Volume 231, number 1, 164–171 FEB 05696 April 1988

The IFI-56K and IFI-54K interferon-inducible human genes belong to the same gene family

Marc G. Wathelet, Isabelle M. Clauss, Jean Content* and Georges A. Huez

Département de Biologie Moléculaire, Université Libre de Bruxelles, 67 rue des Chevaux, B-1640 Rhode-St-Genèse and **Institut Pasteur du Brabant, 642 rue Engeland, B-1180 Bruxelles, Belgium*

Received 28 December 1987; revised version received 1 February 1988

The IFI-56K and IFI-54K human genes are coordinately regulated by interferon, double-stranded RNA and viruses in a number of cell lines. These genes encode polypeptides of 56 and 54 kDa, respectively, whose function remains to be determined. We analysed the possible structural relatedness between these syntenic and similarly regulated genes. We found that they are very closely related at the protein, mRNA and promoter levels. This suggests that the IFI-56K and IFI-54K genes are members of a gene family, which probably arose from duplication of an ancestor gene.

Poly(rI) · poly(rC)-inducible gene; Virus-inducible gene; Sequence comparison; Needleman-Wunsch-Sellers algorithm

1. INTRODUCTION

Viruses and double-stranded RNA activate the transcription of numerous cellular genes [1–5]. Interferon (IFN)-β is one of these cellular inducible genes and the release of the corresponding mature protein in the extracellular medium is in turn followed by the transcriptional stimulation of a set of cellular genes in the neighbouring cells [6].

The study of IFN-inducible gene promoters showed that some of them contain a sequence homologous to part of the IFN-β promoter [7,8]. We observed that the presence of such a homology in two of these genes, namely IFI-56K and IFI-54K, is correlated with their direct inducibility by poly(rI) · poly(rC) and viruses [7,8]. Thus, interestingly, the set of genes that is induced by poly(rI) · poly(rC) and viruses overlaps to some extent the set that is activated by IFN. Such an overlap has also been demonstrated for the proteins induced by various agents such as IFN-α, IFN-γ, interleukin-1 and tumor necrosis factor [9,10] or for poly(rI) · poly(rC) and platelet derived growth factor [2]. The biological activities that these agents have in common might be due to this overlap in the pattern of inducible genes and could reflect the use of identical second messengers.

The IFI-56K and IFI-54K genes have many common features. Indeed, the kinetics of mRNA accumulation in response to either IFN, poly(rI) · poly(rC) or viruses are very similar in the numerous cell lines studied [8,11–13]. This phenomenon is likely to be mediated by the DNA sequence the IFI-56K and IFI-54K genes have in common in their respective promoters [8]. In addition, we showed that the IFI-56K and IFI-54K genes are syntenic, residing both on chromosome 10 (Wathelet et al., submitted). These observations led us to examine if these genes bear any other structural relationships.

Here we show that the sequence of the IFI-56K and IFI-54K genes are very closely related at both the protein and DNA levels. Moreover, we have identified in human DNA two pseudogenes which

Correspondence address: M. Wathelet, Département de Biologie Moléculaire, Laboratoire de Chimie Biologique, Université Libre de Bruxelles, 67 rue des Chevaux, B-1640 Rhode-St-Genèse, Belgium

Abbreviations: IFN, interferon; poly(rI) · poly(rC), polyribo-inosinic-polyribocytidylic acid

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession number Y009186

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

are related to the IFI-56K and IFI-54K genes. One of them is located on chromosome 13 (IFI-56K-1) (Wathelet et al., submitted) and the other (IFI-56K-2) maps close to the IFI-56K locus on chromosome 10. The IFI-56K-1, IFI-56K-2, IFI-56K and IFI-54K genes all belong to the same gene family.

2. MATERIALS AND METHODS

We used the algorithm of Needleman and Wunsch [14], modified by Sellers [15], and by Goad and Kanehisa [16] to compare various sequences. Amino acid and nucleotide sequences were derived from previously published cDNA and genomic DNA sequences [7,13,17].

For nucleotide sequence alignments, the statistical significance of a homology is determined as follows: a weight is assigned to each kind of discrepancy and allows one to calculate the density of weighted discrepancies; all alignments produced by the algorithm are classified according to the calculated value and those for which this value is above a parameter set by the user are eliminated.

