Maternal-specific footprints at putative CTCF sites in the *H19* imprinting control region give evidence for insulator function

Piroska E. Szabó, Shih-Huey E. Tang, Altan Rentsendorj, Gerd P. Pfeifer and Jeffrey R. Mann

Parent-of-origin-specific expression of the mouse insulin-like growth factor 2 (*Igf2*) gene and the closely linked *H19* gene are regulated by an intervening 2 kb imprinting control region (ICR), which displays parent-specific differential DNA methylation [1,2]. Four 21 bp repeats are embedded within the ICR and are conserved in the putative ICR of human and rat *Igf2* and *H19*, suggesting that the repeats have a function [3,4]. Here, we report that prominent DNA footprints were found *in vivo* on the unmethylated maternal ICR at all four 21 bp repeats, demonstrating the presence of protein binding. The methylated paternal ICR displayed no footprints. Significantly, the maternal-specific footprints were localized to putative binding sites for CTCF, a highly conserved zinc-finger DNA-binding protein with multiple roles in gene regulation including that of chromatin insulator function [5,6]. These results strongly suggest that the maternal ICR functions as an insulator element in regulating mutually exclusive expression of *Igf2* and *H19* in *cis*.

Address: Division of Biology, Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, California 91010-3011, USA.

Correspondence: Piroska E. Szabó
E-mail: pszabo@coh.org

Received: **2 March 2000**
Revised: **20 March 2000**
Accepted: **21 March 2000**

Published: **5 May 2000**

Current Biology 2000, 10:607–610

0960-9822/00/$ – see front matter

Results and discussion

The chromatin structure of the *Igf2* and *H19* ICR *in vivo* was investigated in mouse primary embryo fibroblasts (PEFs) that carry only maternally or paternally inherited copies of the distal chromosome 7 region on which these genes reside (MatDup.d7 and PatDup.d7, respectively). In MatDup.d7 embryos, on each of the two distal chromosome 7 regions, *Igf2* is silent whereas *H19* is active in *cis*, and in PatDup.d7 embryos, the opposite is true [7]. PEFs were first treated with DNase I, dimethyl sulfate (DMS) or ultraviolet (UV) light. The modification of DNA by these reagents is sensitive to bound protein, and areas of protein–DNA interaction show as footprints on ligation-mediated-PCR (LMPCR) genomic sequencing ladders. LMPCR analysis of DNA treated with these reagents after its isolation from PEFs, or *in vitro* analysis, served as a reference.

Evidence for protein binding at all four of the 21 bp repeats of the maternally derived ICR was obtained using all three types of *in vivo* treatment. Following DNase I treatment, MatDup.d7 PEFs displayed a prominent footprint after LMPCR over a large area within and adjacent to the repeat (Figure 1; compare lanes 3,4). In contrast, PatDup.d7 PEFs showed no such area of protection (compare lanes 5,6). Following DMS treatment, MatDup.d7 PEFs displayed a number of A and G residues of much higher or lower relative intensity within and adjacent to the repeats (compare lanes 7,8) whereas no footprints were observed for PatDup.d7 PEFs (compare lanes 9,10). Following treatment with UV light, MatDup.d7 PEFs showed a number of C and T bands of much higher or lower relative intensity within and adjacent to the repeats (6–4 photoproducts, compare lanes 11,12; pyrimidine dimers, compare lanes 15,16) whereas no footprints were observed for PatDup.d7 PEFs (compare lanes 14,18 with 13,17, respectively). Similar results were obtained for the other strands of these three repeats and for both strands of the fourth repeat (data not shown) and all data obtained are summarized in Figure 2b.

Inspection of the sites at which we observed protein binding revealed that all four possess a likely binding site for the zinc finger protein CTCF, as previously defined for vertebrate insulators [6] (see alignment in Figure 2c). Insulator elements have been found to lie between independently regulated genes and act to prevent interactions of genes with inappropriate enhancers [6,8]. CTCF–DNA binding is required for function of insulators of the chicken β-globin and human T cell receptor α/δ genes, and copy number enhances the activity of the chicken β-globin insulator *in vitro* [6]. Therefore, the presence of protein binding at each of four putative CTCF sites in the maternally derived ICR strongly suggests that it functions as an insulator element in regulating mutually exclusive expression of *Igf2* and *H19* in *cis*. The *Igf2* gene lies proximal to the ICR, with the *H19* gene and a shared set of enhancers lying distal. Thus, as in a previously proposed model [9], by virtue of this location, the ICR/insulator may prevent *Igf2* from interacting with the shared enhancers while, at the same time, allowing *H19* enhancer access.

