# Comparative sequence analysis of mammalian factor IX promoters

C.-P.Pang, M.Crossley, G.Kent. and G.G.Brownlee
Chemical Pathology Unit, Sir William Dunn School of Pathology, University of Oxford, OX1 3RE, UK

Functionally important nucleotides tend to be conserved between species. The advent of PCR has made it possible to rapidly obtain corresponding sequences from numerous species. We have used this approach to obtain sequence data on the factor IX promoter from various mammals, in the hope of determining which nucleotides contribute to factor IX expression *in vivo*.

The results of CAT assays carried out in transiently transfected HepG2 cells, suggest that the promoter lies in the region −98 to +21 (3) (numbering throughout is relative to the most 5′ start point of transcription in the human gene, +1 (1, 2)). This region has been shown to contain sequences which can bind NF1 (−99 to −77) and C/EBP (+1 to +18) *in vitro* (3). Further evidence that this region is involved in factor IX expression comes from the study of individuals with altered levels of factor IX, and who suffer from a form of haemophilia, known as haemophilia B Leyden. Mutations, in these patients, have been detected at −20, −6 and at +13 (4, 5, 6). Moreover, it has been shown that the A→G mutation at +13 interferes with the binding of C/EBP to its recognition site (+1 to +18) (3).

It is known that the factor IX structural gene is well conserved in mammalian species and cross-species amplification of coding regions has been achieved (7). We took the same approach in an effort to amplify 5′ flanking sequences. Two downstream and three upstream primers were prepared. One of the downstream primers, +84 to +106 (A) was homologous to the 5′ region of the recently reported mouse cDNA (8), the other was homologous to the human sequence, +71 to +111 (B) (9). The three upstream primers, −133 to −110 (C), −229 to −206 (D) and −323 to −298 (E), were all homologous to the human sequence. Successful amplification of the mouse factor IX 5′ region was achieved with primers A and C as well as B and C, whilst the rat factor IX 5′ region was amplified only with B and C. In the case of the monkey, all combinations of upstream and downstream primers resulted in successful amplification. The amplified fragments were gel purified and directly sequenced (10). They were also cloned into pBluescript sk+ (Stratagene) and several independent plasmid isolates were sequenced. Difficult regions were checked by reamplifying with one 5′ [32]P end-labelled and one unlabelled primer (11) and sequencing the resulting fragment by the Maxam and Gilbert method.

```
human     GTCCCAAGAG GCCATTGGAA ATAGTCCAAA GACCCATTGA GGGAGATGGA CATT-ATTTC CCAGAAGTAA
macaque   GTCCCGAGAG GCCATTGGAA ATAGTCCAAA GACCCATTGA GGGAGATGGA CATT-ATTTC CCAGAAGTAA
dog       AACCCAAGAG GCCACTGGAA ATAGCCCAAA GACCCACTGA GGGAGATGGA CACT-GTTTC CCAGAAGTAA
rat       GACCCAAGAA ACAACTGGAA ATAGCCCAAA GATACACTGA GGGAGATGGA CACCAATTTC CCAGAAGTAA
mouse     GATCCAAGAA CCAACTGGAA ATAGCCCAAA GATACACCGA GGGAGATGGA CAACAATTTC CCAGAAGTAA
          -100        -90        -80        -70        -60        -50
                    NF1

                                              -6                    +13
                              -20             G->A                   A->G
                              T->A            G->C                   A->A
                              ●               ●         ┌─► ┌─►      ●              ┌─►
human     ATACAGCTCA GCTTGTACTT TGGTACAACT AATCGACCTT ACCACTTTCA CAATCTGCTA GCAAAGGTTA
macaque   ATACAGCTCA GCTTGTACTT TGGTACAACT AATCGACCTT ACCACTTGCA CAATCAGCTA GCAAAGGTTA
dog       ATACAGCTCA ACTTGTACTT TGGAACAACT GGTCAACCTT GCTCCTTGCA CAATTCATCA GCAAAGGTTA
rat       ATCCCATTCA GCTTGTACTT TGGAACAATT GATTAGCCCT GACACTTGCA CAATCTCCTA ACAAAGGTCA
mouse     GTCCCATTCA GCTTGCACTT TGGAACGATT GATTAGCCCT GACGCTTGCA CAATCTCCTA ACAAAGGTCA
          -40         -30        -20        -10         +1        +11        +21
                                        C/EBP

human     TGCAGCGCGT GAACATGATC ATGGCAGAAT CACCAGGCCT C
macaque   TGCAGCGTGT GAACATGATC ATGGCAGAAT CACCAGGCCT C
dog       GGCAGCGCCT GAATAGGATC ATGGCAGAAG CATCGGGCCT C
rat       GGCGGCACCT GAACGCCATC ATGGCAGACG CTCCGGGTCT C
mouse     TGAAGCACCT GAACACCGTC ATGGCAGAAT CCCGGCCTCT C
          +31         +41        +51        +61        +71
```
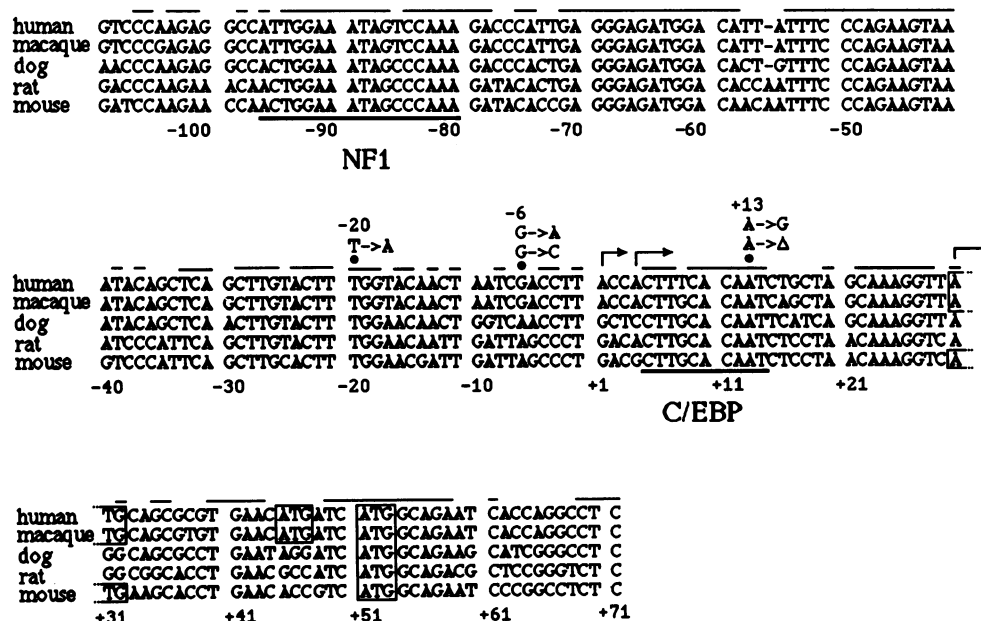
Figure 1. The sequence of the putative promoter regions from five mammalian species is shown. Nucleotides conserved in all species are overlined, the core NF1 and C/EBP recognition sequences are underlined, mutations associated with reduced expression of factor IX and the haemophilia B Leyden-like conditions are shown by dots (4, 5, 6), transcriptional start points are indicated by arrows (1, 2) and potential initiation codons are boxed.

