© 1992 Oxford University Press
Nucleic Acids Research, Vol. 20, No. 16 4199–4204

Structure of the gene encoding hepatocyte nuclear factor 1 (HNF1)

Ingolf Bach, Marco Pontoglio and Moshe Yaniv*
Unité des Virus Oncogènes, UA 1149 du CNRS, Département des Biotechnologies, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris cedex 15, France

Received May 28, 1992; Revised and Accepted July 16, 1992 EMBL accession nos X66548–X66577 (incl.)

ABSTRACT

Genomic clones have been isolated that cover the entire gene for the transcription factor HNF1 (hepatocyte nuclear factor 1). This protein governs the expression of many genes, synthesized in the liver in a tissue-specific manner. We have determined the intron/exon structure of the *HNF1* gene, which is strictly conserved between rat and mouse and estimate that it spans not more than 40kb in the rat genome. Whereas most homeoprotein genes do not contain introns within the homeodomain, *HNF1* displays an intron between the regions encoding the second and the third helices. We discuss possible evolutionary mechanisms leading to this homeobox intron/exon pattern.

INTRODUCTION

Hepatocyte Nuclear Factor 1 (HNF1, also called HNF1α, LF-B1, APF or HP-1) is a transcription factor required for the liver-specific expression of a variety of genes including the albumin gene (for review see 1–3). cDNAs encoding the rat, mouse and human HNF1 proteins have been isolated (4–8) and it was shown that the HNF1 amino acid as well as nucleotide sequence is highly conserved among these species (7, 8). In a family of rat hepatoma cell lines HNF1 mRNA expression is restricted to the differentiated phenotype, that is cells, that express liver-specific proteins (9). However, analysis of the tissue distribution of HNF1 mRNA by Northern blotting, RNAse protection and *in situ* hybridization has revealed that the HNF1 transcript is not only found in the liver, but also in kidney, intestine and stomach (5, 8, 10).

The HNF1 protein is able to homodimerize (11, 6) or heterodimerize with variant HNF1 (vHNF1 also named LF-B3, HNF1β or vAPF), a second member of the HNF1 homeoprotein family, for which a cDNA was recently isolated (12–15). The dimerization domain was mapped to a short segment in the N-terminal portion of the protein. A region designated the B-domain, which displays weak homologies to the POU domain, is located between the dimerization and the DNA binding domain (11). HNF1 is thought to contact DNA via the most diverged homeodomain identified so far. This differs from the homeodomain of other homeoproteins in that a loop of 21 additional amino acids is located between the predicted helices 2 and 3 (4, 16, 5–7). An alternative structure with a loop comprising 18aa within the second helix and three extra amino acids between helices 2 and 3 has been proposed as well (11). Finally, the transcriptional activation function of HNF1 is contained in its C-terminal half, a region rich in serine and threonine residues (11, 8, 17).

In this paper we describe the isolation of genomic clones, bearing inserts that contain the rat and mouse *HNF1* genes. We show that the size of the coding part of the rat *HNF1* gene does not exceed 40kb. We have established the intron/exon structure of the *HNF1* gene which is identical in both species. The first exon contains the entire dimerization domain as well as the N-terminal part of the B-domain. Exon 2 contains the main part of the B-domain. The homeodomain is included in exons 3 and 4. The first and second helices are encoded by exon 3, whereas exon 4 contains most or all of the 18–21 amino acid loop, the third helix and the proline and glycine-rich hinge region. The transactivation domain is located in exons 5 to 9.

Comparison of the intron/exon structure of the homeobox region of the *HNF1* gene with those of other homeoprotein genes shows that it does not correspond to any other intron/exon gene structure known to date. The results show that HNF1 is not only organized differently at nucleotide and protein sequence level but also at the level of genomic structure. This provides further evidence of the divergence of the *HNF1* gene in relation to other homeobox genes.

The mechanisms by which so many different homeotic genes have evolved remain speculative, although we favour a model in which intron gain might have played at least some role.

