Vol. 201, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 30, 1994 Pages 63-70

EXPRESSION AND cDNA CLONING OF HUMAN HMGI-C PHOSPHOPROTEIN

Umesh A. Patel[1], Antonella Bandiera[2], Guidalberto Manfioletti[2], Vincenzo Giancotti[2], Kai-Yin Chau[1] and Colyn Crane-Robinson[1*]

[1] Biophysics Laboratories, University of Portsmouth, PO1 2DT, UK

[2] Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Università di Trieste, 34127, Italy

Received March 31, 1994

SUMMARY. The HMGI family contains three members: I, Y and I-C. HMGI and HMGY are alternative splicings of the same gene and are essential transcription factors at several genetic loci. HMGI-C is transcribed from a different gene and is observed only in highly transformed cells. This work shows that human I-C is present in a more restricted range of cell types than I/Y and is absent from hemopoietic cells, as noted for mouse I-C. However, high expression in a human hepatoma line allowed the cloning of the cDNA and 812 bp of 5'-untranslated, 330 bp of coding and 58 bp of 3'-untranslated DNA were sequenced. The open reading frame showed 4 amino acid substitutions and one additional amino acid when compared to mouse I-C, none of them in the basic DNA binding motifs.

The high mobility group (HMG) proteins consist of three distinct families, HMG1/2, HMG14/17 and HMGI, a common feature being their solubility in 5% perchloric acid (1). The mouse HMGI family consists of three members: HMGI and HMGY are alternatively spliced forms (2), whilst HMGI-C is closely related to I and Y in structural organisation having about 50% sequence identity and expressed from a different gene (3). Rodent HMGI, Y and I-C proteins are preferentially expressed in rapidly dividing, malignant and/or undifferentiated cells and levels of expression in cultured cells correlate with the degree of transformation (4-6). Whereas I and Y may be present in untransformed cells at very low levels, I-C only appears on transformation (5,7). Several studies have implied or directly shown that HMGI/Y are ancillary transcription factors (8-12). In particular, Thanos and Maniatis (12) showed that HMGI/Y are components of the human β-interferon (IFN-β) promoter complex and together with NF-κB are essential for viral induction of the gene. Furthermore, HMGI/Y is capable of binding to ATF-2 and c-Jun, both bZIP proteins required for viral induction and it was proposed that HMGI/Y facilitates the formation of a multiprotein complex at the IFN-β promoter (13). The binding of HMGI/Y to promoters is mediated by three basic segments of 9 amino acids each (14) which are also present in mouse HMGI-C (3). All three HMGI proteins have a highly acidic C-terminal tail, the site of multiple

* Corresponding author. **Fax:** (44)-705 842053.

0006-291X/94 $5.00

phosphorylations (15–17), which together with the remaining segments of the proteins must be the sites of interaction with other transcriptional factors participating in the multiprotein complexes of the combinational promoters in which the HMGI family participates. Human HMGI–C has been detected in a hepatoma cell line (18) and the present work was undertaken to define more fully the expression pattern of HMGI–C in human cells and clone the human gene as its cDNA.

MATERIALS AND METHODS

Cell culture. Human hepatoma cell lines Hep G2, Hep 3B, PLC/PRF/5 were grown in Dulbecco's modified Eagles medium supplemented with 10% heat inactivated foetal calf serum (FCS). Human erythroleukaemic K562 and lymphoblastoid leukaemic MOLT–4 cells were grown in RPMI 1640 medium supplemented with FCS.

