RESULTS
(-716 to -314) region upregulates human apo(a) gene.

Transcriptional regulation in eukaryotes is mediated by sequence specific DNA-protein interactions between regulatory trans-acting factors and the 5' and 3' flanking regions of the gene. To study the role of 5' flanking region in transcriptional regulation of human apo(a) gene, constructs were made by cloning (−313 to +13), (−716 to +13) and (−1432 to +13) regions in pGL2-Basic vector. These constructs were transfected in HepG2 and HeLa cells (as negative control) and subjected to luciferase reporter gene assay. It was observed that (−313 to +13) region had basal promoter activity whereas (−716 to +13) had high activity and (−1432 to +13) had near basal activity in HepG2 cells whereas basal activity was observed with all the three constructs in HeLa cells. The results of reporter gene assay of transfections indicated involvement of both the regions, (−1432 to -717) and (−716 to -314) in transcriptional regulation. To begin with, it was decided to study the two regions separately.

Six imperfect dyad symmetry elements found in the positive regulatory region (-716 to -314).

In several eukaryotic genes, elements with sequence of imperfect dyad symmetry have been shown to bind transcriptional regulatory proteins and cause transcriptional regulation (Greenberg, 1987; Gacy and McMurray, 1994; Spiro and McMurray, 1997; Kim, 1998). The nearly palindromic regions can switch between cruciform and linear duplex structures. Formation of the cruciform creates an alternative binding site which may have different affinity for the binding protein. This mechanism also contributes to regulation (Spiro and McMurray, 1997). Considering involvement of imperfect dyad symmetry elements in transcriptional regulation, the sequence of (−716 to −314) region of human apo(a) was analysed on GCG_71 package for dyad symmetry sequences using Stemloop programme. The Stemloop programme finds stems (inverted repeats) within a sequence. The minimum stem length, minimum and maximum loop sizes, and the minimum number of bonds per stem can be specified for desired result. The output of the programme was displayed as all loops or only the best loops saved into a file. This programme was found to be adept at analysing DNA sequence as well as RNA sequence and also considers G-T as base
pair. Using different specifications of the above mentioned four parameters, six imperfect dyad symmetry elements (DSE), DSE1 (-702 to -689), DSE2 (-667 to -658), DSE3 (-650 to -641), DSE4 (-640 to -619), DSE5 (-698 to -681) and DSE6 (-690 to -678) were considered for further analysis. The sequences and the specifications are shown in the figure (fig. 3).

A HeLa specific protein binds to a synthetic DNA fragment (-643 to -616) containing DSE4.

The imperfect dyad symmetry of 22 bp (-640 to -619) (DSE3) with 7 Watson-Crick (WC) matches, 1 non-WC match and 3 mismatches was opted for DNA-protein interaction studies. Sense and anti-sense sequence and three bases each on the flanking sides of this region were used for commercial synthesis of two oligonucleotides. The synthetic oligonucleotides were end-labelled with $^{32}$P $\gamma$ATP using polynucleotide kinase enzyme and then annealed. The annealed fragment was used as probe and subjected to EMSA with HepG2 and HeLa (as negative control) nuclear extracts in presence of poly(dl-dC)·poly(dl-dC). A retarded band was observed with HeLa nuclear extract that faded away when 30-fold molar excess of unlabeled probe was used. No band was detected in the binding reaction with HepG2 nuclear extract (Singh, S. K., data unpublished). This result indicated the involvement of a HeLa specific protein factor, namely aTR, binding to the DSE4 region. Absence of this factor in HepG2 nuclear extract indicated aTR to be a tissue specific repressor as apo(a) promoter is expressed in HepG2 cells but not in HeLa cells.

Characterisation of the cis-element binding aTR.

We used a 219 bp fragment spanning (-716 to -498) region of apo(a) gene that was end labelled on -484 end, for DNase I footprinting and found a 41bp footprint spanning (-656 to -616) region with HeLa nuclear extract. Three protein-induced DNase I hypersensitive sites were observed in the footprint indicating close proximity of the binding factor and the cis-element. The protected sequence was found extended further on distal side when compared to the 28 bp region involved in
Fig. 3 Sequences of sense strand of dyad symmetry elements found by 'stemloop' programme of GCG_71 package. Vertical lines indicating the base pairing (including non-WC bp, G-T). The three numbers on the right side of each sequence indicate size of stem, number of hydrogen bonds between bases in the stem symmetric about the axis of symmetry and loop size respectively.
binding in EMSA. This may be explained by DNase I binding with the minor groove of DNA in which DNA binding surface covers about 10 bp resulting in extended footprint and hence overestimation of the size of DNA-binding site with DNase I footprinting method (Oefner and Suck, 1986). Large size of aTR factor might be enhancing the effect (fig. 4).