For amino acid sequence alignments, the same algorithm is used together with a scoring matrix derived by Dayhoff and coworkers [18]. This matrix reflects the probability of the mutation of each amino acid into any other for a given evolutionary distance (250 point accepted mutations in this case, see [18]). When one protein is compared to another, one should multiply the odds for each position to calculate the odds for the whole protein. For convenience, the logarithm (multiplied by ten) of the odds matrix is used and the logarithms of the matrix elements are added, with a penalty of −8 for deletions. This sum is termed the 'distance', and its increase corresponds to an increase in the evolutionary relatedness between the two polypeptides.

3. RESULTS

3.1. *The IFN-inducible 56 kDA and 54 kDA polypeptides are closely related*

The human 56 kDa and 54 kDa proteins deduced amino acid sequences were compared using the algorithm of Goad and Kanehisa (see section 2). Fig.1 shows that these polypeptides could be aligned from the initiating methionine up to 13 and 14 amino acids from the carboxy-terminus, respectively. The program calculates the evolutionary relatedness, the 'distance', for an alignment between two polypeptides by adding the values attributed to each amino acid pair. Matching amino acids scored between +2 and +17, depending on the mutation probability of the given amino acid. Different amino acids scored between +7 and −8, depending on their relatedness, and a deletion has a weight of −8. The distance obtained for the

```
   4         14        24        34        44        54        64        74        84        94
MSTNGDDHQVKDSLEQLRCHFTWELSIDDDEMPDLENRVLDQIEFLDTKYSVGIHNLLAYVKHLKGQNEEALKSLKEAENLMQEEHDNQANVRSLVTWGN
:: :        :: :: *:::: :        : :  :    ::          ::::: :::::::: ::  :  :: : : ::  ::  ::::::::
MSEN NKNSLESSLRQLKCHFTWNLMEGENSLDDFEDKVFYRTEFQNREFKATMCNLLAYLKHLKGQNEAALECLRKAEELIQQEHADQAEIRSLVTWGN
   4         14        24        34        44        54        64        74        84        94

104       114       124       134       144       154       164       174       184       194
FAWMYYHMGRLAEAQTYLDKVENICKKLSNPFRYRMECPEIDCEEGWALLKCGGKNYERAKACFEKVLEVDPENPESSAGYAISAYRLDGFKLATKNHKP
 :: :::::::   : : :::   * : : :  :: : :: :*::::  ::*::   :::: *::: ::  : :::   : ::  ::::
YAWVYYHMGRLSDVQIYVDKVKHVCEKFSSP  YRIESPELDCEEGWTRLKCGGNQNERAKVCFEKALEKKPKNPEFTSGLAIASYRLDNW  PP SQNA
103       113       123       133       143       153       163       173       183       193

204       214       224       234       244       254       264       274       284       294
FSLLPLRQAVRLNPDNGYIKVLLALKLQ   DEGQE AEGEKYIEEALANMSSQTYVFRYAAKFYRRKGSVDKALELLKKALQETPTSVLLHHQIGLCYK
    ::::: :::::: : ::::::::     :: :  ::::  ::::      : : : ::::::::   ::: :::::::   :    :: ::: *:
ID  PLRQAIRLNPDNQYLKVLLALKLHKMREEGEEEGEGEKLVEEALEKAPGVTDVLRSAAKFYRRKDEPDKAIELLKKALEYIPNNAYLHCQIGCCYR
  198       208       218       228       238       248       258       268       278       288

300       310       320       330       340       350       360       370       380       390
AQMIQIKEATKGQPRGQNREKLDKMIRSAIFHFESAVEKKPT FEVAHLDLARMYIEAGNHRKAEENFQKLLCMKPVVEETMQDIHFYYGRFQEFQKKSD
:   :          :  : ::   :  :  :   : :     : :    ::     :     ::  :::    :       :  :  :: ::  : :
AKVFQVMNLRENGMYGK R KLLELIGHAVAHLKKADEANDNLFRVCSI LASLHALADQYEDAEYYFQKEFS KELTPVAKQLLHLRYGNFQLYQMKCE
296       306         316       326       336        346       356        366       376       386

399       409       419       429       439       449       459
VNAIIHYLKAIKIEQASLTRDKSINSLKKLVLRKLRRKALDLESLSLLGFVYKLEGNMNEALEYYERAL
  :: :     :: : :    :    : :     :     : : : :  : :   :   :  : :  :: :
DKAIHHFIEGVKINQKSREKEKMKDKLQKIAKMRLSKNGADSEALHVLAFLQELNEKMQQADEDSERGL
392       402       412       422       432       442       452
```

Fig.1. Alignment of the 56 kDa and 54 kDa deduced amino acid sequences. Conserved amino acids between the 56 kDa (top line) and 54 kDa (bottom line) putative polypeptides are indicated by (:) and the six conserved cysteines by (*). Numbering is from the initiating methionine.

alignment between the IFI-56K and IFI-54K putative polypeptides was +1109.