The lack of *in vivo* footprints at the four putative CTCF-binding sites of the paternally derived ICR suggests that it

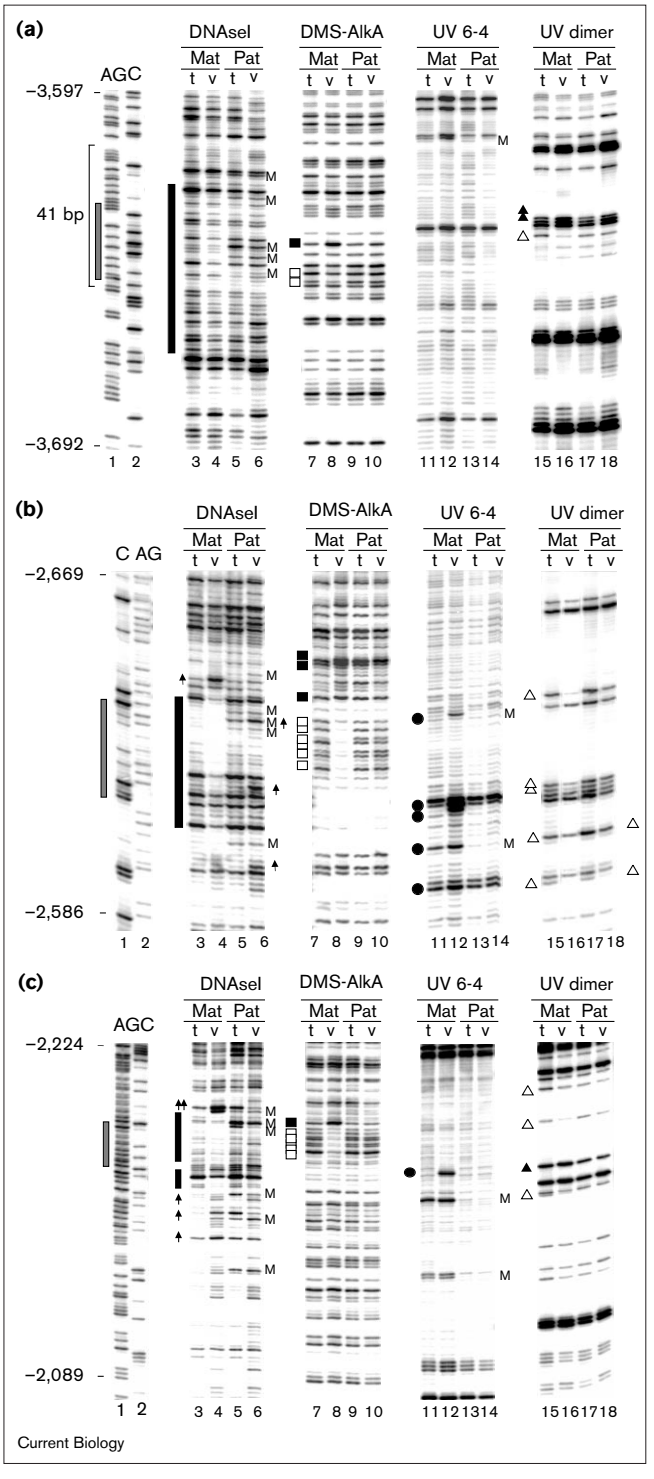


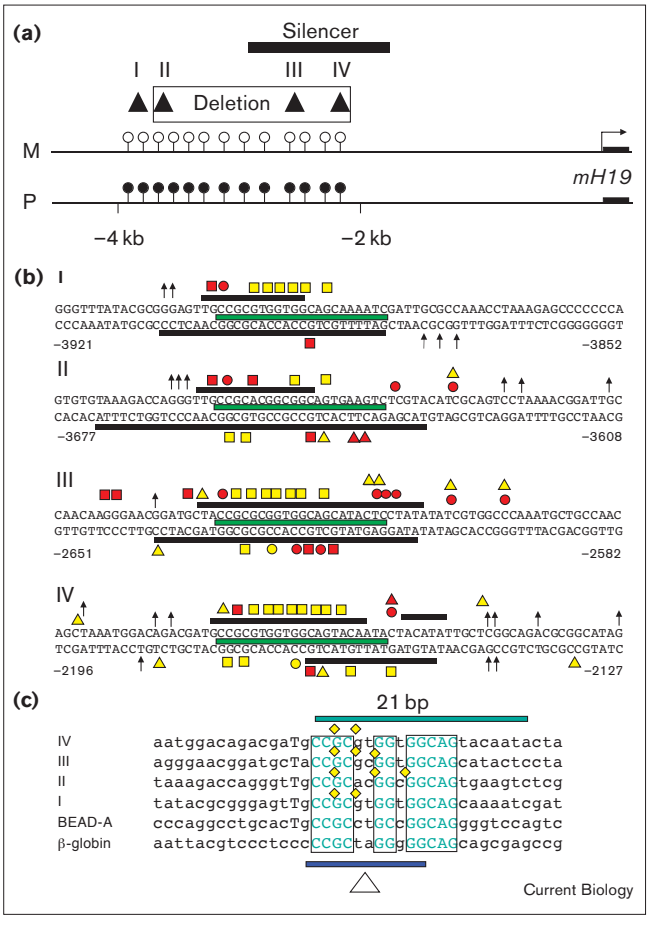
Figure 1

In vivo footprinting of ICR repeats. LMPCR was performed as described previously [12,16]. For sequences of oligonucleotides used, see the Supplementary material. **(a)** Repeat II (see Figure 2a), lower strand. **(b)** Repeat III, upper strand. **(c)** Repeat IV, upper strand. Distance in base pairs from the *H19* transcription start site is indicated (left of lane 1). Lanes 1 and 2, Maxam–Gilbert sequencing controls; shaded bar on left of lane 1, the 21 bp repeat [3]; bracket, 41 bp conserved sequence [4]. The type of treatment is indicated (top) and was either applied to DNA *in vivo* (v), that is, to intact MatDup.d7 (Mat) and PatDup.d7 (Pat) PEFs, or *in vitro* (t) after DNA isolation from PEFs. Prominent *in vivo* footprints are indicated for the maternal ICR (left) and paternal ICR (right). DNase I nicking was inhibited at regions of protein binding and showed as areas of protection (solid bar). Often, hypersensitive sites appeared next to protected areas (arrows). Methylation enhanced DNase I cleavage [10], and CpGs where methylation could be discerned are indicated (M). DMS reacted with protein-bound DNA with a higher or lower rate of methylation of A and G residues. Following cleavage at these residues with the AlkA protein [16], footprints indicative of protein binding showed as stronger (closed squares) or weaker (open squares) A and G bands. Induction of 6–4 photoproducts was inhibited or enhanced, and likewise for cyclobutane pyrimidine dimers, by protein binding upon exposure to UV light and, following cleavage with piperidine and T4 endonuclease V, respectively, stronger (closed circles and triangles) or weaker (open circles (seen on other strands, see Figure 2) and triangles) C and T bands were visualized [17]. Methylation at TCG and CCG trinucleotides results in an absence of bands after UV treatment because of the inhibition of 6–4 photoproduct formation [18], and CpGs where methylation could be discerned are indicated (M). For PEF derivation, MatDup.d7 and PatDup.d7 embryos at 13.5 days *post coitum* (dpc) were obtained as described previously [7] except that the latter were rescued from death at 10.5 dpc by the introduction of a functional mammalian achaete-scute homologue 2 (*Mash2*) transgene (A.R. and J.R.M., unpublished data). PEFs from normal embryos exhibited strict parental-specific expression of *Igf2*, *H19* and other imprinted genes (P.E.S. and J.R.M., unpublished data).