Construction and screening of a human hepatoma cDNA library. An oligo(dT) primed and a random plus specific primed cDNA library was constructed using a cDNA synthesis kit (Pharmacia). Two separate first strands were synthesised using reverse transcriptase on human hepatoma mRNA template from PCL/PRF/5 cells. One first strand was synthesised using an oligo(dT) primer and the other synthesised using a mixture of random hexamers plus UP13, a gene specific primer (5'–GAGGATGTCTCTTCAG T–3') corresponding to a segment near the carboxy terminus of the mouse I–C sequence. The cDNAs from the two syntheses were mixed, size selected, purified on a Sephacryl S–400 spun column and then ligated with phosphorylated *EcoR1/Not1 adaptors into EcoR1 digested λmax1 arms (Clontech) before in vitro* packaging using Giga Pack II Gold (Stratagene). $8x10^5$ plaques from the unamplified cDNA library were transferred to four duplicate sets of Hybond–N (Amersham) membranes. The DNA was UV crosslinked and screened for the presence of HMGI–C sequences using as probe a 224 bp fragment (clone HC29, see below), labelled by random priming to a specific activity of $3x10^8$ cpm/μg. Membranes were hybridised in QuickHyb (Stratagene) containing the denatured labelled probe at $1x10^6$ cpm/ml and 400 μg/ml sonicated salmon sperm DNA at 65°C for 4 hr. Membranes were then washed twice with 2xSSC at room temperature, once with 2xSSC, 0.1% SDS at 65°C for 30 min and finally once with 0.2xSSC, 0.1% SDS at 65°C for 10 min and autoradiographed.

Generation of a human 224 bp DNA probe using reverse transcriptase and PCR (RT–PCR). Two degenerate oligonucleotide primers were designed on the basis of the mouse HMGI–C sequence: 5'–CGCGGGATCCCAGCA(G/A)CA(A/G)GA(G/A)CCAACCTG(T/C)GA(G/A)CC–3' (UP5) and 5'–GCGCGAATTCTC(T/C)TC(T/C)TCNGCNGA(T/C)TC(T/C)TG–3' (UP8). First strand synthesis was at 42°C for 1 hr in 20 μl containing 1 μg total human hepatoma RNA from PLC/PRF/5 cells, 50 pmol of UP8, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 1 mM dNTP and 200 units M–MLV reverse transcriptase (BRL). The reaction mixture was then heated at 95°C for 3 min and made up to 100 μl containing 50 pmol of UP5, 50 pmol of UP8, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$ and 2.5 units Amplitaq DNA polymerase (Cetus). 30 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1 min plus one final cycle of extension at 72°C for 5 min were carried out in a temperature cycler (Coy, USA). The resulting 224 bp DNA fragment was ligated into BlueScript KS^+ (Stratagene) and transformed into *E. coli* XL1–blue. The sequence of clone HC29 showed a high level of identity to the mouse HMGI–C cDNA, demonstrating that it represented a segment of the human homologue. Clone HC9 which was used as probe for Northern analysis was also prepared by the above RT–PCR method.

RT–PCR cloning of the 3' end of the coding sequence. Two antisense oligonucleotide primers (UP14 & UP15) were designed on the basis of the mouse HMGI–C 3'–untranslated sequence at 58 and 90 bases downstream of the stop codon. The sense primer represented the human HMGI–C sequence QQQEPTGEP. The RT–PCR reaction was carried out as described above. The resulting 306 bp fragment from the use of UP14 was ligated into M13mp18 and sequenced. Clone HC37 gave identical sequence to HC29 in the common coding region plus an additional 36 bp sequence up to the stop codon, followed by 58 bp of 3'–untranslated sequence up to UP14.

Protein extraction and polyacrylamide gel electrophoresis. HMG proteins and histone H1 were selectively extracted from culture cells and tumours with 5% (mass/vol) perchloric acid (PCA) and acetone precipitated. First–dimension PAGE (250 mm x 0.7 mm), second–dimension SDS gel electrophoresis (250 mm x 1 mm) and silver staining were carried out as described (18,19).

Preparation and purification of anti-HMGI-C antibodies. Rabbits were injected with 200 μg of HPLC-purified recombinant mouse HMGI-C protein, (expressed in bacteria), mixed with complete Freund's adjuvant. They were boosted with the same amount of protein in incomplete Freund's adjuvant every 3 weeks for 2 months. For affinity purification of antibodies, Affi-Prep 10 (BioRad) was coupled with 0.8 mg of HMGI-C recombinant protein. 5 ml of immune serum, diluted 2x with phosphate-buffered saline (PBS), was incubated with the affinity matrix in the presence of 0.1% Triton X-100. Elution of antibodies used 0.2 M glycine-HCl buffer (pH 2.8). Rabbit polyclonal antibodies against the peptide WPQQVVQKKPAQ of the mouse HMGI-C were obtained from Research Genetics (USA) and affinity purified as above.