MATERIALS AND METHODS

Screening the genomic libraries

8×10^5 plaques of a rat genomic library in λ charon 4a (Clontech) were screened with a rat cDNA probe (CD26) containing a partial HNF1 sequence missing the first 200 coding nucleotides (6). 5×10^5 plaques of a mouse Balb/c genomic library in λ EMBL3 (Clontech) were screened with the

* To whom correspondence should be addressed

BglII/HindIII fragment of the expression vector pRHP, containing all the coding sequences of rat HNF1 (Sourdive et al., in preparation). The different probes were radioactively labelled by random priming with a Multiprime DNA labelling system kit (Amersham) according to the manufacturers instructions. Both screenings were carried out as previously described (7).

Analysis of positive clones

The plaques of all positive λ clones were first analyzed by hybridizing 5′-end labelled oligonucleotides which contained homologous sequences of the very 5′ end to the very 3′ end of the rat and mouse HNF1 cDNAs. The 5′-endlabelling of the oligonucleotides was carried out as described by Sambrook et al., 1989 (18). The hybridizations were done essentially under conditions described previously (6).

To prepare the DNA of positive phage isolates the liquid culture DNA preparation method according to Sambrook et al. 1989 (18) was applied. A southern blot analysis of this DNA was performed as described elsewhere (18) on the inserts of rat and mouse genomic clones. Several fragments containing sequences corresponding to different regions of the HNF1 cDNA were isolated and subcloned into Bluescribe.

The HNF1 encoding regions of the subcloned inserts were sequenced directly from double stranded plasmid DNA using the Sequenase DNA sequencing kit (USB) and successive oligonucleotides as primers.

Elucidating the intron/exon structure

The inserts of the different phage clones were subcloned into Bluescribe and sequenced 5′ to 3′ with primers corresponding to cDNA sequences of rat or mouse HNF1. Introns were indicated when the genomic *HNF1* sequence differed from that of the HNF1 cDNA. Each intron was confirmed and its borders defined by sequencing from downstream of the intron boundary in the opposite direction.

RESULTS AND DISCUSSION

The entire coding sequences for the rat and mouse HNF1 protein are contained on inserts of the isolated λ clones

To analyse the structure of the *HNF1* gene we screened rat and mouse genomic libraries. From a rat genomic library, 32 positive clones were detected and isolated. Four overlapping genomic clones RGL 5, 14, 25 and 26 with insert lengths of 8, 18, 15 and 9kb, respectively were chosen for further analysis (see Fig. 1D). Since the inserts of the three partially overlapping clones RGL 14, 25, and 5, hybridized to all oligonucleotides containing the HNF1 cDNA sequence, the rat *HNF1* gene cannot be larger than 40kb; the sum of the insert lengths of the three clones.

10 positive clones were obtained by screening a Balb/c mouse genomic library. Of those, the 3 clones H11, H2 and H6 (with inserts of 17kb, 15kb and 17kb, respectively) were chosen for

