Structure of the human DNA ligase I gene

Patricia Noguez⁺, Deborah E. Barnes, Harvey W. Mohrenweiser¹ and Tomas Lindahl^{*}

Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK and ¹Human Genome Center, Biomedical Sciences Division L-452, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

Received May 29, 1992; Revised and Accepted July 2, 1992

ABSTRACT

The gene encoding DNA ligase I, the major DNA ligase activity in proliferating mammalian cells, maps to human chromosome 19q13.2–13.3. We have determined the complete structure of the gene, which is composed of 28 exons spanning 53kb on this chromosome. The first exon is untranslated, and utilises a GC dinucleotide instead of the canonical GT splice donor. The 5' flanking region lacks a TATA box and is highly GC-rich, as is characteristic of a 'housekeeping' gene. In common with the promoters of genes encoding other DNA replication enzymes, such as DNA polymerase α , the 5' flanking region of the DNA ligase I gene contains recognition elements for several transcription factors which may mediate increased expression in quiescent cells in response to growth factors.

INTRODUCTION

DNA ligases are involved in DNA replication and are also required for DNA repair and genetic recombination (1). Three distinct DNA ligases have been identified in mammalian cell nuclei (2). DNA ligase I represents the major DNA ligase activity in proliferating cells and its physical and biochemical properties have been extensively characterised (for a recent review, see 3). A cDNA encoding DNA ligase I has been isolated and identifies a 3.2kb transcript by northern hybridisation analysis, encoding a protein of 919 amino acid residues (4). The DNA ligase I gene (*LIG1*) has been localised to human chromosome 19q13.2–13.3 and is distal to the DNA repair gene, *ERCC1* (5), as determined by fluorescence *in situ* hybridisation (6). Three genes have been localised immediately distal of *ERCC1* in 19q13.2–13.3 by automated DNA sequencing, but none of these is identical with the *LIG1* gene (7). The myotonic dystrophy (DM) locus is distal of this sequenced region (8).

Human DNA ligase I is able to complement the replication defect of both *Saccharomyces cerevisiae* and *Escherichia coli* DNA ligase mutants (4, 9). Direct evidence of a role for DNA ligase I in DNA replication, and DNA repair, comes from analysis of a mutant human cell line that exhibits both a reduced

rate of joining of Okazaki fragments during DNA replication, and hypersensitivity to DNA damaging agents. Two missense mutations occurring in different alleles of the DNA ligase I gene were detected in these cells (10), which were obtained from a patient with immunodeficiencies (11). One of the mutations occurred in the active site region of the enzyme and caused functional inactivation. The other allele, encoding a malfunctioning but partially active enzyme, was also present in the mother of the proband and represents the first example of an inherited mutation in a mammalian replication enzyme.

Levels of DNA ligase I are much higher in proliferating than in non-growing mammalian cells and are up-regulated when quiescent cells are induced to enter S-phase (3, 12–15). In contrast, DNA ligase I is present throughout oogenesis and early development of *Xenopus laevis* and other amphibia (16, 17). The DNA ligase encoded by the *CDC9* gene of *S. cerevisiae* is, in common with other budding yeast proteins required for DNA synthesis, periodically expressed under cell cycle control (18). In contrast, the *Schizosaccharomyces pombe* DNA ligase gene, *cdc17⁺*, does not show such periodic transcription (19). Little is known about the regulation of DNA ligase I expression in human cells. Post-translational control of DNA ligase I activity may be achieved through phosphorylation of the N-terminal domain of the protein (20). Transcriptional regulation would presumably be ineffective in cycling cells due to the long half-life of the protein (21) but may be important in allowing stationary cells to respond to growth factors or DNA damage. Here, we have analysed the structure of the DNA ligase I gene and its 5' flanking region as a first step towards characterising the regulation of this enzyme in DNA replication and repair.

MATERIALS AND METHODS

DNA ligase I genomic clones

Three positive clones (f20031, f22104, f22135) have previously been isolated from an arrayed human chromosome 19-specific cosmid library screened with the full-length human DNA ligase I cDNA (6). A linked cosmid clone (f24839) was also identified (6). Restriction digest and Southern hybridisation analysis by standard techniques revealed the extent of DNA ligase I sequences

* To whom correspondence should be addressed

⁺ Present address: Laboratoire de Mutagenese et Pathologie Humaine, Institut Jacques Monod, 2 Place Jussieu, 75251 Paris Cedex 05, France

contained within these cosmid clones (figure 1). For the completion of this study, an additional cosmid (clone f21689, figure 1) was identified by screening arrayed cosmids from the library (22) with an oligonucleotide corresponding to nucleotides 28 to 57 of the cDNA sequence, and with a cDNA fragment (nucleotides 28 to 369) generated by the polymerase chain reaction (PCR; 23).