Western blotting. First-dimension acetic acid-urea gels (without fixing or staining) were incubated for 30 min in 0.75 M Tris base, 0.1% SDS (pH≈11). Proteins were transferred for 40 min to a membrane (PVDF, Immobilon P) using a Trans-Blot SD Semi-Dry Transfer Cell (BioRad) in 0.075 M Tris, 0.01% SDS and stained with 0.2% Ponceau S in 3% trichloracetic acid. Membranes for development with antibody were first saturated with 0.05 M Tris pH 7.5, 0.5 M NaCl, 5% (mass/vol) non-fat dry milk, 0.05% Tween 20, 0.2% BSA, then incubated with anti-HMGI-C affinity purified antibodies (diluted 1:30) for 2 hr at room temperature. Membranes were then washed three times with the above saturation solution and incubated with goat anti-rabbit IgG conjugated with horseradish peroxidase (Pierce), diluted 1:1000, for 1 hr at room temperature. After washing with the same buffer, they were rinsed with 0.2 M Tris pH 7.5, 0.5 M NaCl and developed with 0.3% (mass/vol) 4-chloro-1-naphthol in cold methanol mixed 1:5 with 0.02 M Tris pH 7.5, 0.5 M NaCl. Finally, 30% H_2O_2 solution was added (1:2000) and the reaction stopped by washing with PBS.

RESULTS AND DISCUSSION

Expression Patterns. Fig. 1A shows Coomassie stained gels of perchloric acid extracted proteins from 3 hepatoma cell lines (PLC/PRF/5, Hep G2 and Hep 3B) and 2 lines representative of a total of 5 human transformed hemopoietic cells (K562, MOLT-4, SUP T1, KG1 and KG1A). Protein I-C appears to be present in PLC/PRF/5 and Hep 3B cells and can just be seen in Hep G2. To make the presence of I-C in Hep G2 cells more evident, the 2-D gel was also stained with silver. In contrast, no I-C was detected in the extracts from K562 or MOLT-4 cells and this was confirmed by silver staining (data not shown). To be certain of the identity of the spots, western blotting was carried out using antibodies against recombinant mouse HMGI-C and a peptide representing an I-C specific sequence in the mouse protein,

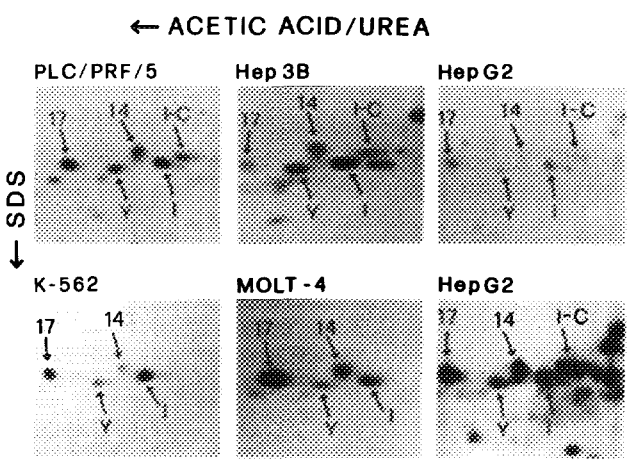

Figure 1A. Two-dimensional gel electrophoresis. Proteins were extracted with PCA from 3 human hepatoma cell lines (PLC/PRF/5, Hep 3B and Hep G2) and 2 hemopoietic cell lines, K562 and MOLT-4. The bottom right panel of Hep G2 was silver stained and the other 5 were Coomassie Blue stained.