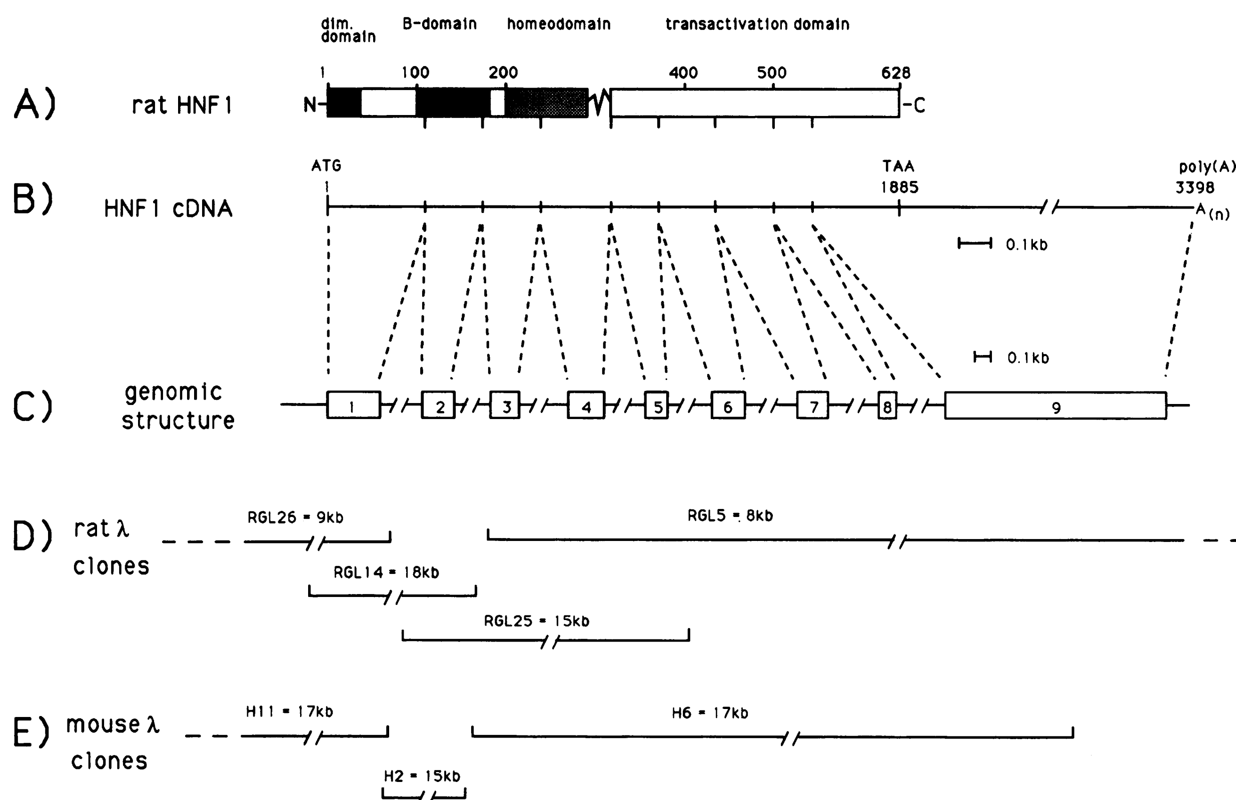

Fig. 1. The *HNF1* gene structure. A) Diagram of the protein structure. The numbering corresponds to the number of amino acids. B) The rat HNF1 cDNA drawn to the same scale as that used for part A. Numbers indicate nucleotides. The vertical bars along the sequence in A and B indicate the position of introns. C) Genomic structure of the rat and mouse *HNF1* genes. The exons are boxed and numbered; the size of the introns were not determined and are not drawn to scale. The dotted lines illustrate the location of each exon on the cDNA. D) Inserts of the isolated λ clones of rat *HNF1* covering the entire gene locus. Clone names and sizes of inserts are indicated and are not drawn to scale. E) Inserts of the isolated λ clones of mouse *Hnf-1* covering all the exonic regions of the protein coding part of the gene. Sizes of inserts are not drawn to scale.

further analysis. The inserts of the three clones cover all the coding sequences of the mouse *Hnf-1* gene (Fig. 1E) but since the insert of clone H2 does not overlap with that of clone H6, we cannot estimate the total size of the mouse *Hnf-1* gene.

The *HNF1* gene consists of 9 exons and 8 introns

To elucidate the intron/exon distribution in the rat and mouse *HNF1* genes we determined the sequence of the entire coding and 5′ and 3′ non-coding regions on the genomic clones corresponding to the HNF1 cDNA sequence.

The genomic clones of the rat *HNF1* gene: With an insert size of 8kb, clone RGL 5 starts in the second intron which is located at a region on the HNF1 cDNA corresponding to the B-domain. This clone covers all the gene sequences from exon 3 to the polyadenylation site. Exon 2, 3, 4 and 5 are contained within the 15kb insert sequences of clone RGL25. This clone shares exons 3 to 5 with RGL 5. Clone RGL 14 contains an insert of about 18kb and covers the gene sequences from 168bp of the 5′ non-coding region until intron 2, thus sharing exon 2 with clone RGL 25. The 9kb insert of clone RGL 26 starts in the 5′ flanking sequence and ends in the first intron, sharing exon 1 with clone RGL 14.

The genomic clones of the mouse *Hnf-1* gene: The insert of clone H6 is 17kb in size and covers the gene region from intron 2 to the 3′ non-coding region, thus containing exons 3−8 and all the coding part of exon 9. The 15kb insert of clone H2 contains exon 2. Finally, clone H11 contains 12kb of 5′ non-coding gene sequences, exon 1 and overlaps for 500bp with clone H2. The overall insert length of this clone is 17kb.