The derived sequences and the previously reported region of the dog (12) and human factor IX genes (9) are shown in figure 1. Neither the NF1 nor C/EBP binding sites are completely conserved in all five species but no residues are found that depart from accepted binding sites (13, 14, 15). These findings are compatible with the hypothesis that these factors bind and contribute to factor IX transcription *in vivo*.

The residue at −20 is conserved, as are the surrounding bases, but it is not clear what protein, if any, binds this region. The −6 residue, on the other hand, is not conserved. In fact, whilst a G→A change at −6 is associated with a mild form of haemophilia B Leyden in humans (5) and has been shown to interfere with promoter activity by CAT assays in HepG2 cells (3), an A instead of the expected G is present at this location in the dog, rat and mouse. There are at least two possible explanations for this unusual finding. As fig. 1 shows, the nucleotides around −6 are not perfectly conserved. It may be that one of the other changes in this area compensates for this G→A change at −6 in these animals. Alternatively, it is possible that transcription of the factor IX gene is differently regulated in these species and the −6 residue is not involved. The recent isolation of a dog cDNA, which begins at −179, makes this an intriguing possibility (12). There is, however, no information about the relative importance of this upstream initiation site. It may be that this cDNA originated from a rare, uncharacteristic initiation event. On the other hand, there may be a major or alternative start point in the vicinity of −179 in the dog. It has also been reported that a human factor IX−CAT fusion construct, when transfected into HepG2 cells, does initiate transcription from −150 and various other upstream start sites (16). Subsequent studies, however, by independent workers, have failed to detect any upstream starts in the endogenous human gene (2), suggesting that the reported upstream starts are actually artefacts of the CAT fusion construct.

Finally, these results extend the information about the start point of translation of factor IX (12). Although there are three potential, in-frame initiation codons in the 5' transcribed region of the human factor IX gene, only the third of these is conserved in all five species (Fig. 1). Moreover, this codon falls into a more extensive conserved sequence which closely matches the optimal initiation codon defined by Kozak (17). These results do not prove that the third initiation codon is used in all these species but they do suggest that amino-acids upstream of this are not required for factor IX activity.

## REFERENCES

1. Anson,D.S. *et al.* (1984) *EMBO J.* **3**, 1053−1060.
2. Reijen,M.J. *et al.* (1990) *FEBS Letters*, **270**, 20.
3. Crossley,M. and Brownlee,G.G. (1990) *Nature* **345**, 344−346.
4. Reitsma,P.H. *et al.* (1988) *Blood* **72**, 1074−1076.
5. Crossley,M. *et al.* (1990) *Nucl. Acids Res.* **18**, 4633.
6. Attree,O. *et al.* (1989) *Thromb. Haemostas.* **62**, 154.
7. Sarkar,G. and Sommer,S.S. (1989) *Science* **244**, 331−334.
8. Wu,M.-H. *et al.* (1990) *Gene* **86**, 275−278.
9. Yoshitake,S. *et al.* (1985) *Biochemistry* **24**, 3726−3750.
10. Winship,P.R. (1989) *Nucl. Acids Res.* **17**, 1266.
11. Romeo,P.-H. *et al.* (1990) *Nature* **344**, 447−449.
12. Evans,J.P. *et al.* (1989) *Blood* **74**, 207−212.
13. Ryden,T.A. and Beamon,K. (1989) *Mol. Cell. Biol.* **9**, 1155−1164.
14. Nowock,J. *et al.* (1985) *Nucl. Acids Res.* **13**, 2045−2061.
15. Gronostajski,R.M. (1987) *Nucl. Acids Res.* **15**, 5545−5559.
16. Salier,J.P. *et al.* (1990) *J. Biol. Chem.* **265**, 7062−7068.
17. Kozak,M. (1987) *Nucl. Acids Res.* **15**, 8125−8149.