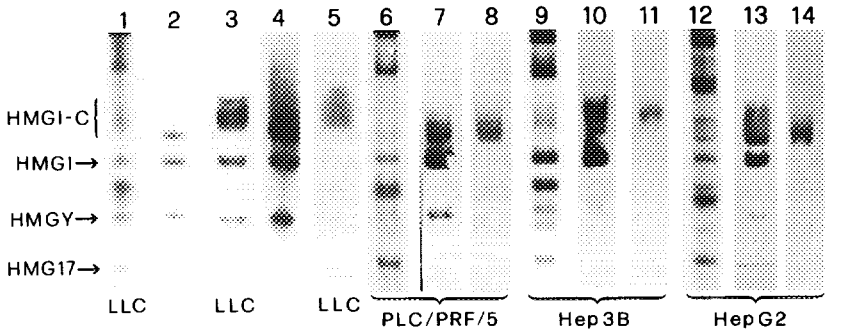

Figure 1B. Western transfers of HMG proteins. Separation was on acetic acid–urea gels, developed with anti (mouse) HMGI–C antibodies (lanes 3, 4, 7, 10, 13) or with anti–I–C peptide antibodies (lanes 5, 8, 11, 14). The Coomassie stained gels from each cell type are shown in lanes 1, 6, 9, 12. Lanes 2 (Coomassie stained) and 4 are an artificial mixture of mouse HMGI, HMGY and HMGI–C proteins.

not present in HMGI or Y (3). Proteins from the three human hepatoma cell lines were compared with extracts from mouse Lewis lung carcinomas (LLC) and with an artificial mixture of HMGI, HMGY and HMGI–C proteins purified by HPLC from LLC cells. Lanes 1 and 2 of Fig. 1B show Coomassie staining and lanes 3 to 5 show immunostaining of the mouse proteins. It is clear that the anti–recombinant–I–C antibodies recognise all 3 proteins (lanes 3 and 4). In contrast, the anti–I–C peptide recognises only I–C protein (lane 5) and emphasises the fact that extracted I–C protein consists of several components, probably a consequence of multiple phosphorylations. A similar situation is seen with the human proteins and the data show the presence of I–C in all 3 hepatoma lines. However, western transfers of perchloric extracts from K562, MOLT–4, SUP T1, KG1 and KG1A showed no indication of the presence of I–C protein when treated with the anti–peptide antibody (data not shown). Northern analysis of RNA from the above cell types was conducted using as probe the human HMGI–C cDNA clone HC9. Total RNA was extracted from cells in exponential growth and the data confirmed the results of western blotting: HMGI–C mRNA was detected in PLC/PRF/5, Hep G2 and Hep 3B cells but not in K562, MOLT–4, SUP T1, KG1 or KG1A cells (data not shown). Thus although high levels of I and Y are seen in all human transformed cells, I–C is expressed in a restricted range of cell types. This is similar to mouse cells for which I–C expression is also more restricted than I and Y (7).

cDNA Cloning. A cDNA library was constructed in two parts as described and these mixed before ligating into λmax1 arms and *in vitro* packaging. The resulting cDNA library was probed with the 224 bp fragment of human I–C cDNA (clone HC29) obtained from RT–PCR by using as template another RNA sample from PLC/PRF/5 cells (see Materials and Methods). Primary and secondary screening of the library produced 8 positive clones. After restriction mapping, the two longest clones HC31 (1075 bp) and HC11 (1095 bp) were sequenced, HC31 on both strands and HC11 fully on one strand and partially on the other. This showed that both clones had resulted from priming with the specific oligonucleotide (UP13) and all the remaining 6 clones were likewise found to have been primed by the specific oligonucleotide. To obtain sequence 3' to the specific primer, RT–PCR was used with total RNA from PCL/PRF/5 cells and reverse primers selected from the 3'–untranslated sequences of mouse I–C. One of these RT–PCR amplifications (clone HC37) was successful and this was sequenced to reveal the remainder of the coding sequence and 58 nucleotides of 3'–untranslated sequence.

