at position −381. One of these clones contained an additional 7 bp on the 5′ end that were not present in the KS32 I cDNA sequence. Finally, (4) three clones (180–202 bp; KS32 III, marked ▲ in Figure 1) contained KS32 untranslated sequence between positions −201 and −349, or −201 and −371. Although both KS43 and KS32 sequences were identified in the 5′ RACE products, the unique 76 bp 5′ UTR for the partial (1400 bp) cDNA designated KS6 [10] was not detected.

Primer extension analysis of RFC transcription initiation sites

The RFC transcription initiation sites were additionally mapped by primer extension assays. Analysis of extension products from CCRF-CEM and RFC transport-up-regulated K562.4CF cells on a sequencing gel gave results similar to those by 5′ RACE, with numerous extension products ranging from 68 to 175 bp, the most prominent being 85 and 91 bp (Figure 2). There were no obvious differences in the patterns of extension fragments between the CCRF-CEM and K562.4CF sublines. Because of their large sizes (more than 200 bp), no attempt was made to assay the

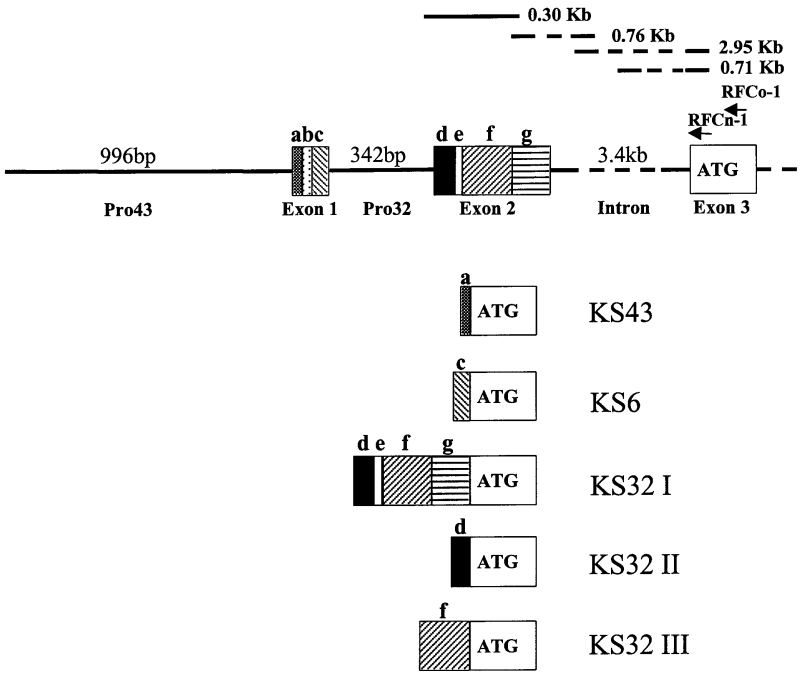

Figure 4 Organization of the human RFC gene upstream region

A schematic is shown depicting the overlapping genomic walking PCR products (0.3–2.95 kb; shown at the top), two putative promoter regions (Pro43 and Pro32), the structures of the upstream exon 1, exon 2, intron and exon 3 (containing the translation start site), and the probable structures of the RFC transcripts (KS32 I-III, KS43 and KS6) deduced from the longest 5′ RACE products in each group or obtained by cDNA library screening. As described, KS6 was isolated as a cDNA [10] but was not detected by 5′ RACE. The variably spliced KS43, KS6 and KS32 I-III exon segments are represented by fragments a (56 bp), b (26 bp), c (82 bp), d (70 bp), e (9 bp), f (171 bp) and g (151 bp). The annealing sites for the RFCn-1 and RFCo-1 primers are also shown.