DNA amplification (PCR)

Fragments of the DNA ligase I gene were amplified from cosmid clones by PCR using Amplitaq (Perkin-Elmer Cetus). Thirty cycles of 1 min at 91°C, 2 min at 55°C, 2 or 5 min (depending on the anticipated length of the product) at 72°C were carried out in a DNA thermal cycler (Perkin-Elmer Cetus). Oligonucleotide primers were based on the cDNA sequence (4) and synthesised on an Applied Biosystems model 380B DNA synthesiser. In some cases, primers incorporated Eco RI or Bam HI restriction sites and PCR products were cloned into the M13 mp18 or mp19 vectors (24) following digestion with the appropriate restriction enzyme. Alternatively, PCR products were end-repaired with the Klenow fragment of *E. coli* DNA polymerase (Boehringer Mannheim) and ligated into M13 vectors that had been linearised with Sma I. PCR products were sequenced in M13 by the dideoxy chain termination method (25) using Sequenase (United States Biochemical). Direct sequencing of double strand (ds) PCR products or cosmid clones was performed by cycle sequencing (26), essentially by the dsDNA Cycle Sequencing System (BRL, Life Technologies, Inc.).

S1 mapping and primer extension analysis

A probe for primer extension analysis was prepared by cloning a cDNA PCR product as a 256bp Eco RI-Bam HI restriction fragment in M13mp18. This fragment spans the translation start site with the 5' end defined by an Eco RI site 15bp upstream of the initiation methionine. A ³²P-labelled single strand probe was isolated after primed synthesis and Eco RI digestion by the 'prime-cut' method (27, 28). Similarly, a 354bp PCR product (encompassing the entire first untranslated exon and additional 5' upstream sequences from cosmid f21689) was used to generate a probe for S1 mapping. Each probe was incubated with 1μg human poly A⁺ RNA in 80% formamide, 500mM NaCl, 40mM Pipes pH6.4, 1mM EDTA at 85°C for 3 min and then incubated for 18hr at 37 or 55°C (primer extension) or 60°C (S1 mapping) prior to extension with AMV reverse transcriptase (Boehringer Mannheim) or digestion with S1 nuclease (Boehringer Mannheim). Products were analysed by electrophoresis in 6% denaturing polyacrylamide gels using sequencing ladders of known templates as size markers.

RESULTS

Exon-intron structure of the DNA ligase I gene

Four DNA ligase I genomic clones have previously been isolated from an arrayed human chromosome 19-specific cosmid library (6). Restriction enzyme digestion and Southern hybridisation analysis with oligonucleotides complementary to the DNA ligase I cDNA, or with cDNA fragments, revealed the extent of DNA ligase I sequences contained within these cosmid clones. The results of this analysis are shown in figure 1. Cosmids f22104 and f22135 have inserts of approximately 30kb, spanning exons 9 to 26 of the gene. Cosmids f24839 and f20031 have larger inserts of approximately 35 and 42kb, spanning exons 7 to 27,

and 3 to 24, respectively. Cosmid f24839 extends beyond the previously identified polyadenylation signal (4), with exon 27 containing the stop codon and entire 3' untranslated region (UTR) of the DNA ligase I cDNA, and so completes the 3' end of the gene. However, all four cosmids lack the 5' exons of the gene, with cosmid f20031 defining the 5' limit of the contig formed by the four overlapping clones. In order to obtain the 5' end of the DNA ligase I gene, the arrayed chromosome 19-specific cosmid library was re-screened with both an oligonucleotide complementary to the 5' UTR of the DNA ligase I cDNA and a 342bp cDNA PCR fragment extending from the 5' UTR and including the first 249 nucleotides of the coding sequence. A single cosmid (f21689) was identified by both probes. Cosmid f21689 extends as far as exon 4 at its 3' limit and so overlaps cosmid f20031.