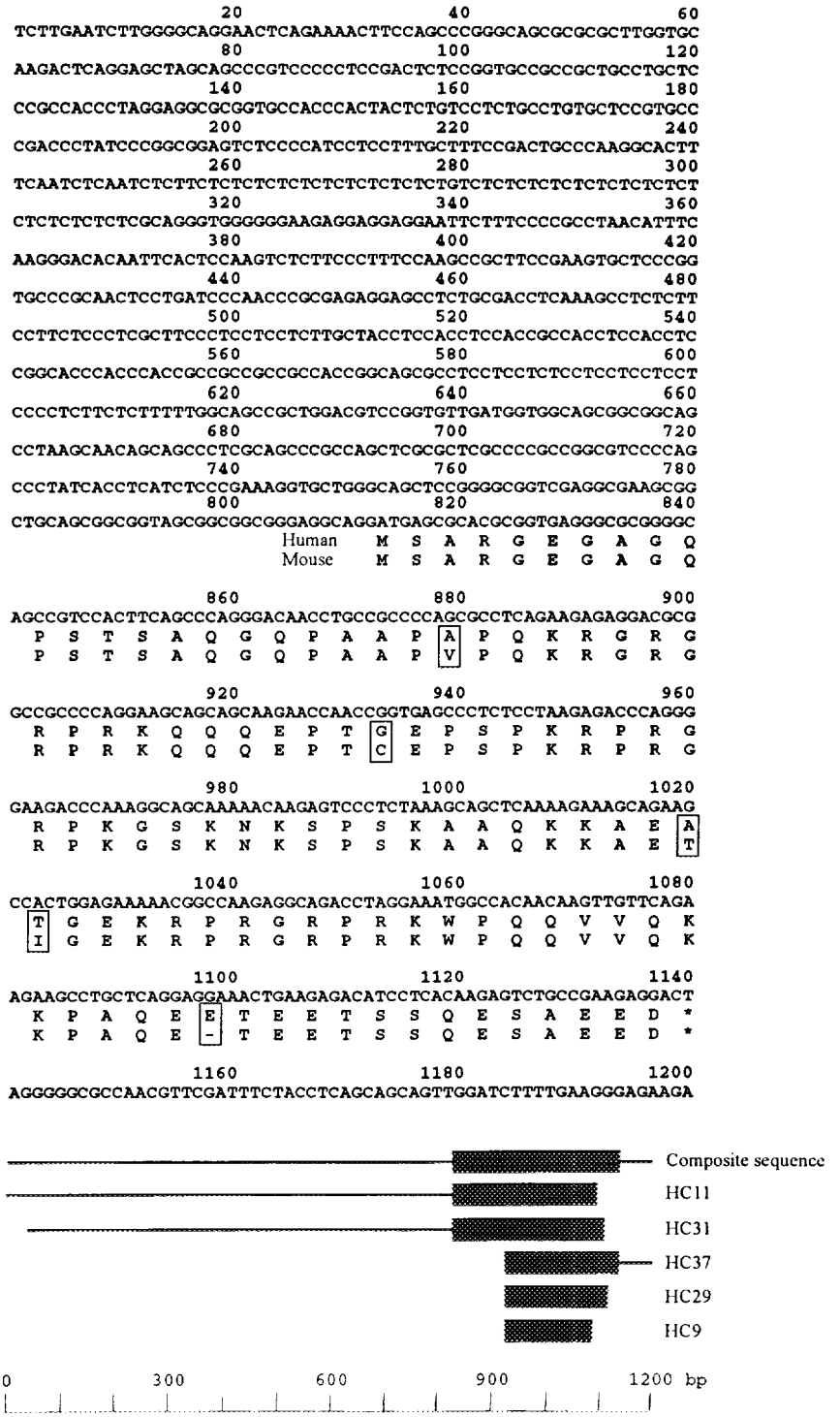

Figure 2. The composite nucleotide and deduced amino acid sequence of the human HMGI–C cDNA. Comparison between the human HMGI–C and the mouse HMGI–C amino acid sequences is given with differences shown in boxes. Shown below the sequences is an alignment of the human HMGI–C cDNA clones analysed. Filled segments represent coding regions.

The composite sequence is shown in Fig. 2. The translational start assumed corresponds to that determined for mouse HMGI–C. The next ATG is at 171 nucleotides upstream (and therefore in frame), but there is a TAA stop codon 7 triplets downstream of this ATG. The cDNA sequence obtained contains an unusually long 5'–untranslated segment, a striking feature of which is a continuous tract of $(CT)_{28}$ (starting at C252), interrupted by a single GT (at 279) and omission of a single C at 256. This sequence could adopt an unusual structure at either the DNA or RNA level and be responsible for some aspect of transcriptional or translational control. Since only a total of 1.2 kb has been sequenced out of an mRNA of 3.8–4.0 kb (Northern analysis, not shown), an extended 3'–untranslated segment of about 2.8 kb is also present. The 22 nucleotides located just upstream of the translational start are identical in human and mouse I–C, and in the 58 nucleotides 3'–of the stop codon there are only 4 differences between the two mRNAs.