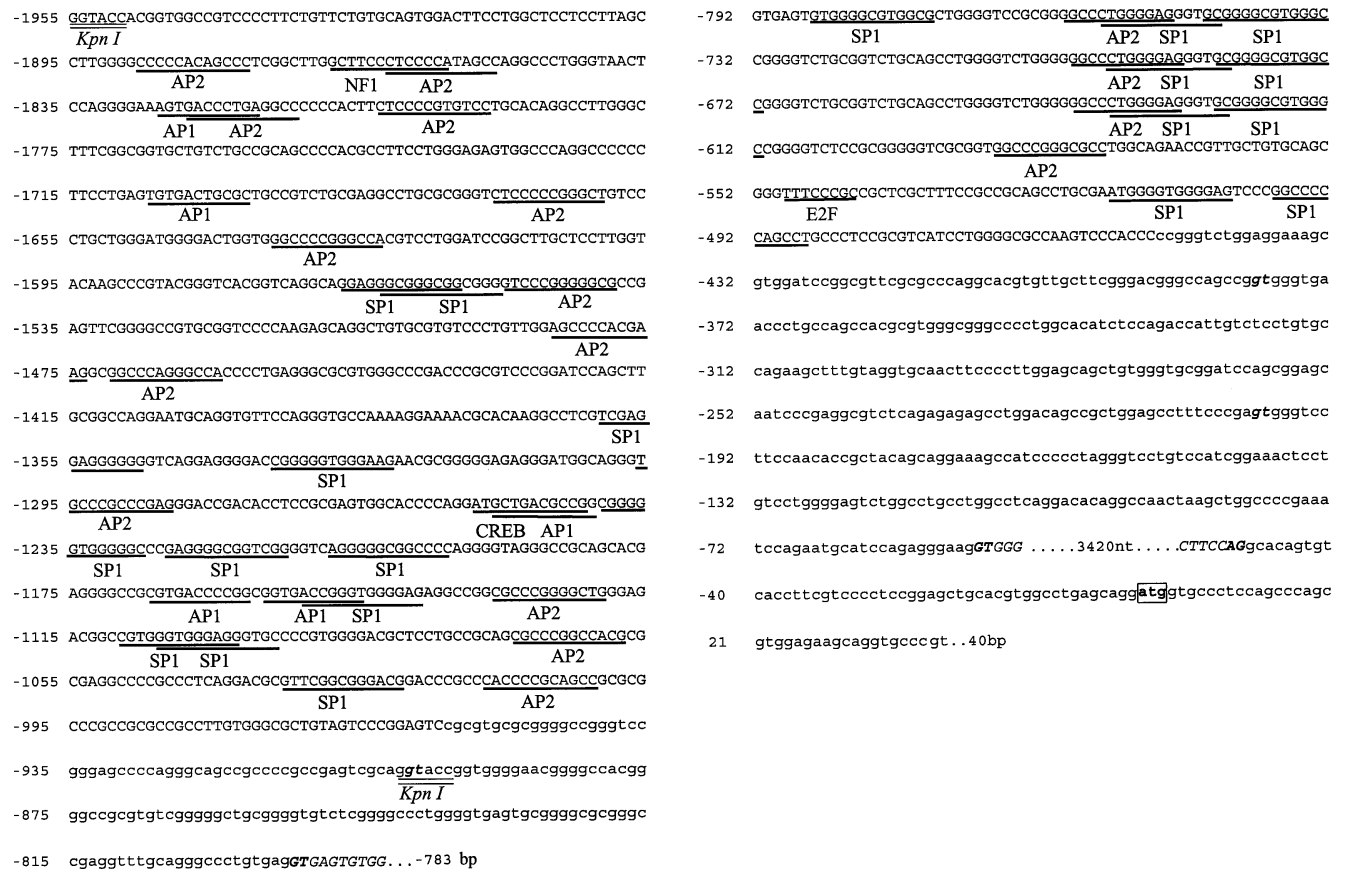

Figure 5 Nucleotide sequence for the human RFC gene upstream region

Sequence data are shown for the putative RFC promoters [Pro43 (left panel) and Pro32 (right panel), noted with capital letters], the intron–exon junctions, and exon 1, exon 2 and part of exon 3, as depicted in Figure 4. The numbers designate the non-intron genomic sequence where the first base (A) of the putative translation start (shown in bold type and rectangle in the right panel) is +1 and are positive and negative in the 3′ and 5′ directions respectively. Putative splice donor and acceptor sites are shown in bold italics. Putative regulatory elements/transcription factor-binding sites (AP2, SP1, NF1, CREB and EF2 in the sense strand) are underlined. The *Kpn*I cutting sites are also noted.

upstream initiation sites represented by the KS32I and KS32II 5′ RACE products (Figure 1) by primer extension analysis.