The statistical significance of this homology was evaluated as follows. Using the amino acid composition of the IFI-56K and IFI-54K sequences presented in fig.1, twenty sequences of 464 and 457 residues respectively, were randomly generated. They were compared to each other and the distances were calculated using the same algorithm. The distances collected in this way average $+55.8 \pm 13.9$. The +1109 distance found between the IFI-56K and IFI-54K polypeptides is thus at 75.8 standard deviation units from the mean value.

Furthermore, the high homology between these polypeptides is reflected in some of their features: (i) the 56 kDa and 54 kDa putative polypeptides have a very similar amino acid composition, rich in charged residues (32.6 and 33.7%, respectively), at the expense of uncharged polar amino acids (28.4 and 29.0%, respectively) as compared to the average composition of proteins (38.9% apolar, 25.1% charged and 36.0% uncharged polar amino acids); there is a striking bias for positively charged residues (17.6 and 17.8%, respectively) in agreement with the isoelectric point determined for the in vitro synthesized 56 kDa protein ($pI = 7.85$, M.W., unpublished). (ii) In the 54 kDa polypeptide, 218 out of 472 (46%) amino acids are perfectly conserved with the 56 kDa polypeptide; this homology is particularly striking in the first two third of each molecule. (iii) Remarkably, six cysteines (out of the eight present in the 56 kDa polypeptide) are conserved between the two sequences.