lacks insulator activity. It is possible that protein binding at these sites may be inhibited by the presence of DNA methylation. Each site contains at least two CpGs (Figure 2c) and methylation of these sites in midgestation embryos has been previously described [1]. Also, we observed methylation at these CpGs in PatDup.d7 PEFs (Figure 1, compare lanes 13 with 11, and 5 with 3). These results are consistent with a previously proposed model in which the paternal ICR lacks insulator activity because of methylation [9]. In this model, for paternal alleles, *Igf2* is allowed enhancer access and *H19* is repressed because of spreading of methylation from the ICR into the promoter. We found that, whereas the maternal *H19* promoter is occupied by a set of transcription factors, the paternal promoter is packaged into rotationally positioned nucleosomes (Figure 3). Such a chromatin structure of the promoter is repressive as it can contribute directly to the exclusion of transcriptional machinery [10,11]. Therefore, it probably plays an important role in the repression of the paternal *H19* allele. In this instance, it appears to represent a maturation of chromatin structure as it was not observed for the paternal *H19* promoter in mouse embryonic stem cells [12]. Based on the observation that the chicken β-globin insulator sequence is protective against heterochromatinization in *Drosophila* and human cells [13], the maternal ICR may also function as a 'barrier' [14] to heterochromatinization. The insulator model of ICR function [9] incorporating our present findings is given in Figure 4.

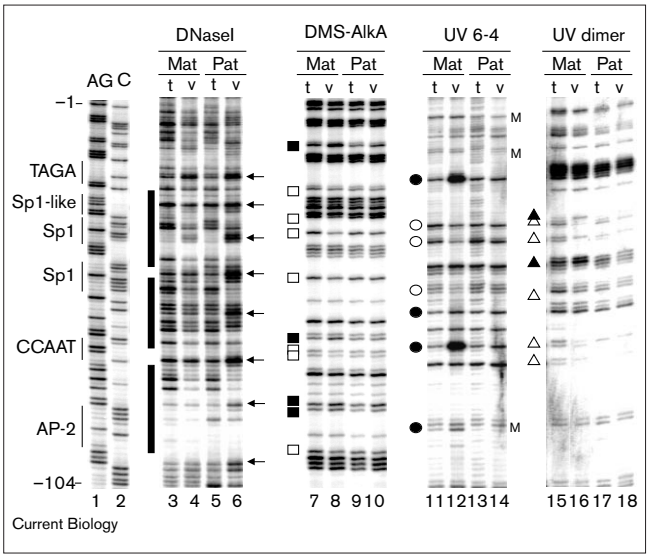
Our results provide a possible explanation for the observation that a segment of the mouse ICR (see Figure 2a)

Figure 2



Putative CTCF-binding sites in the maternal ICR are occupied by protein factors. **(a)** Definition of the ICR. A region −2 to −4 kb from the transcription start site (right-pointing arrow) of the mouse *H19* (*mH19*) gene is non-methylated (open circles on stems) on the maternal (M) chromosome, and methylated (closed circles on stems) on the paternal (P) chromosome. The region contains four conserved 21 bp repeats (closed triangles, I–IV) [3]. Maternal inheritance of a deletion in the ICR (rectangle) causes activation of *Igf2* and downregulation of *H19* in *cis*, whereas its paternal inheritance causes activation of *H19* and downregulation of *Igf2* in *cis* [2]. The ICR acts as a silencer in *Drosophila* [15]. The shortest region exhibiting silencer function is indicted (bar). **(b)** Summary of *in vivo* footprinting as described in the text and in Figure 1. Regions of sequence containing the four (I–IV) 21 bp repeats (green bar) are shown. Upper and lower strands of each repeat were analyzed. The upper line of the sequence is in the 5′ to 3′ direction, and distance from the *H19* transcription start site is given in base pairs. Prominent footprints are indicated by symbols as described in the legend to Figure 1 except that open symbols are now yellow and closed symbols are now red. **(c)** Subregion of the sequence shown in (b) at which footprints were congregated. The sequence was aligned with the CTCF-binding site of the insulators of the chicken β-globin and human T-cell receptor α/δ (BEAD-A) genes [6], revealing a common motif in repeats I–IV, which is likely to be a binding site for mouse CTCF (residues in blue). The 21 bp repeat is indicated by the bar (above), and the 14 bp deletion that disrupts chicken β-globin insulator function and CTCF binding [6] is indicated by the bar and open triangle (below). CpGs within the putative CTCF-binding sites I–IV are indicated (yellow diamonds).