A single RFC gene locus encodes multiple RFC transcripts

Several possibilities were considered to explain the RFC transcript diversity suggested from the results of the 5′ RACE and primer extension assays. These included (1) the existence of multiple polymorphic RFC genes, (2) a single RFC gene for which the sole primary transcript was alternatively spliced, and/or (3) the transcription of multiple RFC transcripts that differ in their 5′ UTRs from separate promoters within a single human RFC gene, followed by variable splicing to generate a greater diversity of RFC transcript forms.

To assess the possibility that the KS32 and KS43 transcripts are encoded by separate RFC genes, genomic DNA species from CCRF-CEM cells were digested with restriction enzymes (*Hin*dIII and *Bam*HI) and analysed by Southern blotting. A 326 bp genomic probe based on the common nucleotide sequence (nt 622 to 948) for the KS6, KS32 and KS43 cDNA species [10] was used for hybridization. At both low-stringency (not shown) and high-stringency washes, only a single band (12 and 14 kb for the *Hin*dIII and *Bam*HI digests respectively) hybridized with the radiolabelled probe (Figure 3), consistent with the notion that the multiple RFC transcripts are encoded by a single RFC gene locus.

Isolation and sequence analysis of the RFC 5′ genomic sequence

The suggestion from our Southern blotting experiments that a single RFC gene was capable of encoding both the KS32 and KS43 RFC transcripts led us to explore the upstream organization of the RFC gene. Upstream sequence was isolated by a combination of genomic walking and screening of a human genomic library.

Upstream genomic walking was performed with commercial libraries of adaptor-ligated genomic DNA isolated from human leucocytes, starting with primers based on the conserved nucleotide sequences of the KS6/KS32/KS43 RFC cDNAs [10]. Three rounds of upstream walking resulted in overlapping PCR products of 710, 2950, 761 and 300 bp, altogether encompassing 3.9 kb of contiguous upstream genomic sequence (designated RFCg1-1; Figure 4). By aligning the RFCg1-1 nucleotide sequence with those for the KS32 cDNA [10] and the KS32I-III 5′ RACE products, we were able to identify a large intron (approx. 3.42 kb) at position −49 that contained consensus sequences for both splice donor (5′-GT-) and acceptor (-AG-3′) junctions, flanked by an upstream exon (designated exon 2; Figure 4 and Figure 5, right panel). The sequence of exon 2 was identical with that previously published for the KS32 cDNA [10] from positions −50 to −418 (where +1 is the translation start), with the exception of a G omission at −355 and the addition of 32 bases at the 5′ end (Figure 5, right panel). Exon 2 was preceded