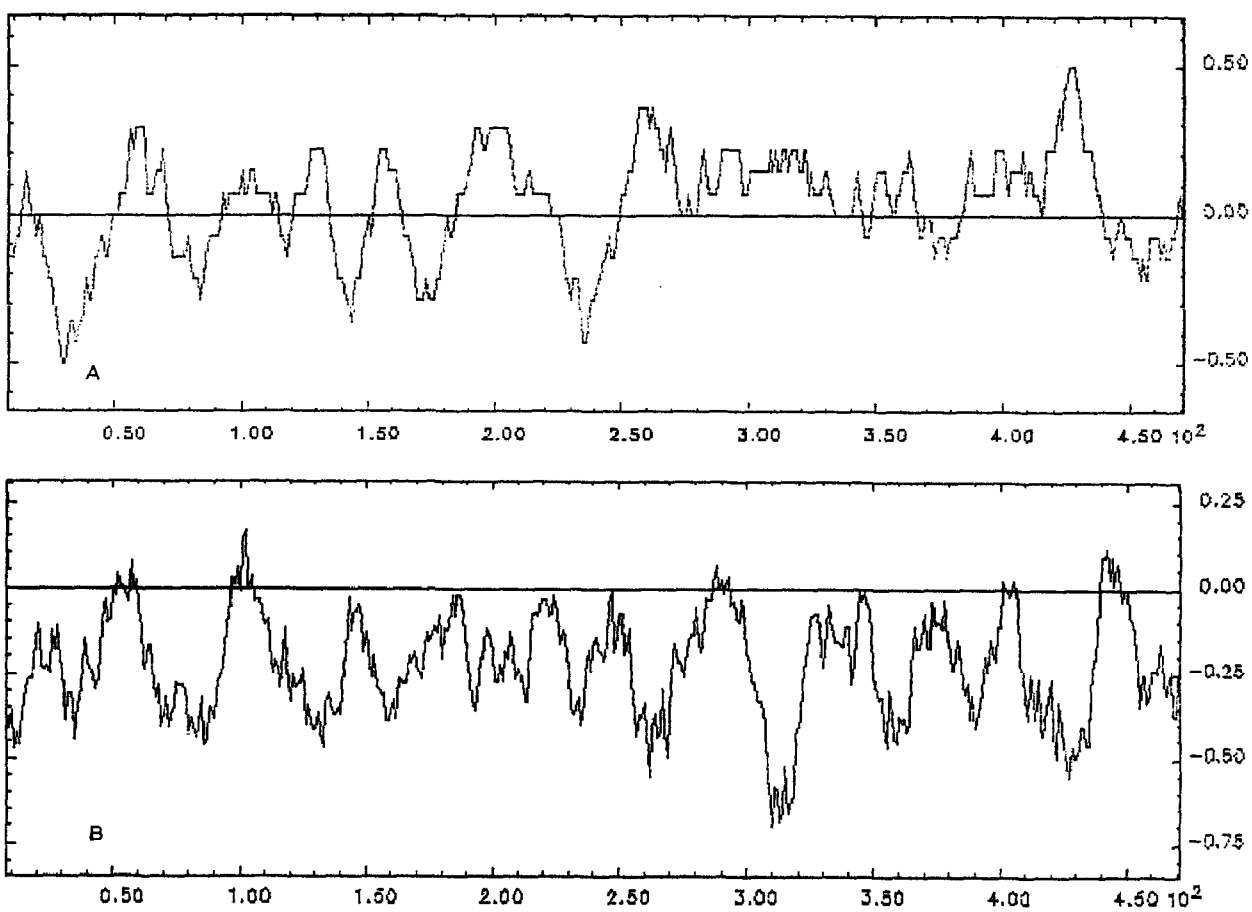


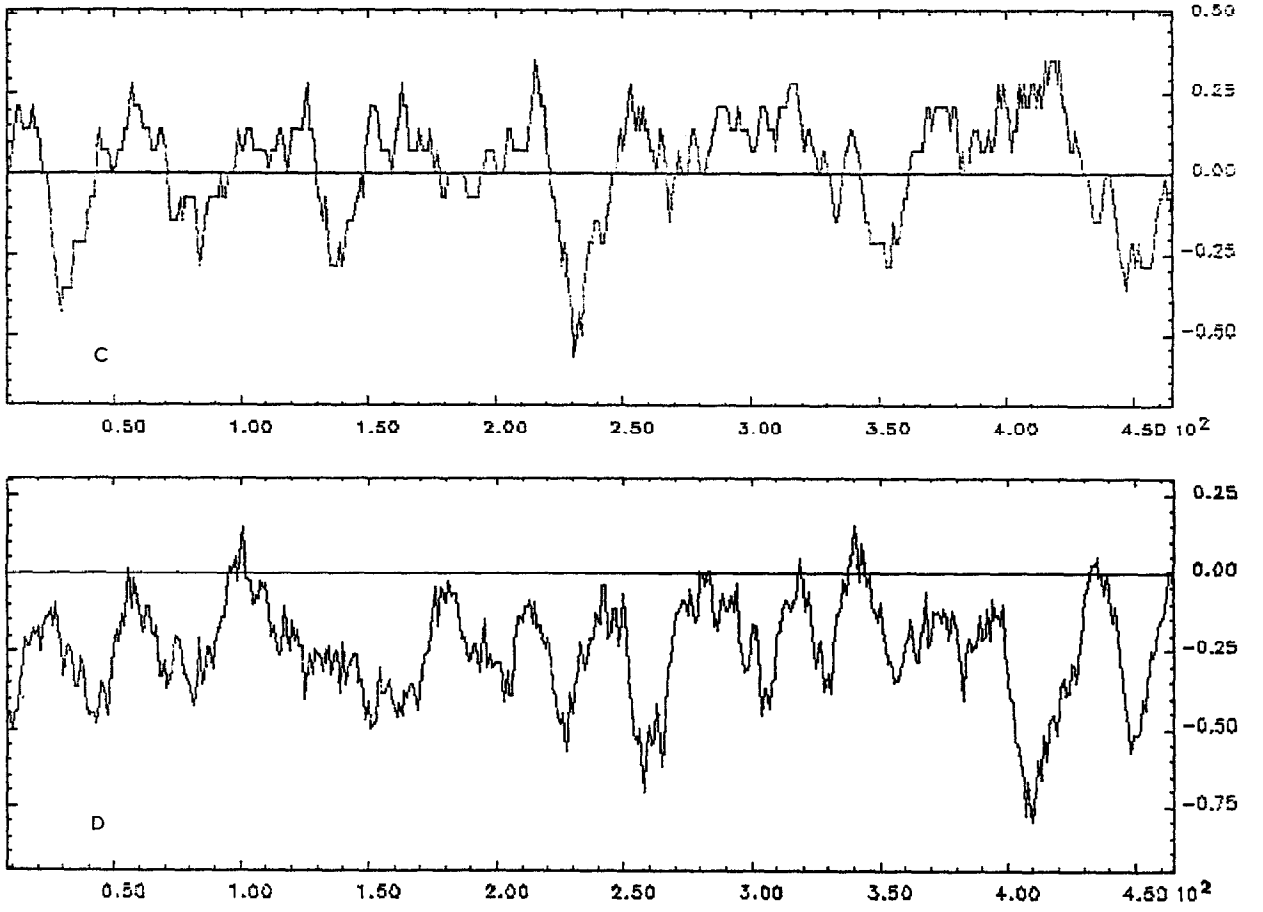
Fig.2. Hydrophobicity profile and charge distribution of the 56 kDa and 54 kDa putative polypeptides. The amino acid sequence of the 56 kDa (A,B) and 54 kDa (C,D) polypeptides plotted by means of a program that progressively evaluates charge (A,C) or hydrophobicity (B,D) of a span of 14 amino acids [29].