Our data demonstrate that the *HNF1* gene of rat and mouse is composed of 9 exons and 8 introns as illustrated in Fig. 1. In the 3′ non-coding regions no intron could be found. Thus, all the introns are located within the protein coding region.

700bp of the 5′ flanking region was sequenced in the rat and 350bp in the mouse gene comprising putative *HNF1* promoter sequences. Around 240 nucleotides of 5′ non-coding sequences can be found between the ATG translational start site and a putative TATA-Box for the *HNF1* promoter in the genes of both species. The assumption that no further intron is located in the 5′ untranslated regions is supported by three observations: 1. It would fit well with results obtained in Northern Blot assays that the HNF1 messenger is 3.5−3.6kb in length (9). 2. cDNAs encoding human HNF1 have been isolated which contain sequences from the 5′ non-coding region highly homologous to the rat genomic sequences up to 40bp downstream of the putative TATA-Box (I.B., unpublished). 3. The transcriptional start site of the rat and mouse *HNF1* genes were mapped and shown to be located between 23−30bp downstream of this TATA-box (19; M.P., unpublished).

Although not determined, a large intron size for one or both of the first two introns of at least the rat *HNF1* gene is suggested by comparing the insert lengths of rat clones RGL 14 (18kb covering 2 exons and 1 entire intron) and RGL 25 (15kb spanning over 4 exons and 3 introns) with that of clone RGL 5 (8kb covering 6 exons and 5 introns).

The occurrence of an intron (intron 2) in the *HNF1* gene at exactly the point where vHNF1-A contains the extra 26 aa (78bp) compared to vHNF1-B (12; I.B., unpublished results; Fig. 1 and 2) suggests that the two genes may be organized in a very similar manner.

	intron sequences	exon sequences	intron sequences	intron positions
rat	5'end	-EXON 1-	CTTCAgtaaggtacc BamHI	1 109 -2
mouse	5'end	-EXON 1-	CTTCAgtaaggaccccaacctgtccccagcaccca	1 109 -2
rat	ttgaacagaccctgtccttgctccctccagGGAGG	-EXON 2-	TCAGCgtaagtaacgaccccagcctgtgtcctcct	2 176 -1
mouse	gacccggtcattgctcccctgctcccccagGGAGG	-EXON 2-	TCAGCgtaagtaacgaccccagcctgggtcctcct	2 176 -1
rat	EcoRI gAATTC	-EXON 3-	AATAGgtagcgcggtggtcccgatcagggcgaggt	3 238 -2
mouse	EcoRI gAATTC	-EXON 3-	AATAGgttagtgtggtcgcccatcagggaggctaa	3 238 -2
rat	gcctggagctcacgggtgccatctttcgagGGCGG	-EXON 4-	CCACGgtgagtcctgtgggtagggaagctgggata	4 319 -1
mouse	ctggaggctcacggggcgccatctttgcagGGCAG	-EXON 4-	CCACGgtgagtgctgtggggtagggaactgggata	4 319 -1
rat	agtggggtgctgacactgctcccaccctagGTGTG	-EXON 5-	AGCTGgtgagtgagcggccctgtcctgtgtagaaa	5 369 -0
mouse	gtggggtgctgacattgctcctcactctagGTGTA	-EXON 5-	AGCTGgtgagtgagcaggggccccctgtcctgtac	5 369 -0
rat	acaccagctcagcttggtttcccttcacagGTCTC	-EXON 6-	TATTGgtaagtgggtggggtgactggagctcctgt	6 436 -1
mouse	actccagctcaacttggtttaccctcacagGTCTC	-EXON 6-	TATCGgtaagtgggtggcatgagtggggtcctgtg	6 436 -1
rat	ggcttggctactcctgcctgtcctccacagGTCTG	-EXON 7-	CCACGgtgagcaatcttcacattggtctgggagag	7 500 -1
mouse	ggctcagctgactctgcctgttctgcccagGTCTG	-EXON 7-	CCACGgtgagcatcccccacgttggcctgggagag	7 500 -1
rat	gtcacagctatctgcaccttctcctcccagCCCTG	-EXON 8-	AGCAGgtaaggccctgcctcctgaagcccggcccc	8 540 -0
mouse	agggagtcagagctctgtgtctcctcccagCCTTA	-EXON 8-	AGCAGgtaaggccctgcctcctgaagccccacccc	8 540 -0
rat	ggctcaggggggttattctgctttacctagGTCTT	-EXON 9-	ACAGTactagtctgggtcacctgataatttatttc	
mouse	ggctcagtggggggttattctgttccctagGTCTT	-EXON 9-		