A single polypeptide of 107,000 Da size is involved in binding to the (-643 to -616) fragment.

In order to find out the number of polypeptides in the trans-acting factor and their respective molecular mass, South-Western blot assay was performed. The nuclear proteins of HeLa cells were separated on SDS-PAGE and then Western transferred onto nitrocellulose filter. The transferred proteins were denatured using guanidine hydrochloride, renatured by stepwise removal of guanidine hydrochloride and then probed with a multimer of the radiolabelled 28 bp fragment. A band was observed in the autoradiogram corresponding to a molecular mass of 107,000 Da in the HeLa nuclear extract lane but no band was observed in the HepG2 nuclear extract lane. This observation was in agreement with the EMSA result. Thus, we concluded that the binding factor is a single protein (fig. 5).

The (-643 to -616) fragment represses luciferase reporter gene expression under heterologous SV40 promoter.

The apo(a) gene is expressed only in liver cells but the binding factor, aTR was not found in HepG2 cells (a hepatoma cell line) in binding studies. Instead, it was present in HeLa cells. It could be inferred from these results that aTR is a tissue specific repressor. To confirm the function of aTR as a repressor, we performed reporter gene assay. The 28 bp synthetic fragment used as probe in EMSA was subcloned in a luciferase reporter gene vector, pGL2-Promoter in which luciferase gene is expressed under a strong promoter (SV40 promoter) (fig. 6). The correct orientation of the subcloned fragment was checked by sequencing (fig. 7). The effect of aTR was determined on the heterologous SV40 promoter. This construct was used for transient transfection in HeLa cells. A 27.5% reduction in luciferase activity was observed when compared to the control (pGL2-Promoter vector) (fig. 8).
fig. 4 DNase I footprinting assay for characterising cis-element that binds aTR factor in HeLa nuclear extract. First lane is A+G Maxam-Gilbert sequencing, second lane is control ladder without nuclear extract and following lanes are DNase I digested ladders with increasing amount of HeLa nuclear extract. The sequence of the probe is written on the top of control lane. The amount of HeLa nuclear extract used (in µg) in the
reaction lanes is written on the top of respective lanes. The sense strand of the probe was radio-labelled with $^{32}$P $\gamma$ATP using T4 polynucleotide kinase at distal end at position (-716). The probe was generated by PCR amplification using end-labelled sense primer.
Fig. 5 South-Western blot assay of HeLa nuclear extract probed with multimer of synthetic fragment spanning (-643 to -616). The arrow indicates the molecular mass of the protein bound by the probe.
Fig. 6 Schematic representation of plasmid construct used for reporter gene assay to see the effect of (-643 to -616) region on a heterologous promoter (SV40 promoter) activity.
Fig. 7 Sequencing of sub-cloned (-643 to -616) fragment in pGL2-Promoter vector. End labelled anti-sense primer was used.
Fig. 8 Plot showing luciferase reporter gene assay.
Repressor activity of the 28 bp fragment proved the earlier hypothesis we had made on the basis of the results of EMSA.

**HepG2 specific factors bind to the (–716 to –616) fragment.**

–716 to –616 region of apo(a) gene was PCR amplified using forward primer used for amplification of (–716 to +13) fragment and anti-sense oligonucleotide involved in aTR binding as reverse primer. This fragment was subjected to EMSA using HepG2 nuclear extract and HeLa nuclear extract as control. One retarded band was obtained with HeLa nuclear extract. It was in agreement with the earlier result of EMSA with HeLa nuclear extract and (–643 to –616) region. HepG2 nuclear extract exhibited multiple bindings. None of the retarded bands in the HepG2 lane corresponded to the retarded band found in HeLa lane. These bands appeared to be involving at least three small proteins. All the bands in HepG2 lane were exhibiting higher mobility than the single band in HeLa lane (fig. 9). The protein factors are HepG2-specific and this observation is consistent with our initial result in which it was shown that (–716 to –314) region upregulated the expression of reporter gene in HepG2 cells but not in HeLa cells.