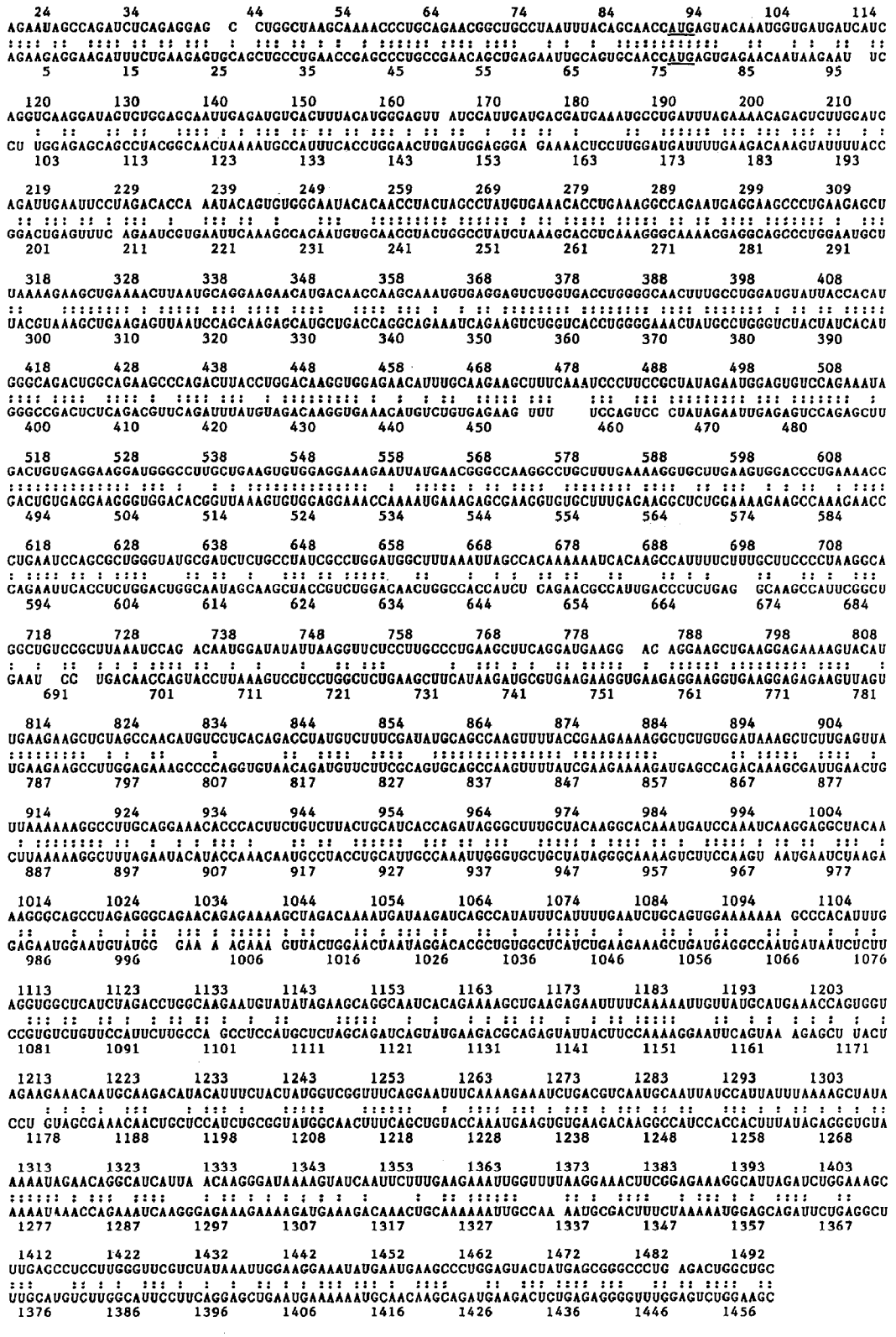
(iv) The hydrophobicity and charge plots of each amino acid sequence are similar as shown in fig.2.