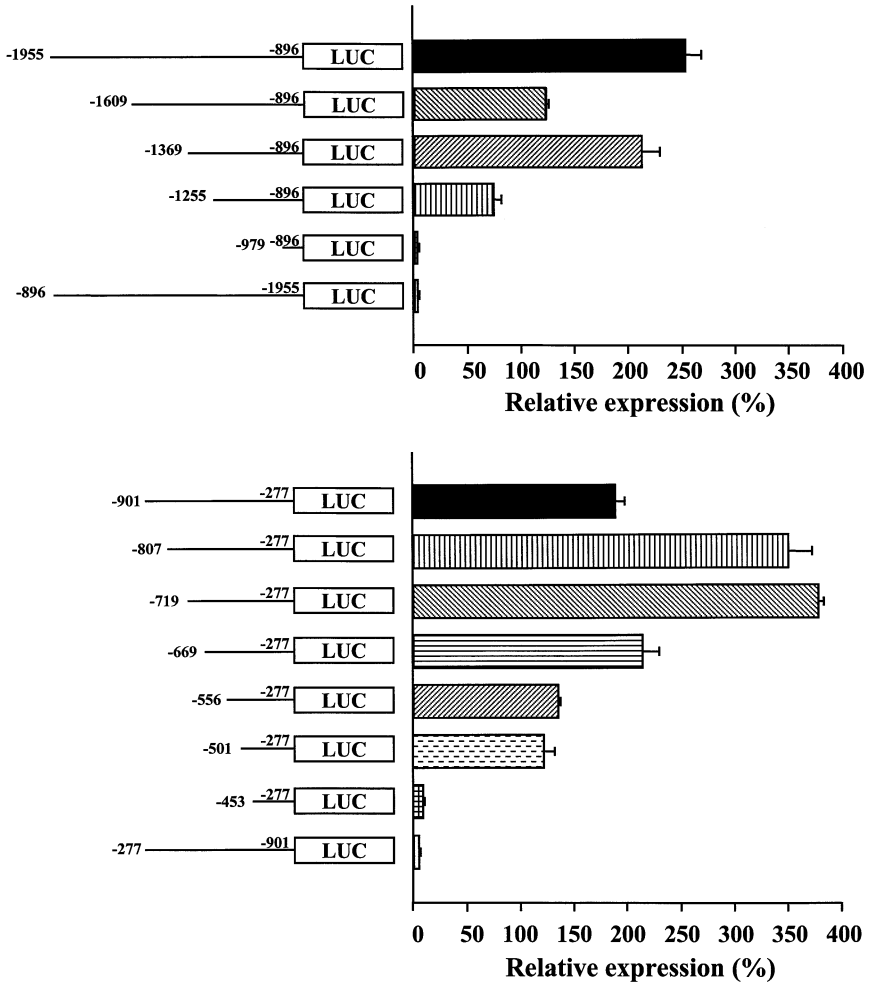

Figure 6 Activity assays for the dual RFC promoters in HT1080 cells

5′ Deletions in the 5′ flanking regions including the putative Pro43 (upper panel) and Pro32 (lower panel) promoters were generated by successive exonuclease III and S1 nuclease digestions. The 5′ termini for the promoter fragments are shown. These sense promoter fragments were ligated 5′ to a firefly luciferase reporter gene in the promoter-less pGL3-Basic plasmid and the deletion constructs were transfected into HT1080 cells. The anti-sense constructs were generated by ligating the full-length fragments, including the Pro43 and Pro32 promoters, into the pGL3-Basic luciferase reporter plasmid in the anti-sense orientation. Firefly luciferase activity was assayed as described in the text and relative activity levels (reported as means ± S.E.M. for three or four experiments) in the different constructs were normalized to *Renilla* luciferase activity and expressed as a percentage of the pGL3-Basic plasmid containing the SV40 promoter (positive control), whose value was arbitrarily set at 100.

immediately upstream by 57 bp of unique sequence. The size of the 3.42 kb intron was confirmed by PCR amplification from CCRF-CEM genomic DNA. Similarly, the junction between the 3.42 kb intron and exon 3 (containing the ATG translational start) was verified in genomic DNA by PCR amplification and sequencing of the PCR products. In spite of this success in genomic walking over approx. 3.9 kb of upstream sequence, additional walking was not possible because of the high GC content further upstream from exon 2.