The exon-intron structure of the DNA ligase I gene was determined by PCR amplification and DNA sequencing. DNA ligase I sequences were amplified from cosmid clones using primers based on the cDNA sequence. The presence or absence of introns was then determined by comparison of the length of amplified products with the distance separating the two primers on the cDNA sequence. The exact positions of exon-intron boundaries were defined by sequencing PCR products and identifying splice consensus sites (29). The results of this analysis are summarised in figure 2. The DNA ligase I gene is composed of 28 exons ranging in size between 54 and 173bp. The 27 introns range in size from 140bp to 5kb.

Identification of the transcription initiation site

We have now obtained a longer DNA ligase I cDNA clone from the human hepatoma (Hep G2) library from which a putative full-length clone was originally isolated (4). This longer clone extends the 31bp 5' UTR of the original clone by 89bp and places an in-frame stop codon 99bp upstream of the initial methionine, confirming that this was correctly assigned. This extended sequence appears in the databases (Genbank M36067) and is shown in figure 3. The 5' UTR of this extended cDNA is interrupted in the gene sequence by an intron of approximately 4.2kb. The first exon (exon -1) is therefore untranslated, with exon 1 containing the translation initiation methionine. Comparison of the 5' UTR of the cDNA with the gene sequence indicates that the untranslated exon -1 utilises a GC dinucleotide instead of the canonical GT donor splice site (29). A GT dinucleotide occurs 4bp downstream of the GC splice site in the gene sequence (figure 2) but even if readthrough occurred such that this alternative splice site was used in a proportion of transcripts, this would not place an upstream ATG in-frame with the coding sequence. The length of the extended cDNA sequence is in good agreement with the size of the mRNA detected by northern hybridisation analysis (4). Both this cDNA and the slightly shorter clone previously described (4) contain the translation start ATG and the entire coding sequence.

S1 mapping and primer extension analysis were employed to try to identify the major transcription initiation site and confirm that the extended cDNA sequence represents a full-length clone. The cDNA sequence would correspond to a transcription initiation site 120bp upstream of the translation start. Primer extension analysis identified multiple products, corresponding to heterogeneous initiation at various sites between 120 and 200bp upstream of the initial methionine (data not shown). Some extension products are likely to reflect incomplete strand synthesis by the reverse transcriptase due to local secondary structure. For

Figure 1. Schematic representation of the human DNA ligase I gene. Exons are represented by solid boxes, numbered -1 to 27 from 5' to 3' in the direction of transcription. Exon -1 is untranslated. The connecting line represents intervening sequences. The span of DNA ligase I sequences in the five cosmid clones analysed is shown, together with a partial restriction map.

EXON			INTRON			EXON		
No.	Size	5'donor	No.	Size	3'acceptor	No.		
-1	63	CGC AAG gaa gt---	1	4200	---gtccag GAG CAG	1		
1	74	ATC AT gtg agc---	2	2800	---ctctag G TCA	2		
2	90	CCA AA gtat gt---	3	700	---ttccag G GCG	3		
3	136	GGC CAG gtag gt---	4	3500	---ttccag AAG CCT	4		
4	127	ACA G gtg aga---	5	3800	---ccccag CT CGG	5		
5	96	GAA G gtc agc---	6	1000	---cgtcag AA GAG	6		
6	108	TCC A gtg agt---	7	1000	---ccgcag AA GCA	7		
7	123	TTC A gtg agt---	8	242	---tcttag CC CCC	8		
8	79	GAG GG gtg agt---	9	5000	---ccgtag A CCC	9		
9	81	CAG AA gtg agg---	10	280	---ccctag G GTT	10		
10	57	GCT CG gta act---	11	3500	---ctccag G CTC	11		
11	173	ACA G gta agg---	12	2900	---gtgaag GT CGG	12		
12	167	AGT GCT gtg agt---	13	500	---ctccag TCC ACA	13		
13	77	GCT AG gta agc---	14	800	---ccccag G TCC	14		
14	92	CAA G gtg agc---	15	300	---tcttag AA TTC	15		
15	100	TTC TG gtg aga---	16	1200	---ccccag C GAG	16		
16	86	CCA G gtc agg---	17	800	---tccag GG ATT	17		
17	116	GCA CAG gtat gg---	18	1800	---tgacag ATC CAC	18		
18	96	CCC AAG gtg ggc---	19	2900	---tttcag ATT AAA	19		
19	111	CGC AAG gtag ca---	20	600	---accag GAG GTG	20		
20	72	GGA GAG gtg agt---	21	4800	---ctgcag TCC CTG	21		
21	145	AAA G gtg atg---	22	1400	---cctcag AC TCC	22		
22	83	CTC AAG gtg att---	23	1000	---tccag CTG AAG	23		
23	153	TGC AAG gtc ctg---	24	2000	---ctgcag CTT GGA	24		
24	54	CTC AAG gtg agc---	25	1400	---ctgcag GCG CTG	25		
25	144	GGC CTG gtg agt---	26	1400	---ccgcag GTG GAT	26		
26	93	GCT CAG gtg agg---	27	140	---ccgcag GTG GCC	27		
27	81	ACC TAC TAA						

