Vol. 205, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1994 Pages 545-551

# Co-occurrence of CArG boxes and TCF sites within viral genomes

Michael A. Cahill, Alfred Nordheim, and Ralf Janknecht[1]

Institute for Molecular Biology, Hannover Medical School,
D-30623 Hannover, Germany

Received October 7, 1994

The transcription factor SRF is involved in the transduction of extracellular signals into nuclear responses, often in conjunction with ternary complex factors (TCFs). Here we report the identification of CArG box SRF binding-sites, and neighboring TCF binding-sites, in viral genomes. SRF binds and recruits TCFs to CMV, RSV and HTLV-1 viral genomes. At least one of two specific CArG boxes occurred in cytomegaloviruses in the 5' proximal region of the major immediate early gene, one always accompanied by a TCF site. This conservation was striking since neither the flanking sequences nor the spacing to the CAP site were conserved. Thus the ubiquitous SRF and TCF molecules may control events in the life cycle of viruses. 

Many cellular immediate early genes contain the CC(A/T)$_6$GG "CArG box" consensus SRF binding site, with an adjacent GGA(A/T) Ets binding site capable of recruiting Ternary Complex Factors (TCFs). The combination of a CArG box with adjacent TCF binding site is referred to as a serum response element (SRE), after the prototypical SRE of the c-*fos* gene (1). The spatial requirements for SRF and TCF binding sites necessary for ternary complex formation are flexible at the DNA level, with ternary complexes formed when the TCF site is up to 25 bp distant from the CArG box (2). However there have been no functional studies to indicate whether all CArG boxes bind SRF, or whether all ternary complexes can activate transcription. Recently a functional SRE has been identified in the mouse *junB* gene 2 kb downstream of the site of transcription initiation (3), indicating that SREs may exert influence over long distances in an enhancer-like manner.

[1]*Correspondence to* R. Janknecht, Institut für Molekularbiologie, Medizinische Hochschule Hannover (OE 5250), D-30623 Hannover, Germany. Fax: +49 - 511 - 532 4283.

**Abbreviations**: CArG box, CC(A/T)$_6$GG; EMSA, electrophoretic mobility shift assay; hCMV, human cytomegalovirus; HSV-1 to HSV-6, herpes simplex virus types 1-6; LTR, long terminal repeat; mCMV, murine cytomegalovirus; MIE, major immediate early; MmCMV, *Macaca mulatta* cytomegalovirus; RSV, Rous sarcoma virus; sCMV, simian cytomegalovirus; SRE, serum response element; SRF, serum response factor; TCF, ternary complex factor.

0006-291X/94 $5.00

Viral replication is preceded by transcription of viral immediate early genes. Some viruses may couple their immediate early induction programs to signal pathways activating the host cell immediate early response governed by SRF, since CArG motifs have been identified in the promoters of the human and simian cytomegalovirus (hCMV and sCMV) major immediate early (MIE) promoters (4) and the long terminal repeats (LTRs) of Rous Sarcoma Virus (RSV) and related avian retroviruses (5-8). By scanning the EMBL data base for all possible CArG boxes we identify here that immediate early loci of cytomegaloviruses and some other members of the herpesviridae contain CArG motifs whose distribution suggests biological importance. We also show for the irst time recruitment of TCFs to viral CArG boxes by SRF.

## Methods

pEQ+RSV was made by replacing the CMV promoter from pEQ176 (9) with the RSV promoter from pRSVneo (10). Electrophoretic mobility shift assays (EMSA) were performed as described (11). Mouse fibroblast NIH3T3 cells were transfected with 2 μg of pEQ176, pmEQ176, pEQ+RSV or pEQ+mRSV, and 4 μg of the luciferase expressing ptk80-luc. Where indicated, 1.5 μg of each BXB and MAP kinase expression plasmids were additionally transfected. Measurement of β-galactosidase and luciferase activities were as described (11). Luciferase activity was used to normalize β-galactosidase activity for transfection efficiency. pET3b-Tax was a gift of Dr. Peter Beimling (Berlin), and contains the HTLV-1 Tax cDNA 5' from the Hinc II site.

## Results and Discussion

### *CArG boxes in viral genomes from the EMBL data base*

The viral data base (CDEM35VO, December 1993) was searched with 10 base pair motifs corresponding to all 64 possible CArG boxes using the PC/GENE program (Intelligenetics, Mountain View, CA). 24 bp on each side of the 375 different viral CArG boxes found, excluding the residues immediately adjacent to the outer C:G base pairs of the CArG box, were then scanned for the presence of GGAA, GGAT, ATCC, or TTCC motifs. The observed frequency of these TCF sites per CArG box was 0.65, approximately double the random expected frequency. However this was CArG box independent since a random search using control sequences instead of CArG boxes gave a corresponding frequency of 0.73.