Fig. 2. Nucleotide sequence of intron/exon junctions in the rat and mouse *HNF1* genes. The exon sequences are shown in upper case whereas intron sequences in lower case letters. Sequence common to the 3′ end of the HNF1 message (excluding poly A tail) is typed in upper case standard letters. The genomic sequence downstream of the message is typed in lower case standard letters. BamHI and EcoRI restriction sites used for subcloning genomic fragments are underlined in the first and second intron/exon junction, respectively. The intron numbers and positions are indicated to the right. The positions are described by both codon and phase. Codons followed by −1 or −2 are split by an intron after the first and second base, respectively. A codon followed by −0 has an intron between it and the previous codon.

The rat and mouse *HNF1* gene structures are strictly conserved

The sequences of the intron/exon junctions in the rat and mouse *HNF1* gene are shown in Fig. 2. The sequence of the first 30 nucleotides of most introns was determined. Restriction sites which were used for subcloning genomic fragments are underlined. Intron sequences were not determined beyond these sites. The intron numbers are indicated and their position relative to the cDNA and protein sequence is given by both amino acid number and translation phase. Our sequence analysis shows that the numbers, positions and translation phases were conserved between the *HNF1* genes of rat and mouse. The dinucleotides GT at the 5′ and AG at the 3′ intron boundaries, respectively, are conserved for all introns investigated. In addition pyrimidine-rich sequences which immediately precede the 3′ AG boundaries are present in all introns which were sequenced. These observations match well with the previously deduced consensus sequences for donor and acceptor sites (20).

The results show that HNF1 is not only highly conserved between different species at nucleotide and amino acid sequence levels but also at the level of genomic structure. Intronic gene sequences closer to intron/exon junctions are in general more conserved than those which are more distant. Nevertheless, this conservation is lower than that of the exon sequences (see Fig. 2; 8).

It is noteworthy, that the overall intron/exon pattern does not really fit a functional pattern in that functional modules of the protein are not encoded by single exons (see also Fig. 1 and 3). The only domains to which unique functions can be attributed and that are located on a single exon are the dimerization domain at the very N-terminus (exon 1) and a region important for efficient transactivation (17; D. Sourdive et al., manuscript in preparation) that contains 26% of the serine residues in the C-terminal part of the protein (exon 9).

The homeodomain of the *HNF1* gene shows an unusual intron/exon pattern

The intron/exon structure in the region encoding the homeodomain is quite conserved amongst homeoproteins: there are many examples of genes in which the homeodomain is encoded by a single exon e.g. the genes for antennapedia, Ubx, Scr in Drosophila or most vertebrate *Hox* genes (21–25, see also Fig. 3). In almost all of these genes an intron is located immediately upstream of the homeodomain. Other homeodomain encoding genes like *AbdB, Lab, Evx1, 2, Xhox-3* and *goosecoid* display an intron at the beginning of the third helix of their homeodomains (26–30).

The homeodomain of the rat and mouse *HNF1* genes contain an intron at an unusual location: exactly between the second helix and the 21 amino acid loop (loop A) or within the 18 amino acids of the alternative structure loop B (see Fig. 3). Both possible intron locations have not been previously observed in any other homeodomain containing genes, emphasizing the distance between HNF1 and the other homeoproteins. The loop A alternative has two advantages: Helix 2 remains intact and since the intron boundary coincides with the beginning of the loop,