Figure 2. Exon-intron organisation of the DNA ligase I gene. Exons and introns are numbered from 5' to 3' in the direction of transcription. Sizes of exons, and approximate sizes of introns (as determined by gel electrophoresis of PCR products obtained with flanking primers), are given (bp). Splice donor and acceptor sites are shown. Exon sequences are in uppercase letters, intron sequences in lowercase letters. Only partial intron sequence, flanking splice junctions, was obtained. Dinucleotides corresponding to the splicing consensus are shown in bold.

example, a shorter product equivalent to strand termination within the palindromic sequence -TGCCCGGGCA- (located 81 to 90bp upstream, in exon -1) was consistently seen. There was no evidence from primer extension experiments or northern

hybridisation analysis for an alternatively spliced mRNA that would lack the untranslated exon -1 and commence at exon 1. S1 mapping experiments gave inconclusive results due to incomplete digestion of the probe, even at large enzyme excess

Figure 3. Structure of the 5' flanking region of the DNA ligase I gene. **a.** Restriction map of the region for the enzymes shown. The gene is shown schematically; exons are denoted by boxes, with coding sequences filled. **b.** Sequence of the region upstream of the start ATG (shown in outline letters). The cDNA sequence is shown in bold with the extent of the shorter cDNA originally described (4) underlined. The position of the first intron is denoted by an asterisk. Putative Sp1 and ATF binding sites are underlined by a dashed line, or boxed, respectively.

and incubation temperatures up to 45°C. This is probably due to the high GC content of the 5' flanking region (see below).

Sequence of the 5' flanking region

Southern hybridisation analysis of cosmid f21689 with an oligonucleotide complementary to the extreme 5' end of the DNA ligase I cDNA identified a 15kb positive Eco RI fragment (data not shown). As there is an Eco RI site 4.3kb downstream of this oligonucleotide in the DNA ligase I gene, it is estimated that cosmid f21689 extends at least 11kb upstream of the 5' limit of the cDNA sequence and contains the functional promoter of the gene. The sequence of the gene was determined for approximately 400bp upstream of the cDNA sequence/putative transcription start site (+1) by direct sequencing of cosmid f21689 (figure 3). This upstream region has a GC content of 72% with a high concentration of CpG dinucleotides. There is no Goldberg-Hogness TATA box (30) or CCAAT box (31). However, several potential transcription factor binding sites were identified by their homology to consensus binding sequences. Four inverted Sp1 binding sites, CCGCCCC (32) are located between -405 to -399, -151 to -146bp, -72 to -66bp, -55 to -48bp and -49 to -44bp. From -40 to -35, the sequence GACGTCT

differs by only one nucleotide from the ATF binding consensus, (T/G) (T/A) CGTCA, found in several adenovirus and cellular promoters (33).