### *CArG boxes within herpesviruses*

The herpesvirus superfamily consists of herpesviruses, CMVs, Varicella-Zoster Virus, and Epstein-Barr Virus (12). The MIE promoter of sCMV contains three CArG boxes, at -702, -575, and -517 with respect to the mRNA CAP site (4; Figure 1A). The latter two sites each have identical, perfectly dyad symmetrical CArG boxes with an adjacent consensus TCF site three bp from the CArG box (Figure 1B). This CArG sequence was identified as a high affinity SRF binding site by random site selection (13). At least one of these two specific CArG motifs were found in the proximal MIE promoter of all CMV viruses: CCATATATGG motifs with an adjacent TCF site in sCMV (-575, -517), MmCMV (-451) and hCMV (-501), and CCATTATTGG (or CCAATAATGG minus strand) without a TCF site in sCMV (-702) and mCMV (-300) (Figure

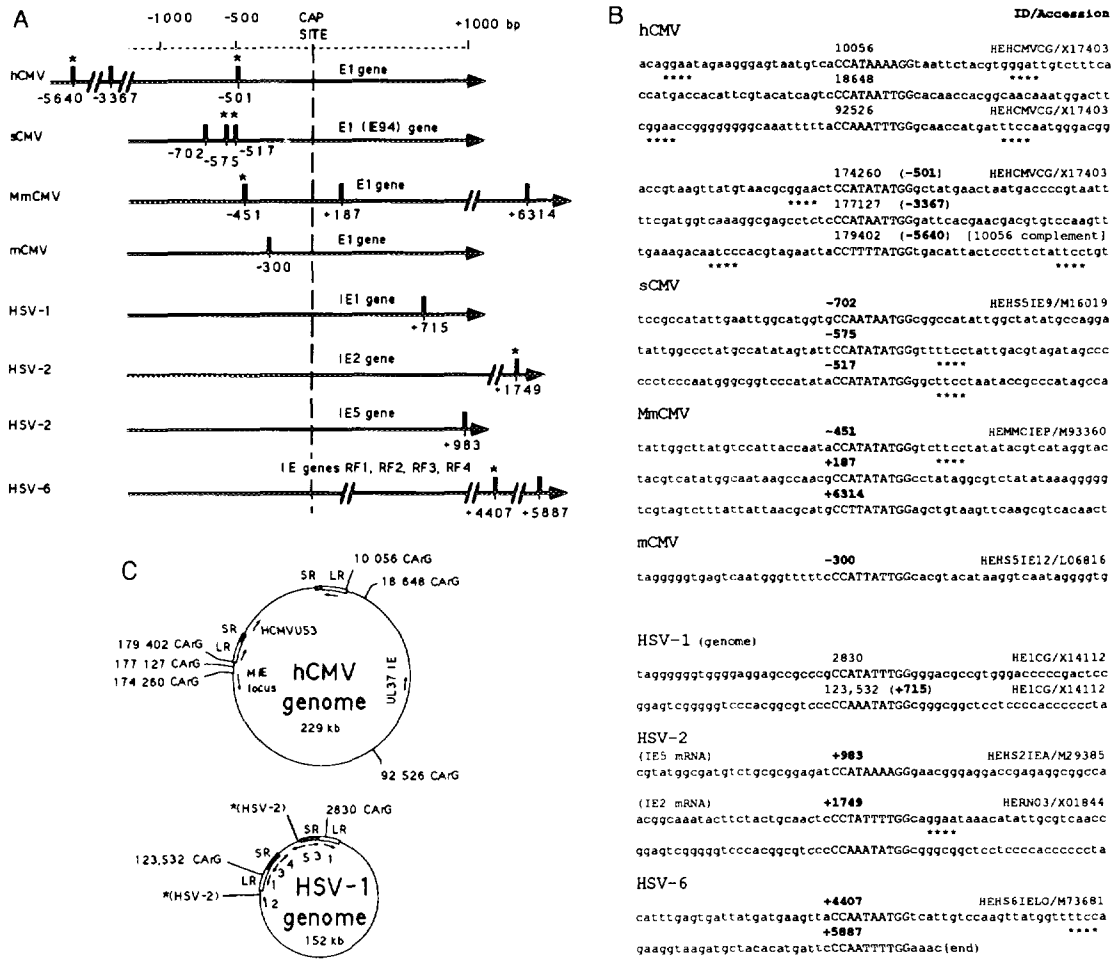


**Figure 1.** *Colocalization of CArG boxes and immediate early genes.*
A. The positions of CArG boxes with respect to the mRNA start site (CAP site) of the indicated viral immediate early genes are shown schematically. "E1 gene" refers to the homologue of the hCMV MIE gene. CArG boxes with neighboring TCF sites (within 25 bp) are asterisked.
B. CArG boxes from "A" are shown with 25 flanking nucleotides. Consensus TCF sites are asterisked. Identification (ID) and accession numbers are shown. Boldface numbers refer to the position relative to the CAP site from "A", and other numbers to those from the data base.
C. The linear genomes of hCMV and HSV-1 are depicted schematically in the circularized state. The short repeat (SR; filled boxes) and long repeat (LR; open boxes) of both genomes are indicated. Arrows indicate the sites of immediate early transcription. hCMV immediate early loci follow (12), and HSV-1 and HSV-2 loci follow (18). The position of HSV-2 CArG boxes, at homologous sites within the HSV-1 genome, are indicated by asterisked brackets *(HSV-2).

1A,B). Notably, neither the position of the CArG box relative to the CAP site nor the flanking sequences are conserved, which may contribute to differences in the biology of these viruses. Furthermore, 5 out of 6 CArG boxes present in the hCMV circularized genome are within 10 kb of immediate early loci, indicating clustering of CArG boxes around these loci (Figure 1C).

Herpes simplex virus type 6 (HSV-6), which is closely related to hCMV, has two CArG boxes within the immediate early locus encoding the RF1-RF4 proteins, at +4407 and +5887 (Figure 1A,B). However it does not have a CArG motif in the 5' region of the MIE gene (14). The published HSV-1 genome has only one CArG motif in the long repeat, which is present twice in the circularized genome in a position where it could provide enhancer functions for immediate early loci 1 through 5 (Figure 1C). HSV-2 shares about 50% sequence similarity with HSV-1, and very similar genome organization. There are also two CArG boxes in the HSV-2 genome (bracketed at their homologous positions in Figure 1C) where they could also perform enhancer-like functions for the immediate early genes. This preponderance of CArG motifs in the vicinity of immediate early genes suggests that certain herpesviruses may utilize SRF in the transcriptional regulation of the viral immediate early response.

***hCMV, RSV and HTLV-1 CArG boxes can recruit TCFs***

We examined by EMSA whether the TCF Elk-1 could be recruited to the good consensus TCF site at -501 in the hCMV promoter. Figure 2A shows that this is the case. Elk-1 could not bind to the hCMV promoter alone but was recruited by SRF, similarly to the ternary complex formed on the c-*fos* promoter. The binding of both transcription factors was specific, as indicated by competition studies (lanes 11-19).

Figure 2B shows that SRF which was bound to the CArG boxes in the RSV promoter (5-8) or the open reading frame of the HTLV-I Tax gene (15) could also recruit Elk-1. These sites in the reading frame of HTLV-1 Tax have not been previously reported. An internal HTLV-1 promoter between the *pol* and *env* genes can direct transcription of Tax mRNA (16). The Tax protein transactivates transcription via protein interactions with SRF (17), suggesting a role where Tax may regulate its own transcription through this CArG box.

***Transcriptional activity of the hCMV and RSV CArG boxes***

Chang et al. (4) observed that a segment of the sCMV MIE promoter which contained SRE-like sequences was not serum inducible, yet served as an efficient enhancer. To assess the role SRF may play in the enhancer function of the hCMV MIE promoter, we mutated the CArG box and assayed the effects using a β-galactosidase reporter gene under hCMV promoter control in transient transfection experiments. Basal levels of transcription were unaffected by the presence or absence of the hCMV SRE (Figure 3A). The promoter was not serum inducible in the presence or absence of the SRE (data not shown). However the hCMV promoter could be stimulated by activated MAP kinase in an SRE-independent manner. To our knowledge this is the first example where the MAP kinase pathway has been demonstrated to activate viral transcription.

Next we assayed the single CArG box present in the RSV promoter of the plasmid pEQ+RSV. Consistent with the results of Habel et al. (7), destroying the RSV-LTR CArG sequence (pEQ+mRSV) lowered basal activity of the RSV promoter, and similarly to the results of Lang et al. (8), the distal RSV promoter employed in this study was only marginally inducible by activated MAP kinase or serum in a CArG box-independent fashion (Figure 3B; data not shown).