With the aim to determine if the homology between the 56 kDa and 54 kDa polypeptides could be biologically relevant, we compared amino acid sequences whose biological relatedness is known, using the same algorithm. For instance, the alignments between some of the human IFN polypeptides score as follows (distance and standard deviation units; percentage of perfect matches; see also [18–21]): IFN-αA and αB, +761, 60 SD, 80.3%; IFN-αA and β1, +341, 29 SD, 33.9%; IFN-β1 and β2, +53, 2.3 SD, 5.8% (values on the best alignment obtained, which covers only one half of the molecule in the latter case). The 56 kDa and 54 kDa polypeptides appear thus to be more related to each other than IFN-α and β, suggesting that this homology may have a physiological meaning. The screening of data banks for other homologous sequences does not reveal any other striking relationships; among the best alignments obtained, we found a homology with two other IFN-inducible proteins, namely the murine Mx (as previously noted by Levy and co-workers [17]) and human 2-5A synthetase, and with human IFN-β itself; however, the distances for these alignments range between 70 and 84, values which are at the limit of statistical significance.

3.2. *The IFI-56K and IFI-54K mRNA are homologous in their 5'-untranslated sequence and their coding region*

The IFI-56K and IFI-54K complete mRNA sequences were derived from cDNA and genomic sequences [7,13,17]. Comparison of these sequences using the algorithm of Goad and Kanehisa [16] indicated that they are closely related. However, both nucleotide sequences could not be aligned on all their length in contrast with the amino acid sequences. This is due to the fact that the parameters





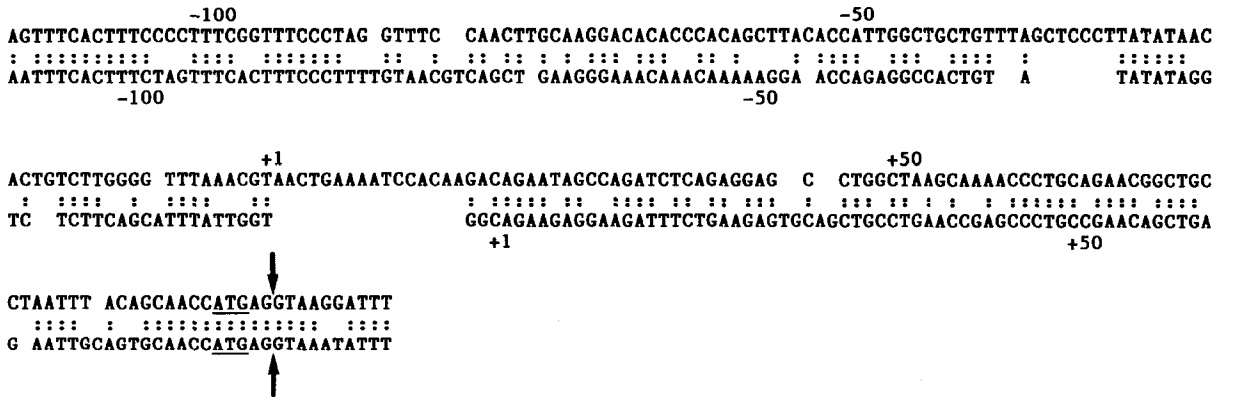


Fig.4. Alignment of the IFI-56K and IFI-54K genes 5′-flanking regions. The 5′-flanking region sequence of the IFI-56K and IFI-54K genes were aligned using the algorithm of Goad and Kanehisa, with parameter values of −1, 0.75 and 3 for match, mismatch and deletion, respectively. This produces two alignments which were brought together to form the longer one displayed here, leaving two gaps. Numbering is from the first transcripted nucleotide. The initiating ATG is underlined and the splicing point is indicated by an arrow.

of the algorithm were chosen in order to find locally highly homologous segments. By reducing the penalty for mismatches, we obtained the alignment shown in fig.3, with an overall homology of 60%. This value is higher than the one calculated for the amino acid sequences, suggesting that the IFI-56K and IFI-54K genes are derived from a common ancestor gene through a duplication event.