Translation of the coding region shows, in comparison to mouse I–C, 4 changes of amino acid, 22: V to A, 40: C to G, 69: T to A, 70: I to T, and the addition of an extra glutamic acid residue at position 95. In addition to these alterations, there are 18 base changes that result in no change of amino acid. Comparison with the mouse sequence shows that the principal characteristics of HMGI–C that distinguish it from I/Y are maintained in the human protein, in particular the segment WPQQVVQKKPAQ and the specific N–terminal domain of 24 amino acids before the first basic segment. The cysteine residue (C40) characteristic of mouse I–C, is however replaced by glycine in human I–C and this accords with the absence of I–C dimers seen in polyacrylamide gels of the human I family proteins (18).

The sequence obtained for human HMGI–C completes the 3 members of the human HMGI family and Fig. 3 shows them aligned for maximum homology. Human HMGI–C is a phosphoprotein but the sites of phosphorylation have not yet been determined, nor have those in mouse I–C. However, for mouse HMGI/Y, *in vivo* phosphorylation of serines 98, 101 and 102 has been established (17). *In vitro* these serines are substrates for casein kinase II (15,16,20). Both human and mouse I–C contain serine at sites

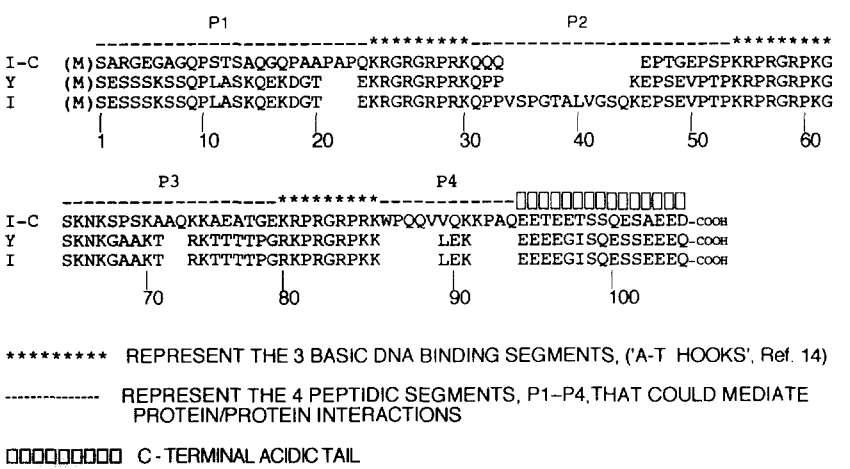

Figure 3. Human HMGI–C amino acid sequence aligned with human HMGI/Y (24) for maximum homology. The numbering is for HMGI and does not include the initial methionine which is known to be absent in all 3 mouse proteins.

corresponding to S98 and S102 of human HMGI, but S101 is replaced by glutamic acid in both mouse and human I–C and an additional serine in I–C replaces Q99. Phosphorylation due to p34/cdc2 kinase has been reported for T52 and T77 of HMGI (21–23). The potential p34/cdc2 site TPKR (residues 52–55) in I/Y is retained as SPKR in both human and mouse I–C but the site TPGR (residues 77–80) in I/Y is not present in I–C. However, another potential p34/cdc2 phosphorylation site can be found at S67 which is present in both human and mouse I–C but not in I/Y. The total potential sites of phosphorylation for both casein kinase II and p34/cdc2 are thus conserved through the three proteins, but with differences in the localization along the sequence.

Comparison of the mouse and human I–C sequences with those of I and Y shows that the DNA binding regions are fully conserved. Protein I–C can be found in several mouse, rat and human transformed cells but is not detectable either in human transformed cells of the hemopoietic system (this paper) or in mouse lymphomas (7). Protein I–C might therefore replace I/Y at specific promoters or act together with them at promoters requiring more than one copy of an HMGI family molecule.