~50 bp deletion constructs of (–716 to +13) region, from -716 side were sub-cloned in luciferase reporter gene vector (pGL2-Basic, Promega). These constructs were used for transient transfection in HepG2 cells. In reporter gene assay it was found that there was a 90% drop in luciferase activity with the construct containing (–639 to +13) region and 62% drop with the construct containing (–668 to +13) when compared to the luciferase activity with the construct containing (–716 to +13). The luciferase activity of (–639 to +13) was comparable to the basal activity which was exhibited by (–593 to +13) region (Pati, data unpublished). The reporter gene assay results were in agreement with the result of EMSA. The functional role of the binding factors as activators was also confirmed.
Fig. 9 EMSA of (-716 to -616) region with HeLa and HepG2 nuclear extracts. First lane is binding reaction without nuclear extract (control), second and fourth lane show binding reactions with HeLa and HepG2 nuclear extract respectively.
Separation of binding factors on SDS-PAGE renders inability to bind to the
(-716 to -616) region.

To determine the number of polypeptides involved in binding and their respective
molecular masses, South-Western blot assay was performed. HepG2 nuclear extract
proteins along with HeLa nuclear extract proteins as a control, were separated on
SDS-PAGE, western transferred onto nitrocellulose membrane, denatured with 6 M
guanidine hydrochloride, renatured by stepwise removal of guanidine hydrochloride
and probed with (-716 to -616) labelled fragment. One band was observed in HeLa
lane but no band appeared in HepG2 lane (fig. 10). It led to the following inference.
Probably more than one protein is involved in all the complexes formed with (-716 to
-616) probe and HepG2 nuclear extract and complex formation is co-operative in
nature. No protein was capable of binding to the probe independently. It implied that
at least two proteins were required for the formation of all the complexes observed in
the EMSA. Another explanation for not being able to detect any band in HepG2
might be sensitivity of the involved proteins to the harsh treatments given to them
such as boiling in the presence of strong detergent SDS, and reducing agent β-
mercaptoethanol and later denaturation in presence of 6 M guanidine hydrochloride.

Characterisation of cis-element binding to the HepG2 specific trans-acting
factors.

The further characterisation of cis-elements was performed by DNase I footprinting
of the region involved in DNA-protein complex formation. On the basis of the results
obtained from EMSA and luciferase reporter gene assay of 5' deletion constructs of
(-716 to +13) region, we chose a 233 bp fragment (-716 to -484) for DNase I
footprinting assay. A single 95 bp long footprint stretching from (-703 to -609) was
observed. This was another unsuccessful attempt to know more about the nature of
complexes. However, the limits were defined for the cis-elements involved (fig. 11).

To check the limit at the other end, another 292 bp long probe spanning (-775 to
-484) was generated and subjected to DNase I footprinting. This probe also exhibited
a long continuous footprint limiting to -610 at the proximal end but stretched to -717
**Fig. 10** South-Western blot assay with HepG2 and HeLa nuclear extract using (-716 to -616) as probe.
Fig. 11 DNase I footprinting of (-716 to -616) region using HepG2 nuclear extract. The probe was labeled from proximal end. In both the figures first lane was containing Maxam and Gilbert A+G reaction. Second lane was control and subsequent lanes were increasing amounts of HepG2 nuclear extract. The amounts (in µg) of the nuclear extract is written on the top of the lane. The (-716 to -484) probe was generated by PCR amplification. The anti-sense primer
was radio-labelled with $^{32}$P-ATP using T4 polynucleotide kinase to end-label the probe at position (-484). The probe (-706 to -616) was also generated in a similar manner. The anti-sense primer was radiolabelled with $^{32}$P-ATP using T4 polynucleotide kinase to end-label the probe at position (-616).
at the distal end. In this footprint, two DNase I hypersensitive sites were also observed (fig. 12).

In another attempt, b (-706 to -616) was used as probe and a 57 bp long partial footprint spanning (-647 to -703) was obtained (fig. 11).

More than one DSE is required in DNA probe for the formation of DNA-protein complexes with HepG2 specific factors.