To obviate the limitations of genomic walking, we screened a human leucocyte genomic library in EMBL3 bacteriophage with the RFC KS32 cDNA. Three positive clones (RFCg1-31, RFCg1-42 and RFCg1-44; 17, 16 and 20 kb respectively) were isolated. Restriction enzyme mapping and Southern hybridizations indicated that each of the clones contained all three unique 5′ UTRs previously reported for the RFC cDNAs. A 2.6 kb *Pvu*II restriction fragment (designated RFCg1-3d) from RFCg1-31 that hybridized with the three unique 5′ UTRs was completely sequenced in both directions. Sequence analysis indicated that RFCg1-3d contained 174 bp from the 5′ end of exon 2, preceded by a 342 bp 5′ flanking sequence (Pro32) and the 5′ UTRs of both the KS6 and KS43 cDNA species, separated by a 26 bp gap (collectively designated exon 1); exon 1 was preceded immediately upstream by a 996 bp 5′ flanking region (Pro43). The upstream organization of the human RFC gene is depicted in Figure 4. Sequence data are shown in Figure 5.

Assay of promoter activities for the Pro32 and Pro43 5′ flanking regions

The proximity of the highly GC-rich Pro43 and Pro32 fragments to the transcription initiation sites for the KS43 and KS32I-III transcripts suggested that these 5′ flanking fragments might represent dual promoters. Computer analysis [20,21] of these regions with the TRANSFAC database failed to identify either canonical TATA or CAAT motifs within the first 100 bp

upstream from the transcription initiation sites; however, a number of putative transcription factor binding sites (SP1, AP1, AP2, E2F, CREB and NF1) were indicated (Figure 5). The presence of numerous SP1 sites close to the transcription initiation site(s) is characteristic of promoters of TATA-less genes [22].

To assess the capacities of these putative promoter sequences to direct the transcription of a luciferase reporter gene, a series of nested deletion constructs including both the Pro32 and Pro43 fragments were fused to a firefly luciferase reporter gene in the promoterless pGL3-Basic plasmid. The undeleted proximal and distal fragments (positions −277 to −901 and −896 to −1955 respectively) were ligated in both the sense and the anti-sense orientations. Constructs were transiently expressed in HT1080 and K562 cells, both of which express high levels of RFC transcripts on Northern blots. Firefly luceriferase activity was normalized to *Renilla* luciferase activity, encoded by a co-transfected pRL-SV40 plasmid and expressed relative to a pGL3-Promoter control plasmid containing an SV40 promoter.

For both the Pro43 and Pro32 fragments, expression of the longest constructs in HT1080 and K562 cells resulted in luciferase activities far exceeding that for the promoterless pGL3 construct (results not shown) or constructs in which nearly all of the 5′ flanking regions were deleted (Figure 6 shows results for HT1080 cells). Although both promoter activities were orientation-dependent, maximal activity for Pro32 exceeded that for Pro43 by approx. 35% [compare the activities for −719 Pro32 and −1955 Pro43 constructs; Figure 6 (lower panel) and Figure 6 (upper panel) respectively]. Deletion of 182 bp including 109 bp of exon 1 from the longest (−901) Pro32 construct resulted in an approximate doubling in promoter activity that, with continued deletion between −719 and −501, resulted in a progressive approx. 65% decrease from the maximal level. This was accompanied by the loss of two 89 bp highly GC-rich (82%) tandem repeats (one from −762 to −674, the other from −701 to −614 with a G omission at −674), including a series of putative AP2 and SP1 sites. Although a putative E2F site was identified in Pro32 (at −555; Figure 5), deletion of this sequence (between −556 and −501) had no obvious effect on promoter activity (Figure 6, lower panel). However, the loss of an additional 52 bp (to position −453) resulted in the complete abolition of Pro32 promoter activity. For Pro43, both the −1955 and −1369 promoter constructs exhibited near-maximal luciferase activity; a 240 distal region (−1369 to −1609) exerted a repressive effect, suggesting the presence of *cis*-regulatory factors in this fragment. Further deletion of Pro43 (to −1255) resulted in a progressive and, eventually, the complete (at −979) loss of promoter activity (Figure 6, upper panel).