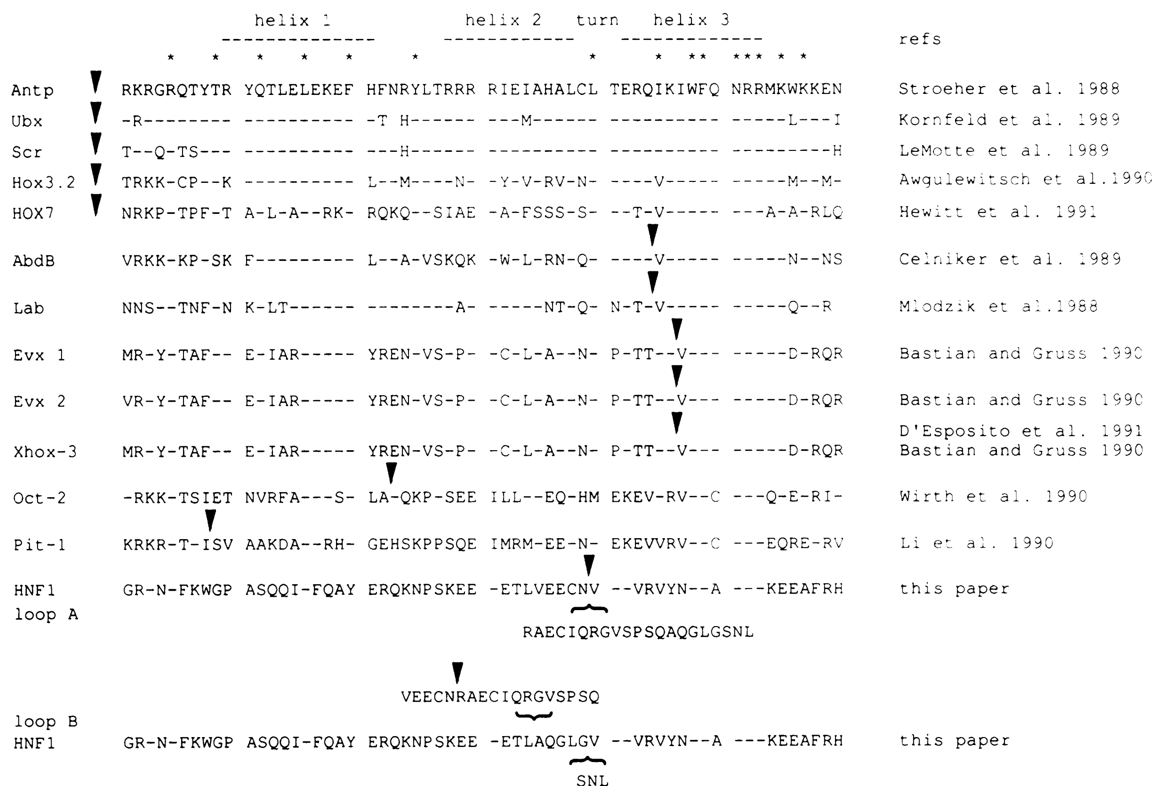

Fig. 3. Intron/exon homeodomain gene structure of different homeoproteins including rat and mouse HNF1. Secondary structure predictions are indicated on top. An asterisk marks residues conserved in most homeodomains and the flash indicates the location of introns. References are indicated on the right. The extra 21 amino acids of the HNF1 homeodomain are listed below (loop A), whereas the 18 amino acid loop of the alternative structure (loop B) is listed above the HNF1 homeodomain.

one can mechanistically explain the occurrence of such a structure more easily e.g. by the creation of a new acceptor site inside a primitive intron.

Unusual intron/exon structures are also found in the cases of the POU proteins *Oct-2* and *Pit-1* genes. These homeoboxes are interrupted by an intron upstream of helix 2 and 1, respectively (31, 32).

Is the homeobox organized as a structural module suitable for exon-shuffling?

An interesting theory for the importance of discontinous gene structure in the evolution of eucaryotic genomes, the so-called 'exon shuffling theory' has been proposed. According to this theory introns would provide non-coding targets for recombination between different genes which would then give rise to novel exon combinations (33; review: 34). Indeed, a large number of membrane and secretory proteins were found to share extracellular domains. These domains are encoded by single exons (for review see 35). This theory has been proposed to be more general in that exons would originally encode stable protein units as modules which could be set together in different combinations to form a whole variety of proteins (36).