The (-716 to -616) region was dissected out into smaller fragments to get more information on complexes observed in EMSA with HepG2 nuclear extract. On the basis of success with DSE4 in binding with a HeLa specific factor, a fragment containing DSE5 and DSE6 and fragments containing DSE1, DSE2 and DSE3 (g) were synthesised and used for EMSA. None of the fragments exhibited any binding with HepG2 nuclear extract. Then five different fragments were generated by PCR amplification using combinations of DSE1-4 oligonucleotides as primers. These five fragments were a (-706 to -652)(55 bp), b (-706 to -616)(91 bp), c (-668 to -616)(53 bp), e (-668 to -636)(33 bp) and f (-662 to -616)(47 bp) (fig. 13). EMSA with the b fragment using HepG2 nuclear extract showed four retarded bands with a pattern similar to the pattern of four retarded bands observed with (-716 to -616) fragment, whereas one retarded band was observed with HeLa nuclear extract. The retarded band in HeLa lane was distinct from any of the four retarded bands observed in HepG2 lane. This retarded band was result of complex of aTR with b that contains DSE4 at the proximal end. This showed that (-716 to -707) region is not playing any important role in complex formation with trans-acting factors. The trans-acting factors binding to the activator region are absent in HeLa nuclear extract, and therefore appear to be tissue specific activators. The c fragment exhibited a pattern of three retarded bands, comparable to the band pattern obtained with b, the only difference was absence of the most retarded band. We conclude that one factor is responsible for the formation of the complex with least mobility along with one of the three other complexes. The c fragment was digested with Apol to give (-668 to -646) d1 and (-641 to -616) d2 fragments, which were together subjected to EMSA. No retarded band was observed with HeLa or HepG2 nuclear extract which, implied that neither of the two digested fragments could bind with any factors (fig. 14). This digestion had
Fig. 12 DNase I footprinting analysis of (-775 to -484) region labelled at distal end using HepG2 nuclear extract. First three lanes are controls without nuclear extract. Next three lanes contained increasing amounts of HepG2 nuclear extract. The amount of nuclear extract (in μg) is written on the top of lanes. Last lane is Maxam and Gilbert sequencing for A+G. The probe was generated by restriction digestion of PCR amplified fragment.
(-995 to -484) by Sau3AI at position (-775). The (-775 to -484) fragment was purified and radio-labelled by filling the site of digestion with $^{32}$P αdCTP in dNTP mixture using DNA polymerase I large (Klenow) fragment.
Fig. 13 Pictorial representation of PCR amplified and synthesised DNA fragments along with their sense sequences. Sequences written with black background highlight the six imperfect dyad symmetries. Length and position of the sequences is written on the right side of each fragment. Underlined sequence written in italics are overhangs.
Fig. 14 EMSA of smaller fragment of (-716 to -616) region, b, c and Apol digested c. First lane of each probe is control (without any nuclear extract) followed by binding reaction with HepG2 and HeLa nuclear extracts. Third lane of b lane consists of binding reaction competed with 30 fold unlabelled probe. The bands above free probe in Apol digested c lanes are of residual undigested c fragment.
resulted in demolishing DSE3 and separated DSE2 from DSE4. This result further strengthened the hypothesis of co-operative binding and significance of DSE3 in complex formation.

The fragment a shares a 17 bp region containing DSE2 with fragment c and includes the region missing in probe c. When a was used as probe in EMSA with HepG2 nuclear extract, two retarded bands were observed (fig. 15). The pattern of the retarded bands was distinct from the pattern of retarded bands with b or c. Since probe a is part of probe b which has been subjected to South-Western blot assay, the previously proposed property of co-operative nature of the complexes may be extended to the complexes involved in retarded bands with a.

When e and f fragments were subjected to EMSA using HepG2 nuclear extract, both the probes showed two retarded bands each. The pattern of retarded bands was similar. Interestingly, g fragment failed to give any retarded band but e and f fragments which were extended on either side of g by a few base-pairs showed positive EMSA (fig. 16).

**Cross-linking of DNA-protein complexes by UV irradiation**

Since no information about the number of size of the trans-factors binding to form DNA-protein complexes in EMSA could be obtained from South-Western blot assay, an alternative approach was used. UV cross-linking assay was performed to determine the size of the DNA-protein complex. Binding reactions with a and c were modified by scaling it up five fold. Body-labelled probes were used instead of end-labelled ones for the UV cross-linking assay. After the incubation period was over, the binding reaction was subjected to UV irradiation in stratalinker (Stratagene). The reaction mixture was analyzed by SDS-PAGE, and exposed for autoradiography. Two bands of 97,000 Da and 112,000 Da were seen with a and two bands of 73,000 Da and 103,000 Da were observed with c in the autoradiogram. It implied that two proteins of 15,000 and 30,000 Da are involved in binding. One half of the reaction with c was subjected to DNasel digestion after UV irradiation, to split the complex into two or more fractions in order to get more information about the number and size of the proteins involved, but the band pattern did not show any difference when
**Fig. 15** EMSA of probe a with HepG2 nuclear extract. First lane is control and second lane is binding reaction with HepG2 nuclear extract.
**Fig. 16** EMSA of fragments g, e and f with HepG2 nuclear extract. First lane of all the three probes represents control reaction and second lane consists of respective binding reactions.
compared to the undigested lane (fig. 17). It implied that the complex was so tight that DNase I enzyme could not reach DNA between the bound proteins. This observation is in agreement with the results of DNase I footprinting that showed continuous long footprints and protein induced DNase I hypersensitive sites indicating protein-protein and protein-DNA compactness respectively, in the DNA-protein complex.