DISCUSSION

The complete exon-intron structure of the DNA ligase I gene has been determined by analysis of five overlapping cosmid clones. The gene is composed of 28 exons of 54 to 173bp separated by 27 introns ranging in size from 140bp to 5kb, the largest introns being located towards the 5' end of the gene. These results are in broad agreement with the basic structure of eukaryotic genes (34, 35). The 28 exons span 53kb, in close agreement with the approximate size of the gene estimated by Southern hybridisation analysis of human genomic DNA with DNA ligase I cDNA probes (6). The first exon is untranslated and utilises a GC dinucleotide instead of a GT splice donor; this is the only exception throughout the DNA ligase I gene to the GT-AG splicing consensus (29). A few cases have been reported where a functional splice site has involved an altered GT dinucleotide (36–38). In these cases, as here, a GC dinucleotide has been used, as described for another human DNA repair gene, *ERCC3* (39). A human DNA ligase I cDNA has been isolated from a tonsillar library that lacks sequences encoding amino acid residues 7 to 36 of the protein (15). The missing sequences correspond to exon 2 as determined here, and this shorter cDNA apparently represents an alternatively spliced form of the DNA ligase I mRNA. The physiological significance of this alternatively spliced form is unclear, although the corresponding truncated cDNA has reduced activity when expressed in *S.cerevisiae* (15). The alternatively spliced message was not detected by northern hybridisation analysis of fibroblast or HeLa RNA (4, 15).

The deduced amino acid sequence of the human DNA ligase I cDNA is 40% homologous to the smaller DNA ligases of *S. cerevisiae* (40) and *S. pombe* (41). This homology is confined to the carboxyl-terminal regions of the respective proteins (4) which presumably define essential enzyme functions. A minimum domain of the human enzyme required for catalytic activity *in vivo* (4, 9) and *in vitro* (42) is encoded by exons 8 to 27 of the DNA ligase I gene, with the active site of the enzyme (43, 44) encoded by exon 17. The *S. cerevisiae* DNA ligase gene has no introns, while the *S. pombe* gene has two small introns towards the 5' end of the gene, in a region with no homology to the human enzyme. The two point mutations identified in the DNA ligase I gene of a mutant human cell line (10) occur near the active site of the enzyme in exon 17, and in an evolutionarily conserved region corresponding to exon 23 of the human gene. Detailed knowledge of the exon-intron structure will allow analysis of the DNA ligase I gene in other mutant human cell lines where defects in DNA metabolism may correlate with distinct forms of inherited disease.

Primer extension analysis indicated heterogeneous initiation of transcription from multiple sites. This is consistent with a GC-rich 5' flanking region which lacks a TATA box, and is typical of a 'housekeeping' gene (45). There is a high concentration of CpG dinucleotides and methylation of these potential acceptor sites may control expression of the gene (46). These features, together with 'GC-box' recognition elements for the cellular transcription factor Sp1 (32), are also features of other genes encoding enzymes involved in DNA synthesis/replication, such as DNA polymerase α (*pol* α ; 47), thymidine kinase (*tk*; 48).

DNA topoisomerase II α (*topo II α* ; 49) and proliferating cell nuclear antigen (*PCNA*; 50), which are regulated by various growth stimuli. Sp1 elements are able to direct transcription when present in either orientation (32) and occur in the inverted orientation with respect to the mRNA in the DNA ligase I gene.

Other ubiquitous features of the *pol α* , *tk*, *topo II α* and *PCNA* genes are inverted CCAAT boxes, which may mediate complex regulation of the respective genes. Cell cycle-specific interaction of DNA-binding proteins with the inverted CCAAT element correlates with increased transcription of the *tk* gene prior to S-phase (51). However, *tk* protein levels are apparently independent of the level of *tk* mRNA (48) and are regulated by specific degradation of the protein (dependent on the presence of carboxyl-terminal residues) at mitosis (52). The expression of *pol α* , in common with most serum-inducible genes, does not vary through the cell cycle of actively dividing cells (53), but an inverted CCAAT box enhances activity of the *pol α* promoter even in cycling cells (47). An inverted CCAAT element located in intron 1 of the *PCNA* gene negatively regulates cell cycle expression of this gene (54). The inverted CCAAT motif was not seen in the putative promoter of the DNA ligase I gene, which differs from the other four DNA synthesis/replication genes in this respect. A putative ATF binding site (33) has been observed here and such sites have also been detected in the *pol α* , *topo II α* and *PCNA* genes.

DNA ligase I levels increase upon proliferation of quiescent cells (12–15), and a report has appeared on the apparent induction of DNA ligase synthesis as a result of cellular exposure to ultraviolet light (55). The transcription factor binding sites in the 5' flanking region of the DNA ligase I gene might mediate induction of this gene in response to such stimuli. Analysis of the 5' flanking region by targeted deletion and *in vitro* mutagenesis will facilitate detailed characterisation of the control of expression of the DNA ligase I gene.