The homeodomains of different homeoproteins are thought to have evolved from a common ancestral domain. We wondered whether the homeodomain could be such a module shuffled during evolution to generate the genes of several dozens of homeoproteins and how this could be compatible with the data summarized in Fig. 3.

As mentioned above, the homeodomains of most proteins are located on a single exon. Nevertheless, there are some examples of genes in which the homeodomain encoding sequence is located on two exons, separated by an intron at different positions (see Fig. 3). Assuming that all homeodomains evolved from a common ancestor gene this observation leads to several speculative ways, which are not mutually exclusive, of how homeodomains could have spread and of intron occurrence in this domain: One possibility would be that the common ancestor gene already contained introns at all the different positions which were subsequently partially or completely lost during evolution. According to this model the original homeobox contained at least 4 different introns which is somewhat unlikely considering its size (60aa). Some introns must have been separated by only a few nucleotides. In addition this model would exclude the possibility, that the homeodomain could have been shuffled on a single exon, as has been the case for many extracellular protein domains (35, 37). It is evident that shuffling of this domain as a whole on several exons is still possible but recombination probabilities are lower.

Another possibility would be that the ancestor homeobox-containing gene did not display any intron and that the introns were inserted later during evolution into a founder of one branch of homeogenes but not into other ancestral subfamilies. This hypothesis implies among other mechanisms the possibility of exon shuffling of the homeodomain on a single exon before the gain of introns. Or a combinatorial alternative would be feasable: The original homeodomain contained one or several introns and then during evolution some introns were lost or gained, depending on the gene family. Exon shuffling could have happened from all of those genes but with different probabilities according to the numbers and sizes of the introns within the homeodomain.

Besides the occurence of introns in some homeodomains, several other observations favour a model where intron gain played at least some role. The characterisic features of the HNF1 protein family are a stretch of 21 extra amino acids between helices 2 and 3 of the homeodomain, and a proline/glycine-rich region which separates the HNF1 protein into two parts. These HNF1-typical structures share the same exon (exon 4) as the most conserved part of all the different homeodomains, namely the third helix. An easy way to explain this finding would be that the loop was introduced into this conserved domain, perhaps as the flanking sequences of a gained intron. This argument is strengthened by the observation that exactly at the point where the 21 extra amino acids start, an intron can be found (see Fig. 3). Evidence for intron insertions has been presented for a variety of genes including serine proteases (38), collagen (39) and actin and tubulin genes (40). Likewise, although speculative, it appears to be more probable that all or most introns which exist in the genes of certain homeoprotein families were inserted later. However, it seems to be clear from the data summarized in Fig. 3, that intron fluidity and the evolution of intronic gene regions is much faster than the evolution of conserved protein structures such as homeodomains.

ACKNOWLEDGEMENTS

We wish to thank J.Ham for critically reading the manuscript and S.Cereghini for valuable discussions. This work was supported by grants from the CEE BAP (0117F), INSERM, ARC, LNFCC and FRMF to M.Y. I.B. is a Boehringer Ingelheim Fonds fellow, M.P. is supported by a fellowship from the Istituto Scientifico S.Raffaele and was also recipient of a Anna Villa Rusconi fellowship.