The fragment 'b' (-706 to -616) forms more stable DNA-protein complexes than the fragments 'a' (-706 to -652) and 'c' (-668 to -616), with HepG2-specific trans-acting factors

Identification of common factors involved in binding with the three different fragments, a, b and c was attempted by competing complexes with cold probes having some common region. Entire sequences of a and c are shared by b, whereas a and c mutually share a 17 bp region. b was competed with cold a and c but without any effect on binding. When c was competed with a, all the binding was removed but when a was competed with cold c, binding was slightly affected. This implied that b formed the most stable complex which was not affected by competition with either a or c, whereas the complex with a was more stable than complex with c as binding with a was hardly affected by cold c but on the contrary the complex with a is completely removed by cold c (fig 18).

Purification of the trans-acting factor(s) using biotinylated cis-element and streptavidin magnetic beads.

Biotinylated probe was generated by PCR amplified fragment using Biotinylated sense primer. Binding reaction was scaled up and after completion to incubation period it was mixed with streptavidin-magnetic beads for biotinylated DNA-streptavidin magnetic bead coupling. Then supernatant was removed after pelleting down streptavidin-biotinylated DNA-protein complex using a strong magnet. The pelleted complex was washed with low salt buffers and protein was released from the complex using increasing concentrations of NaCl. The fractions were used for EMSA and also analysed on SDS-PAGE. SDS-PAGE was silver-stained. Neither any binding EMSA nor any silver-stained bands in the SDS-PAGE were observed.
Fig. 17 UV cross-linking assay of fragments a and c with HepG2 nuclear extract. HepG2 labelled lanes contain UV cross-linked binding reaction with HepG2 nuclear extract whereas control lanes consisted of binding reaction without nuclear extract. First two lanes contain DNase I digested reactions.
Fig. 18 Cold competition of binding reaction of probes a, b and c with HepG2 nuclear extract. Binding reaction of b was competed with unlabelled a, b and c, whereas binding reactions of a and c were competed with unlabelled a and c separately. The unlabelled competitor probes were added in 50 fold molar excess.
Purification of the UV cross-linked DNA-protein complex using biotinylated fragment 'a', HepG2 nuclear extract and streptavidin magnetic beads.

After failing to purify the protein by affinity purification method, purification of UV cross-linked DNA-protein complex was attempted. The purified DNA-protein complex can be used to determine partial protein sequence by mass spectroscopic analysis that can be used to screen cDNA library to clone the gene encoding the protein.

Binding reaction was scaled up thirty fold and biotinylated probe a was used. The complex in binding reaction was exposed to UV radiation as 100 μl aliquots in Stratalinker, after completion of incubation period. The UV cross-linked biotinylated DNA-protein complex was purified using streptavidin magnetic beads and a strong magnet. The purified complex was released from strong binding of Streptavidin and Biotin by boiling the complex in SDS loading dye containing 8 M urea for fifteen minutes. The supernatant was loaded onto a SDS-PAGE and separated. The gel was silver-stained after the run was over. A band of 43,500 Da was observed in the reaction lane. There was no other band of comparable intensity in the lane. The control lane was free of any bands (fig 19). The size of the purified UV cross-linked DNA-protein complex was smaller than the sizes found in UV cross-linking assay owing to partial degradation of the nuclear extract.

To confirm that the band was not of DNA probe, the gel was dried and exposed and two bands of free probe, distinct from the above mentioned bands were observed in the autoradiogram. The purified UV cross-linked DNA-protein complex will be used for partial sequencing of protein involved using mass spectroscopy (Nordhoff, 1999).
**Fig. 19** SDS-PAGE of purified UV cross-linked DNA-protein complex of a and its binding protein(s). First and ninth are molecular mass marker lanes. Third and fourth lanes are nuclear extract and supernatant of UV cross-linked binding complex respectively. Eighth lane contains streptavidin-magnetic bead purified biotinylated UV cross-linked DNA-protein complex. Sixth lane contains control in which purification was performed in absence of HepG2 nuclear extract.