Interestingly, the 5′-untranslated regions, in contrast to the 3′-ones, are also conserved; the preservation of these regions suggests that they could play a physiological role (such as being the target of the posttranscriptional control to which both mRNA are subjected, for instance).

3.3. *Homologies in the genomic 5′-flanking region of the IFI-56K and IFI-54K genes*

We have previously described the existence of a short and highly homologous segment in the IFI-56K and IFI-54K genes promoters [7]. This element is thought to be involved in the responsiveness of these genes to IFN, poly(rI)·poly(rC) and viruses. Comparison of the 5′-flanking structure of the IFI-56K and IFI-54K genes using a reduced penalty for discrepancies showed that these genes are also related at the genomic level: (i) both sequences could be aligned from the highly homologous segment located around −100 up to the 10th nucleotide in the first intron of each gene (both numbering relative to the first transcription initiation site, fig.4). (ii) The first splicing event occurs two nucleotides after the initiating AUG, a characteristic which is reminiscent of other IFN-inducible genes [24,25].

4. DISCUSSION

4.1. *The IFI-56K and IFI-54K genes are members of a gene family*

Convergent evolution is associated with homologies at the amino acid level only. The IFI-56K and IFI-54K genes are closely related both at the amino acid and nucleotide levels, and have a common 5′-genomic structure. This strongly suggests that both genes arose through a duplication of an earlier gene and diverged thereafter. Such a phenomenon is relatively common and is for instance believed to be at the origin of the type I interferon gene family [26].

Several observations indicate that the

Fig.3. Alignment of the IFI-56K and IFI-54K mRNA. The IFI-56K and IFI-54K mRNA sequences were aligned using the algorithm of Goad and Kanehisa [16], with parameter values of −1, 0.75 and 3 for match, mismatch and deletion, respectively. Numbering is from the first transcripted nucleotide and the initiating AUG is underlined. The alignment obtained corresponds to the 5′-untranslated sequence and nearly all the coding region; the mRNA sequences corresponding to the last 13 amino acids and to the 3′-non-coding region are divergent and not presented here.

IFI-56K/IFI-54K gene family is not restricted to these two members. Indeed, we have described the existence of a pseudogene (IFI-56K-1) homologous to the IFI-56K cDNA and located on chromosome 13, whereas the functional IFI-56K and IFI-54K genes reside on chromosome 10 (Wathelet et al., submitted). Furthermore, we recently observed the existence of a second pseudogene, termed IFI-56K-2. This pseudogene, which also maps on chromosome 10, is 84 and 65% homologous to the IFI-56K and IFI-54K genes, respectively (Marinx, O. and Wathelet, M., unpublished). The phenomenon(s) that generate(s) the existence of pseudogenes IFI-56K-1 and -2 remain(s) to be determined, but the fact that the homology is much higher between IFI-56K and IFI-56K-2 than between IFI-56K and IFI-54K sequences, suggests that the appearance of this pseudogene is a more recent event than the duplication that gave rise to the IFI-56K and IFI-54K genes.

4.2. *Biological significance of the relatedness between the IFI-56K and IFI-54K genes*

The statistically very significant homology between the 56 kDa and 54 kDa putative polypeptides, the similarities in their hydrophobicity and charge profiles, together with the conservation of six cysteines, suggest that the two polypeptides may adopt a similar secondary and tertiary structure, and hence might have a common biological activity. Nevertheless, there could be a functional difference between the products of these genes which would account for their preservation throughout the evolutionary process.

We have suggested that both proteins are involved in the antiviral effect of IFN [8]. If it is indeed the case, the existence of multiple forms of a given biological activity could represent a selective advantage by helping the organism to cope with the diversity of viruses.

In this respect, it is interesting to note that in humans, four isoforms of 2-5A synthetase have been detected [27] and that two gene families have been identified, which correspond to IFN-inducible genes 1-8 and 6-26 [28].

Acknowledgements: We thank Dr M. Colet for help in computing and I. Tavitian for critical reading of the manuscript. This work was supported by ULB-Actions de Recherche Concertées of the Belgian Government, by a grant to G.H. and J.C. from the Fund for Medical Scientific Research (Belgium) no. 34.560.85. M.W. is Research Assistant of the National Fund for Scientific Research (Belgium).