Figure 3

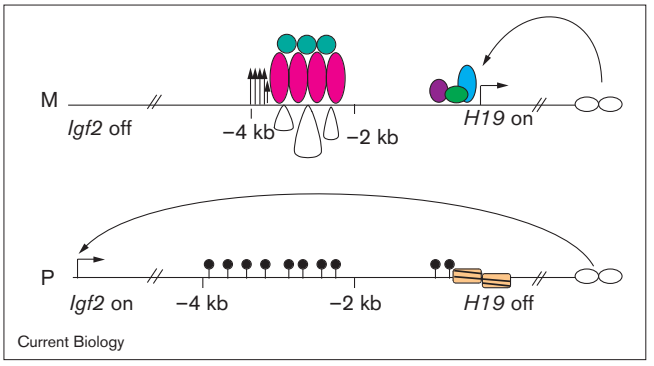


In vivo footprinting of the *H19* promoter. The lower strand is shown. Distance in nucleotides from the *H19* transcription start site and binding sites for common transcription factors are indicated (left of lane 1). All other details are as described in Figure 1. A periodicity of DNase I hypersensitive sites specific to the paternal promoter is indicated (arrows), showing that it is occupied by rotationally positioned nucleosomes. Other footprints were located exclusively on the maternal promoter and provide further evidence for maternal-specific transcription factor binding [12].

acts as a bidirectional silencer when introduced into *Drosophila melanogaster* [15], an organism that lacks DNA methylation. It is plausible that binding of a *Drosophila* protein to the ICR repeats could impart silencer function. That the ICR does not behave like an insulator in this system may be due to the lack of an appropriate interaction of the binding protein with a modifier. In *Drosophila*, the zinc-finger protein suppressor of Hairy-wing (Su(Hw)) is responsible for the insulator activity of the *gypsy* DNA element and this activity is dependent on the presence of the modifier of mdg4 (Mod(mdg4)) protein. This modifier interacts directly with Su(Hw) and, in its absence, the Su(Hw)–*gypsy* interaction has bidirectional silencer rather than insulator activity [8].

It is possible that maternal-specific ICR protein binding could be set up in the female germ line where oocyte-specific expression of the binding protein may inhibit the deposition of methylation as occurs in the male germ line. The protein–DNA interaction would therefore constitute a primary imprint [12]. Wide expression of CTCF and its involvement in the regulation of numerous genes [5,6] argues against such a role for this particular protein. It is, however, likely that CTCF-mediated insulator activity may require the participation of other proteins as does Su(Hw)-mediated insulator activity in *Drosophila* [6], and

Figure 4



Insulator model for the regulation of *Igf2* and *H19* expression in *cis*. Maternal chromosome (M): CTCF (red ovals) binds to four repeat sites in the ICR. A DNase I hypersensitive region (vertical arrows) is located just 5′ of the CTCF-binding region ([12] and data not shown), suggesting the presence of a polymer insulator structure. Other proteins (blue circles) might be necessary for this structure and could be gene-, tissue-, developmental stage-, and species-specific modifiers of CTCF function. UV dimer formation is sensitive to DNA bending and UV dimer-induced footprints were often observed next to DNase I footprints (see Figure 2b), suggesting that DNA could be looped out between the insulator binding sites (loops below red ovals). The insulator prevents *Igf2* from interacting with its enhancers (open circles) and only *H19* gains enhancer access. The promoter is bound by several transcription factors (purple, green and blue ovals) ([12] and Figure 3). Paternal chromosome (P): methylation (closed circles on stems) inhibits CTCF binding, the region lacks insulator activity, and *Igf2* is thereby able to interact with its enhancers. Methylation in the ICR spreads to the *H19* promoter and may inhibit transcription factor binding. The promoter is occupied by rotationally positioned nucleosomes (striped yellow rectangles; Figure 3).

these modifiers may exhibit more specific modes of expression and provide another level of control. It remains possible that methylation of the ICR in the male germ line is the primary imprinting event and this requires a methylation mechanism specific to male germ-cell development. Further insights into the possible roles of DNA modifications by protein binding and methylation in genomic imprinting should be obtained by an investigation of these phenomena in the germ cell lineages.