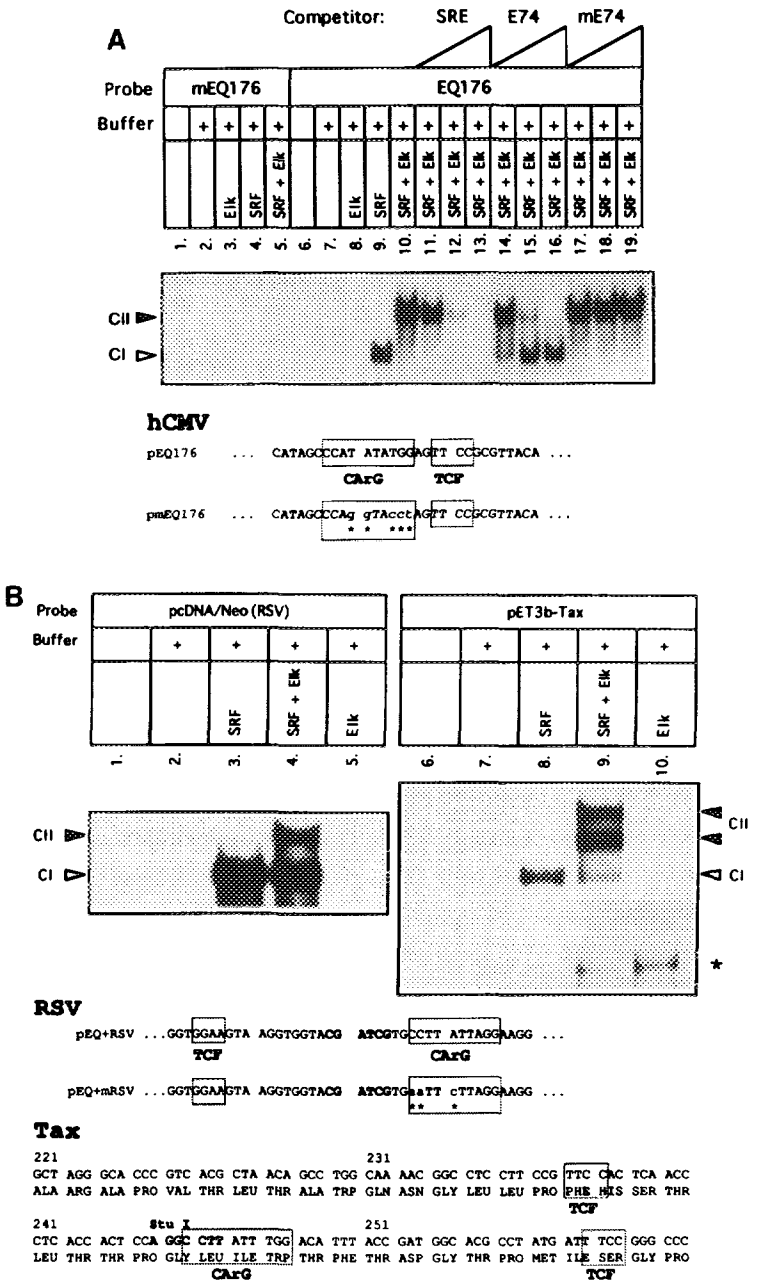


**Figure 2.** *Ternary complex formation on viral genomes.*
A. The Spe I/Nde I fragments of pEQ176 or pmEQ176 were used as EMSA probes. CI, SRF; CII, SRF plus Elk-1. Unlabeled competitor oligonucleotides according to (11) were added at 1 fold, 10 fold, or 100 fold excess over $^{32}$P-labeled probe. These contained the *c-fos* SRE which binds both SRF and Elk-1, the *Drosophila* E74 site which binds Elk-1 but not SRF, or a mutated E74 site which does not bind Elk-1 or SRF. The region of pEQ176 containing the SRF (CArG) and Elk-1 (TCF) sites is depicted underneath, along with the mutated sequence of pmEQ176.
B. EMSA using the Bsu36 I/Nhe I fragment from pcDNA/Neo (Invitrogen) containing the RSV promoter, or the Hinf I/Sca I fragment from pET3b-Tax. Partial nucleotide and/or amino acid sequences of the RSV LTR and open reading frame (ORF) of the HTLV-1 tax gene are depicted below the panel, as is the mutated sequence of pEQ+mRSV. CArG boxes and TCF sites are boxed. Numbering for Tax represents amino acids. Direct binding of Elk-1 is asterisked, and was not observed for the hCMV or RSV probes (not shown).