ACKNOWLEDGEMENTS

We thank Iain Goldsmith and his staff for oligonucleotide synthesis, and Susan Tsujimoto for identification of the 5' cosmid f21689. This work was supported by the Imperial Cancer Research Fund and performed under the auspices of the U.S. Department of Energy, Office of Health and Environmental Research by the Lawrence Livermore National Laboratory under Contract W-7405-ENG-48.

REFERENCES

- Kornberg, A. and Baker, T.A. (1992) DNA Replication, 2nd edition, Ch. 9. W.H. Freeman & Co., New York.
- Tomkinson, A.E., Roberts, E., Daly, G., Totty, N.F. and Lindahl, T. (1991) *J. Biol. Chem.*, **266**, 21728–21735.
- Lindahl, T. and Barnes, D.E. (1992) *Annu. Rev. Biochem.*, **61**, 251–281.
- Barnes, D.E., Johnston, L.H., Kodama, K., Tomkinson, A.E., Lasko, D.D. and Lindahl, T. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6679–6683.
- Van Duin, M., Koken, M., van den Tol, J., ten Dijke, P., Odijk, H., Westerveld, A., Bootsma, D. and Hoeijmakers, J.H.J. (1987) *Nucleic Acids Res.*, **15**, 9195–9213.
- Barnes, D.E., Kodama, K., Tynan, K., Trask, B.J., Christensen, M., de Jong, P.J., Spurr, N.K., Lindahl, T. and Mohrenweiser, H.W. (1992) *Genomics*, **12**, 164–166.
- Martin-Gallardo, A., McCombie, W.R., Gocayne, J.D., FitzGerald, M.G., Wallace, S., Lee, B.M.B., Lamerdin, J., Trapp, S., Kelley, J.M., Liu, L.-I., Dubnick, M., Johnston-Dow, L.A., Kerlavage, A.R., de Jong, P., Carrano, A., Fields, C. and Venter, J.C. (1992) *Nature Genet.*, **1**, 34–39.
- Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevin, M., Tsilfidis, C., Chen, C., Alleman, J., Wormskamp, N.G.M., Vooijs, M., Buxton, J., Johnson, K., Smeets, H.J.M., Lennon, G.G., Carrano, A.V., Korneluk, R.G., Wieringa, B. and de Jong, P.J. (1992) *Nature*, **355**, 548–551.
- Kodama, K., Barnes, D.E. and Lindahl, T. (1991) *Nucleic Acids Res.*, **19**, 6093–6099.
- Barnes, D.E., Tomkinson, A.E., Lehmann, A.R., Webster, A.D.B. and Lindahl, T. (1992) *Cell*, **69**, 495–503.
- Webster, A.D.B., Barnes, D.E., Arlett, C.F., Lehmann, A.R. and Lindahl, T. (1992) *Lancet*, **339**, 1508–1509.
- Soderhall, S. (1976) *Nature*, **260**, 640–642.
- Teraoka, H. and Tsukada, K. (1982) *J. Biol. Chem.*, **257**, 4758–4783.
- Elder, R.H. and Rossignol, J.M. (1990) *Biochemistry*, **29**, 6009–6017.
- Petrini, J.H., Huwiler, K.G. and Weaver D.T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7615–7619.
- Aoufouchi, S., Hardy, S., Prigent, C., Philippe, M. and Thiebaud, P. (1991) *Nucleic Acids Res.*, **19**, 4395–4398.
- Aoufouchi, S., Prigent, C., Theze, N., Philippe, M. and Thiebaud, P. (1992) *Dev. Biol.*, **152**, in press.
- Lowndes, N.F., Johnson, A.L. and Johnston, L.H. (1991) *Nature*, **350**, 247–250.
- White, J.H.M., Barker, D.G., Nurse, P. and Johnston, L.H. (1986) *EMBO J.*, **5**, 1705–1709.
- Prigent, C., Lasko, D.D., Kodama, K., Woodgett, J.R. and Lindahl, T. (1992) *EMBO J.*, **11**, 2925–2933.
- Lasko, D.D., Tomkinson, A.E. and Lindahl, T. (1990) *J. Biol. Chem.*, **265**, 12618–12622.
- Tynan, K., Olsen, A., Trask, B., de Jong, P., Thompson, J., Zimmerman, W., Carrano, A. and Mohrenweiser, H. (1992) *Nucleic Acids Res.*, **20**, 1629–1636.
- Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Ehrlich, H.A. and Arnheim, N. (1985) *Science*, **230**, 1350–1353.