ACKNOWLEDGMENTS

The financial support of the Wellcome Trust, the Wessex Cancer Trust and the EC Science Programme (CT91–0619) are gratefully acknowledged by U.A.P. and C.C.R. K.–Y.C. acknowledges a Studentship from Hong Kong Polytechnic. The research was also funded by grants to V.G.and G.M. from the Associazione Italiana per la Ricerca sul Cancro, the Consiglio Nazionale delle Ricerche and the Ministero dell'Università e della Ricerca Scientifica e Tecnologica (Italy). We are very grateful to Dr. Graham Goodwin (Institute of Cancer Research, London) for many helpful discussions.

REFERENCES

1. Goodwin, G.H., Sanders, C. & Johns, E.W. (1973) Eur. J. Biochem. **38**, 14–19.
2. Johnson, K.R., Lehn, D.A., Elton, T.S., Barr, P.J. & Reeves, R. (1988) J. Biol. Chem. **263**, 18338–18342.
3. Manfioletti, G., Giancotti, V., Bandiera, A., Buratti, E., Sautiere, P., Cary, P., Crane–Robinson, C., Coles, B. & Goodwin, G.H. (1991) Nucleic Acids Res. **90**, 6793–6797.
4. Giancotti, B.V., Berlingieri, M.T., DiFiore, P.P., Fusco, A., Vecchio, G. & Crane–Robinson, C. (1985) Cancer Res. **45**, 6051–6057.
5. Giancotti, V., Pani, B., D'Andrea, P., Berlingieri, M.T., Di Fiore, P.P., Fusco, A., Vecchio, G., Philp, R., Crane–Robinson, C., Nicolas, R.H., Wright, C.A. & Goodwin, G.H. (1987) EMBO J. **6**, 1981–1987.
6. Ram, T.G., Reeves, R. & Hosick, H.L. (1993) Cancer Res. **53**, 2655–2660.
7. Giancotti, V., Bandiera, A., Ciani, L., Santoro, D., Crane–Robinson, C., Goodwin, G.H., Boiocchi, M., Dolcetti, R. & Casetta, B. (1993) Eur. J. Biochem. **213**, 825–832.
8. Lanahan, A., Williams, J.B., Sanders, L.K. & Nathans, D. (1992) Mol. Cell. Biol. **12**, 3919–3929.
9. Fashena, S.J., Reeves, R. & Ruddle, N.H. (1992) Mol. Cell. Biol. **12**, 894–903.
10. Shalnik, D.G. & Neufeld, E.J. (1992) Biochem. Biophys. Res. Commun. **187**, 563–569.
11. Chuvpilo, S., Schomberg, C., Gerwig, R., Heinfling, A., Reeves, R., Grummt, F. & Serfling, E. (1993) Nucleic Acids Res. **21**, 5694–5704.
12. Thanos, D. & Maniatis, T. (1992) Cell **71**, 777–789.
13. Du, W., Thanos, D. & Maniatis, T. (1993) Cell **74**, 887–898.
14. Reeves, R. & Nissen, M.S. (1990) J. Biol. Chem. **265**, 8573–8582.

15. Lund, T., Skalhegg, B.S., Holtlund, J. Blomhoff, K.K. & Laland, S.G. (1987) Eur. J. Biochem. **166**, 21–26.
16. Palvimo, J. & Linnala–Kankkunen, A. (1989) FEBS Lett. **257**, 101–104.
17. Ferranti, P., Malorni, A., Marino, G., Pucci P., Goodwin, G. H., Manfioletti, G. & Giancotti, V. (1992) J. Biol. Chem. **267**, 22486–22489.
18. Giancotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B. & Goodwin, G.H. (1991) Eur. J. Biochem. **198**, 211–216.
19. Giancotti, V. & Goodwin, G.H. (1986) J. Biochem. Biophys. Methods **12**, 265–269.
20. Lund, T., Dahl, K.H., Mork, E., Holtlund, J. & Laland, S.G. (1987) Biochem. Biophys. Res. Commun. **146**, 725–730.
21. Lund, T. & Laland, S.G. (1990) Biochem. Biophys. Res. Commun. **171**, 342–347.
22. Nissen, M.S., Langan, T.A. & Reeves, R. (1991) J. Biol. Chem. **266**, 19945–19952.
23. Reeves, R. Langan, T.A. & Nissen, M.S. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 1671–1675.
24. Eckner, R. & Birnstiel, M.L. (1989) Nucleic Acids Res. **17**, 5947–5959.