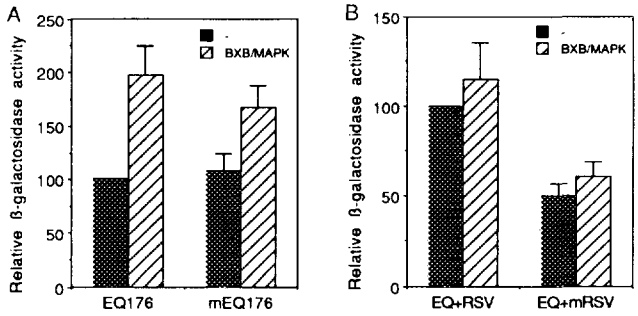


**Figure 3.** *Transactivation via viral SREs.*
A. Transcriptional activity of a β-galactosidase gene under the control of an hCMV promoter (EQ176) or an hCMV with mutated SRE-like motif (mEQ176, see Figure 2A). β-galactosidase activity was determined in the absence (black bars) or in the presence of activated MAP kinase (striped bars). *In vivo* activation of MAP kinase was accomplished using a constitutively active form of Raf kinase (BXB) as detailed (11).
B. Analogous with RSV promoter (pEQ+RSV) or its mutated form (pEQ+mRSV; see Figure 2B).

### *Perspectives*

This study reveals SRF/TCF binding to viral genomes, and a pattern of occurrence of CArG boxes within some Herpesviruses, especially within the cytomegaloviruses where they tend to be clustered within several kb of immediate early genes. We and others (4) have been unable to detect a function for SRE-like sequences in CMV promoters, yet their evolutionary conservation suggests strongly they perform some role. Identification of more CArG boxes within hundreds of viral genomes (data not shown) suggests that the ubiquitous transcription factor SRF is important in the transcriptional regulation of many viruses.

## Acknowledgments

We are very grateful for the excellent technical support of Uschi Wiedemann, for discussions with Wolfram Ernst, and for the financial support of the *Deutsche Forschungsgemeinschaft* (Grant No 120/7-2).

## References

1. Treisman, R. (1994) *Curr. Opin. Genet. Dev.* **4,** 96-101
2. Treisman, R., Marais, R. and Wynne, J. (1992) *EMBO J.* **11,** 4631-4640
3. Perez-Albuerne, E.D., Schatteman, G., Sanders, L.K. and Nathans, D. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11960-11964
4. Chang, Y.-N., Jeang, K.-T., Chiou, C.-J., Chan, Y.-J., Pizzorno, M. and Hayward, S. (1993) *J. Virol.* **67,** 516-529
5. Boulden, A. and Sealy, L. (1990) *Virol.* **174,** 204-216
6. Zachow, K. and Conklin, K.F. (1992) *J. Virol.* **66,** 1959-1970

7. Habel, D.E., Dohrer, K.L. and Conklin, K.F. (1993) *J. Virol.* **67,** 1545-1554
8. Lang, A., Fincham, V.J. and Wyke, J.A. (1993) *Virol.* **196,** 564-575
9. Firzlaff, J.M., Lüscher, B. and Eisenman, R.N. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5187-5191
10. Gorman, C., Padmanabhan, R., and Howard, B.H. (1983) *Science* **221,** 551-553
11. Janknecht, R., Ernst, W.H., Pingoud, V. and Nordheim, A. (1993) *EMBO J.* **12,** 5097-5104
12. Chee, M.S., Bankier, A.T., Beck, S., Bohni, R., Brown, C.M., Cerny, R., Horsnell, R., Hutchison, C.A., Kouzarides, T., Martignetti, J.A., Preddie, E., Satchwell, S.C., Tomlinson, P., Weston, K.M. and Barrell, B.G. (1990) *Curr. Topics Microbiol. Immunol.* **154,** 125-169
13. Pollock, R. and Treisman, R. (1990) *Nucl. Acids Res.* **18,** 6197-6204
14. Thomson, B.J. and Honess, R.W. (1992) *J. Gen. Virol.* **73,** 1649-1660
15. Seiki, M., Hattori, S., Hirayama, Y. and Yoshida, M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3618-3622
16. Nosaka, T., Ariumi, Y., Sakurai, M., Takeuchi, R. and Hatanaka, M. (1993) *Nucl. Acids Res.* **21,** 5124-5129
17. Fujii, M., Tsuchiya, H., Chuhjo, T., Akizawa, T. and Seiki, M. (1992) *Genes Dev.* **6,** 2066-2076
18. Whitton, J.L., Rixon, F.J., Easton, A.J. and Clements, J.B. (1983) *Nucl. Acids Res.* **11,** 6271-6287