- Norrander, J., Kempe, T. and Messing, J. (1983) *Gene*, **26**, 101–106.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Murray, V. (1989) *Nucleic Acids Res.*, **17**, 8889.
- Farrell, P.J., Deininger, P.L., Bankier, A. and Barrell, B. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1565–1569.
- Biggin, M., Farrell, P.J. and Barrell, B. (1984) *EMBO J.*, **3**, 1083–1090.
- Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P. (1986) *Annu. Rev. Biochem.*, **55**, 1119–1150.
- Goldberg, M. (1979) Ph.D. thesis, Stanford University, Stanford, California.
- Efstathiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., Deriel, J.K., Forget, B.G., Weissman, S.W., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) *Cell*, **21**, 653–668.
- Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) *Trends in Genetics*, **11**, 20–23.
- Lee, K.A., Hai, T.Y., SivaRaman, L., Thimmappaya, B., Hurst, H.C., Jones N.C. and Greene, M.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8355–8359.
- Traut, T.W. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2944–2948.
- Hawkins, J.D. (1988) *Nucleic Acids Res.*, **16**, 9893–9905.
- Avvedimento, V.E., Vogeli, G., Yamada, Y., Maizel, J.V.Jr., Fastan, I. and Crombrughe, B. (1980) *Cell*, **21**, 689–696.
- Ebril, C. and Niesing, J. (1980) *EMBO J.*, **2**, 1339–1343.
- Broderick, T.P., Schaff, D.A., Bertino, A.M., Dulsh, M.K., Tischfield, J.A. and Stambrook, P.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3349–3353.
- Weeda, G., Ma, L., van Ham, R.C.A., van der Eb, A.E. and Hoeijmakers, J.H. (1991) *Nucleic Acids Res.*, **19**, 6301–6308.
- Barker, D.G., White, J.H.M. and Johnston, L.H. (1985) *Nucleic Acids Res.*, **13**, 8323–8337.
- Barker, D.G., White, J.H.M. and Johnston, L.H. (1987) *Eur. J. Biochem.*, **162**, 659–667.
- Tomkinson, A.E., Lasko, D.D., Daly, G. and Lindahl, T. (1990) *J. Biol. Chem.*, **265**, 12611–12617.
- Tomkinson, A.E., Totty, N.F., Ginsburg, M. and Lindahl, T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 400–404.
- Yang, S.W. and Chan, J.Y.H. (1992) *J. Biol. Chem.*, **267**, 8117–8122.
- Dynan, W.S. (1986) *Trends in Genetics*, **2**, 196–197.
- Bird, A., Taggart, M., Frommer, M., Miller, O.J. and Macleod, D. (1985) *Cell*, **40**, 91–99.
- Pearson, B.E., Nasheuer, H. and Wang, T. S.-F. (1991) *Mol. Cell. Biol.*, **11**, 2081–2095.

48. Flemington, E., Bradshaw, H.D., Traina-Dorge, V., Slagel, V. and Deininger, P.L. (1987) *Gene*, **52**, 267–277.
49. Hochhauser, D., Stanway, C.A., Harris, A.L. and Hickson, I.D. (1992) *J. Biol. Chem.*, in press.
50. Travali, S., Ku, D., Rizzo, M.G., Ottavio, L., Baserga, R. and Calabretta, B. (1989) *J. Biol. Chem.*, **264**, 7466–7472.
51. Knight, G.B., Gudas, J.M. and Pardee, A.B. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8350–8354.
52. Kaufmann, M.G. and Kelly, T.J. (1991) *Mol. Cell. Biol.*, **11**, 2538–2546.
53. Wahl, A.F.M., Geis, A.M., Spain, B.H., Wong, S.W., Korn, D. and Wang, T. S.-F. (1988) *Mol. Cell. Biol.*, **8**, 5016–5025.
54. Alder, H., Yoshinouchi, M., Prystowsky, M.B., Appasamy, P. and Baserga, R. (1992) *Nucleic Acids Res.*, **20**, 1769–1775.
55. Mezzina, M. and Nocentini, S. (1978) *Nucleic Acids Res.*, **5**, 4317–4328.