REFERENCES

[1] Skup, D., Windass, J.D., Sor, F., George, H., Williams, B.R.G., Fukuhara, M., De Maeyer-Guinard, J. and De Maeyer, E. (1982) Nucleic Acids Res. 10, 3069-3084.
[2] Zullo, J.N., Cochran, B.H., Huang, A.S. and Stiles, C.D. (1985) Cell 43, 793-800.
[3] Wong, G.W. and Goeddel, D.V. (1986) Nature 323, 819-821.
[4] Simon, M.C., Kitchener, K., Kai, H.-T., Hickey, E., Weber, L., Voellmy, R., Heintz, N. and Nevins, J.R. (1987) Mol. Cell. Biol. 7, 2884-2890.
[5] Raj, N.B.K. and Pitha, P.M. (1980) Proc. Natl. Acad. Sci. USA 77, 4918-4922.
[6] Content, J. (1986) in: Molecular Basis of Viral Replication (Perez-Beroff, R. ed.) Plenum Press, New York, in press.
[7] Wathelet, M.G., Clauss, I.M., Nols, C.B., Content, J. and Huez, G.A. (1987) Eur. J. Biochem. 169, 313-321.
[8] Wathelet, M.G., Clauss, I.M., Content, J. and Huez, G.A. (1988) Eur. J. Biochem., in press.
[9] Weil, J., Epstein, C.J., Epstein, L.B., Sedmak, J.J., Sabran, J.L. and Grossberg, S.E. (1983) Nature 301, 437-439.
[10] Beresini, M.H., Lempert, M.J. and Epstein, L.B. (1987) J. Interferon Res. 7, 819.
[11] Larner, A.C., Jonak, G.C., Cheng, Y.S.E., Korant, B., Knight, E. and Darnell, J.E., jr (1984) Proc. Natl. Acad. Sci. USA 81, 6733-6737.
[12] Larner, A.C., Chauduri, A. and Darnell, J.E., jr (1986) J. Biol. Chem. 261, 453-459.
[13] Wathelet, M., Moutschen, S., Defilippi, P., Cravador, A., Collet, M., Huez, G. and Content, J. (1986) Eur. J. Biochem. 155, 11-17.
[14] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443-453.
[15] Sellers, P.H. (1974) SIAM J. Appl. Math. 26, 787-793.
[16] Goad, W.B. and Kanehisa, M.I. (1982) Nucleic Acids Res. 10, 247-263.
[17] Levy, D., Larner, A., Chaudhuri, A., Babiss, L.E. and Darnell, J.E., jr (1986) Proc. Natl. Acad. Sci. USA 83, 8929-8933.
[18] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation 5, s.3, pp. 345-352.
[19] Taniguchi, T., Mantei, N., Schwarstein, M., Nagata, S., Muramatsu, M. and Weissmann, C. (1980) Nature 285, 547-549.
[20] Zilberstein, A., Ruggieri, R., Horn, J.H. and Revel, M. (1986) EMBO J. 5, 2529-2537.
[21] May, L.T., Helfgott, D.C. and Sehgal, P.B. (1986) Proc. Natl. Acad. Sci. USA 83, 8957-8961.
[22] Haegeman, G., Content, J., Volckaert, G., Derynck, R., Tavernier, J. and Fiers, W. (1986) Eur. J. Biochem. 159, 625-632.
[23] Wathelet, M., Moutschen, S., Cravador, A., De Wit, L., Defilippi, P., Huez, G. and Content, J. (1986) FEBS Lett. 196, 113-120.

[24] Hug, H., Staeheli, P., Wehrli, M. and Aebi, M. (1987) Abstr. 19th Ann. Meet. U.S.G.E.B.
[25] Reich, N., Evans, B., Levy, D., Fahey, D., Knight, E., jr and Darnell, J.E., jr (1987) Proc. Natl. Acad. Sci. USA 84, 6394–6398.
[26] Weissmann, C. and Weber, H. (1986) Progr. Nucleic Acid Res. Mol. Biol. 33, 251–300.
[27] Chebath, J., Benech, P., Hovanessian, A., Galabru, J. and Revel, M. (1987) J. Biol. Chem. 262, 3852–3857.
[28] Friedman, R.L., Manly, S.P., McMahon, M., Kerr, I.M. and Stark, G.R. (1984) Cell 38, 745–755.
[29] Eisenberg, D., Weis, R.M., Tenmiller, T.C.S. and Wilcox, W. (1982) Faraday Symp. Chem. Soc. 17, 109–120.