Supplementary material

Supplementary material including the sequences of oligonucleotides used is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

We thank Art Riggs and Mike Reed for valuable discussions, Steven Lloyd for T4 endonuclease V and Tim O'Connor for providing AlkA protein. This work was supported by NIH grant RO1GM48103-04A2 and NSF grant BIR-9220534.

References

1. Tremblay KD, Duran KL, Bartolomei MS: **A 5′ 2–kilobase–pair region of the imprinted mouse *H19* gene exhibits exclusive paternal methylation throughout development.** *Mol Cell Biol* 1997, **17:**4322-4329.
2. Thorvaldsen JL, Duran KL, Bartolomei MS: **Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*.** *Genes Dev* 1998, **12:**3693-3702.
3. Stadnick MP, Pieracci FM, Cranston MJ, Taksel E, Thorvaldsen JL, Bartolomei MS: **Role of a 461–bp G–rich repetitive element in *H19* transgene imprinting.** *Dev Genes Evol* 1999, **209:**239-248.
4. Frevel MA, Hornberg JJ, Reeve AE: **A potential imprint control element: identification of a conserved 42 bp sequence upstream of *H19*.** *Trends Genet* 1999, **15:**216-218.
5. Awadi TA, Bigler J, Ulmer JE, Hu YJ, Moore JM, Lutz M, *et al.:* **Negative transcriptional regulation mediated by thyroid hormone response element 144 requires binding of the multivalent factor CTCF to a novel target DNA sequence.** *J Biol Chem* 1999, **274:**27092-27098.
6. Bell AC, West AG, Felsenfeld G: **The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.** *Cell* 1999, **98:**387-396.
7. McLaughlin KJ, Szabó PE, Haegel H, Mann JR: **Mouse embryos with paternal duplication of an imprinted chromosome 7 region die at midgestation and lack placental spongiotrophoblast.** *Development* 1996, **122:**265-270.
8. Gdula DA, Gerasimova TI, Corces VG: **Genetic and molecular analysis of the *gypsy* chromatin insulator of *Drosophila*.** *Proc Natl Acad Sci USA* 1996, **93:**9378-9383.
9. Schmidt JV, Levorse JM, Tilghman SM: **Enhancer competition between *H19* and *Igf2* does not mediate their imprinting.** *Proc Natl Acad Sci USA* 1999, **96:**9733-9738.
10. Pfeifer GP, Riggs AD: **Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNaseI and ligation-mediated PCR.** *Genes Dev* 1991, **5:**1102-1113.
11. Li G, Chandler SP, Wolffe A, Hall TC: **Architectural specificity in chromatin structure at the TATA box *in vivo*: nucleosome displacement upon β–phaseolin gene activation.** *Proc Natl Acad Sci USA* 1998, **95:**4772-4777.
12. Szabó PE, Pfeifer GP, Mann JR: **Characterization of novel parent-specific epigenetic modifications upstream of the imprinted mouse *H19* gene.** *Mol Cell Biol* 1998, **18:**6767-6776.
13. Bell AC, Felsenfeld G: **Stopped at the border: boundaries and insulators.** *Curr Opin Genet Dev* 1999, **9:**191-198.
14. Sun F-L, Elgin SCR: **Putting boundaries on silence.** *Cell* 1999, **99:**459-462.
15. Lyko F, Brenton MA, Surani MA, Paro R: **An imprinting element from the mouse *H19* locus functions as a silencer in *Drosophila*.** *Nat Genet* 1997, **16:**171-173.
16. Szabó PE, Pfeifer GP, Miao F, O'Connor TR, Mann JR: **Improved *in vivo* DMS footprinting using AlkA protein: DNA-protein interactions at the mouse *H19* gene promoter in primary embryo fibroblasts.** *Analyt Biochem* 2000, in press.
17. Tornaletti S, Pfeifer GP: **UV light as a footprinting agent: modulation of UV-induced DNA damage by transcription factors bound at the promoters of three human genes.** *J Mol Biol* 1995, **249:**714-728.
18. Pfeifer GP, Drouin R, Riggs AD, Holmquist GP: ***In vivo* mapping of a DNA adduct at nucleotide resolution: detection of pyrimidine (6–4)pyrimidone photoproducts by ligation-mediated polymerase chain reaction.** *Proc Natl Acad Sci USA* 1991, **88:**1374-1378.