DISCUSSION

Our previous finding of three separate cDNA isoforms for the human RFC strongly implied the existence of multiple RFC transcripts with distinct 5′ non-coding regions. In the present study we explored the extent and molecular basis of this apparent heterogeneity in RFC transcripts. Interestingly, the frequency of the major RFC transcript corresponding to the KS43 cDNA described by at least three laboratories [10–12] was comparatively low in our 5′ RACE analysis of CCRF-CEM transcripts. Further, a separate KS6 form [10] was not detected at all. Rather, most of the RFC transcripts seemed to be variants of the full-length non-coding KS32 5′ UTR (designated KS32 I–III), apparently generated by alternative splicing.

The possibility that the multiple RFC transcripts might originate from separate but homologous RFC genes was considered by probing a Southern blot of CCRF-CEM genomic DNA with a 326 bp DNA fragment common to all three cDNA species. Our finding of a single hybridizing fragment in two different restriction digests strongly suggests the existence of only a single RFC gene locus and argues that differential splicing of alternative non-coding exons probably accounts for the 5′ UTR transcript heterogeneity.

Indeed, analysis of the 5′ organization of the RFC gene allowed us to localize the unique 5′ UTRs for the KS43 and KS32 transcripts to separate non-coding exons (exons 1 and 2 respectively). These were located immediately upstream from a large (approx. 3.42 kb) intron and exon 3 containing the translation start site. The KS6 cDNA sequence also mapped to exon 1 (within 26 bp of the 3′ end of KS43), suggesting that this minor transcript form in K562.4CF cells [10] was probably a product of alternative splicing of the larger upstream exon. Exons 1 and 2 were contiguous with unique GC-rich 5′ flanking regions (Pro43 and Pro32 respectively), which, although devoid of classical TATA or CAAT motifs, contained numerous putative transcription factor-binding sites (SP1, AP2, E2F and CREB). The 5′ organization of the human RFC gene therefore resembles those for the hamster [16] and murine [14,15] genes in the presence of a large intron upstream from the ATG start site and alternative upstream exons, and a TATA-less 5′ flanking region with numerous SP1 consensus sites. However, there seems to be little sequence similarity in the upstream non-coding exons or the 5′ flanking regions between the rodent RFCs [14–16] and the human RFC gene sequence described here. The human Pro32 and Pro43 fragments both exhibited strong promoter activities when they were fused to a luciferase reporter gene and transiently expressed in HT1080 and K562 cells. In addition to putative transcription factor-binding sites deduced from computer analysis, promoter deletion studies implicated other important regulatory regions, including two 89 bp tandem repeats that appeared to increase Pro32 activity, and an inhibitory 240 bp distal sequence in the Pro43 promoter.

The lack of CAAT or TATA motifs and the presence of numerous SP1 sites preceding multiple transcription initiation sites are features typical of constitutively expressed housekeeping genes or proto-oncogenes [22]. Although the detection of RFC transcripts in a host of normal human tissues (placenta, small intestine, colon, thymus, prostate, testes, ovaries, spleen and peripheral blood leucocytes) [23] is consistent with the notion of the expression of a constitutive RFC, the presence of multiple upstream non-coding exons and promoters nevertheless suggests a complex transcriptional regulation for the human RFC gene. Multiple promoter use could regulate RFC at the transcriptional level and maintain expression in a large number of tissues owing to the presence of unique transcription factor-binding sites that respond differently to endogenous and exogenous stimuli [24]. Further, another level of control could result from alternative upstream exons, including KS32 and KS43 [10], which can be translated with different efficiencies [24]. 5′ UTR length or secondary structure can also influence the processing of primary transcripts to mature mRNA species or the rates of mRNA turnover [24].

In conclusion, our results demonstrate that the heterogeneous 5′ UTRs of the human RFC transcripts arise from at least two promoters and the alternative splicing of dual 5′-non-coding exons. Even though the 5′ sequence divergence in itself does not alter protein structure, these differences can exert post-transcriptional effects on RFC expression at the level of transcript processing, stabilities or translational efficiencies, as noted above. Our studies provide a basis for further analysis of the regulation of RFC gene expression by multiple promoters; additional studies

are undoubtedly needed for a better definition of the transcriptional elements of the basal RFC promoters. The further elucidation of the functional elements of each promoter might shed light on the factors that regulate the gene expression and alternative splicing of RFC transcripts in human tissues and tumours.

This work was supported by NIH grant CA53535.

REFERENCES

1 Goldman, I. D. and Matherly, L. H. (1985) Pharmacol. Ther. **28**, 77–100
2 Sirotnak, F. M. (1985) Cancer Res. **45**, 3992–4000
3 Sirotnak, F. M., Moccio, D. M., Kelleher, L. E. and Goutas, L. J. (1981) Cancer Res. **41**, 4447–4452
4 Gorlick, R., Goker, E., Trippet, T., Steinherz, P., Elisseyeff, Y., Mazumdar, M., Flintoff, W. F. and Bertino, J. R. (1997) Blood **89**, 1013–1018
5 Wong, S. C., McQuade, R., Proefke, S. A. Bhushan, A. and Matherly, L. H. (1997) Biochem. Pharmacol. **53**, 199–206
6 Brigle, K. E., Spinella, M. J., Sierra, E. E. and Goldman, D. I. (1995) J. Biol. Chem. **270**, 22974–22979
7 Gong, M., Yess, J., Connolly, T., Ivy, S. P., Ohnuma, T., Cowan, K. H. and Moscow, J. A. (1997) Blood **89**, 2494–2499
8 Dixon, K. H., Lampher, B. C., Chiu, J., Kelley, K. and Cowan, K. H. (1994) J. Biol. Chem. **269**, 17–20
9 Williams, F. R. M., Murray, R. C., Underhill, T. M. and Flintoff, W. F. (1994) J. Biol. Chem. **269**, 5810–5816
10 Wong, S. C., Proefke, S. A., Bhushan, A. and Matherly, L. H. (1995) J. Biol. Chem. **270**, 17468–17475
11 Moscow, J. A., Gong, M., He, R., Sgagias, M. K., Dixon, K. H., Anzick, S. L., Mettzer, P. S. and Cowan, K. H. (1995) Cancer Res. **55**, 3790–3794
12 Williams, F. M. R. and Flintoff, W. F. (1995) J. Biol. Chem. **270**, 2987–2992
13 Prasad, P. D., Ramamoorthy, S., Leibach, F. H. and Ganapathy, V. (1995) Biochem. Biophys. Res. Commun. **206**, 681–687
14 Brigle, K. E., Spinella, M. J., Sierra, E. E. and Goldman, D. I. (1997) Biochim. Biophys. Acta **1353**, 191–198
15 Tolner, B., Roy, K. and Sirotnak, F. M. (1997) Gene **189**, 1–7
16 Murray, R. C., Williams, F. R. M. and Flintoff, W. F. (1996) J. Biol. Chem. **271**, 19174–19179
17 Matherly, L. H., Angeles, S. M. and Czajkowski, C. A. (1992) J. Biol. Chem. **267**, 23253–23260
18 Matherly, L. H., Czajkowski, C. A. and Angeles, S. M. (1991) Cancer Res. **51**, 3420–3426
19 Triezenberg, S. J. (1992) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 4.8.1–4.8.5, John Wiley & Sons, New York
20 Wingender, E. (1988) Nucleic Acids Res. **16**, 1879–1902
21 Prestridge, D. S. (1991) Comput. Appl. Biosci. **7**, 203–206
22 Azizkhan, J. C., Jensen, D. E., Pierce, D. E. and Wade, M. (1993) Crit. Rev. Eukar. Gene Express. **3**, 229–254
23 Nguyen, T. T., Dyer, D. L., Dunning, D. D., Rubin, S. A., Grant, K. E. and Said, H. M. (1997) Gastroenterology **112**, 783–791
24 Ayoubi, T. A. Y. and Van De Ven, W. J. M. (1996) FASEB J. **10**, 453–460

Received 11 February 1998; accepted 24 March 1998