REFERENCES

1. De Simone, V., and Cortese, R. (1991). Curr. Opinion Cell Biol. 3: 960–965.
2. Mendel D.B., and Crabtree, G.R. (1991). J. Biol. Chem. 266: 677–680.
3. Rey-Campos, J., and Yaniv, M. (1992). Molecular Approaches to the Study and Treatment of Human Diseases. T.O. Yoshida ed., Elsevier, 239–262.
4. Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R., and Cortese, R. (1989). Cell 59: 145–157.
5. Baumhueter, S., Mendel, D.B., Conley, P.B., Kuo, C.J., Turk, C., Graves, M.K., Edwards, C.A., Courtois, C., and Crabtree, G.R. (1990). Genes & Dev. 4: 372–379.
6. Chouard, T., Blumenfeld, M., Bach, I., Vandekerckhove, J., Cereghini, S., and Yaniv, M. (1990). Nucl. Acids Res. 18: 5853–5863.
7. Bach, I., Galcheva-Gargova, Z., Mattei, M.-G., Simon-Chazottes, D., Guénet, J.-L., Cereghini, S. and Yaniv, M. (1990). Genomics, 8: 155–164.
8. Kuo, C.J., Conley, P.B., Hsieh, C.-L., Francke, U. and Crabtree, G.R. (1990). Proc. Natl. Acad. Sci. USA 87: 9838–9842.
9. Cereghini, S., Yaniv, M., and Cortese, R. (1990). EMBO J. 9: 2257–2263.
10. Blumenfeld, M., Maury, M., Chouard, T., Yaniv, M., and Condamine, H. (1991). Development 113: 589–599.
11. Nicosia, A., Monaci, P., Tomei, L., DeFrancesco, R., Nuzzo, M., Stunnenberg, H., and Cortese, R. (1990). Cell 61: 1225–1236.
12. Rey-Campos, J., Chouard, T., Yaniv, M. and Cereghini, S.(1991). EMBO J. 10: 1445–1457.
13. De Simone, V., De Magistris, L., Lazzaro, D., Gerstner, J., Monaci, P., Nicosia, A., and Cortese, R. (1991). EMBO J. 10: 1435–1443.
14. Mendel, D.B., Hansen, L.P., Graves, Conley, P.B., and Crabtree, G.R. (1991). Genes & Dev. 5: 1042–1056.
15. Bach, I., Mattei, M.-G., Cereghini, S., and Yaniv, M. (1991). Nucl. Acids Res. 19: 3553–3559.
16. Finney, M. (1990). Cell 60: 5–6.
17. Raney, A.K., Easton, A.J., Milich, D.R., and McLachlan, A. (1991). J. Virology 65: 5774–5781.
18. Sambrook, J, Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbour, New York; Cold Spring Harbour Laboratory.

19. Tian, J.-M., and Schibler, U. (1991). Genes & Dev. 5: 2225–2234.
20. Breathnach, R., and Chambon, P. (1981). Ann. Rev. Biochem. 50: 349–383.
21. Stroeher, V.L., Gaiser, J.C., and Garber, R.L. (1988). Mol. Cell. Biol. 8: 4143- 4154.
22. Kornfeld, K., Saint, R.B., Beachy, P.A., Harte, P.J., Peattie, D.A., and Hogness, D.S. (1989). Genes & Dev. 3: 243–258.
23. LeMotte, P.K., Kuroiwa, A., Fessler, L.I., and Gehring, W.J. (1989). EMBO J. 8: 219–227.
24. Awgulewitsch, A., Bieberich, C., Bogarad, L., Shashikant, C., and Ruddle, F. (1990). Proc. Natl. Acad. Sci. USA 87: 6428–6432.
25. Hewitt, J.E., Lorraine, C.N., Alasdair, I., and Williamson, R. (1991). Genomics 11: 670–678.
26. Celniker, S.E., Keelan, D.J., and Lewis, E.B. (1989). Genes & Dev. 3: 1424–1436.
27. Mlodzik, M., Fjose, A., and Gehring, W.J. (1988). EMBO J. 7: 2569–2578.
28. Bastian, H., and Gruss, P. (1990). EMBO J. 9: 1839–1852.
29. D'Esposito, M., Morelli, F., Acampora, D., Migliaccio, E., Simeone, A., and Boncinelli, E. (1991). Genomics 10: 43–50.
30. Blum, M., Gaunt, S.J., Cho, K.W.Y., Steinbeisser, H., Blumberg, B., Bittner, D., and De Robertis, E.M. (1992). Cell 69: 1097–1106.
31. Wirth, T., Priess, A., Annweiler, A., Zwilling, S., and Oeler, B. (1991). Nucl. Acids Res. 19: 43–51.
32. Li, S., Crenshaw III, E.B., Rawson, E.J., Simmons, D.M., Swanson, L.W., and Rosenfeld, M.G. (1990). Nature 347: 528–533.
33. Gilbert, W. (1978). Nature 271: 501.
34. Rogers, J.H. (1990). FEBS lett. 268: 339–343.
35. Doolittle, R. (1985). Trends Biochem. Sci. 10: 233–237.
36. Gilbert, W., Marchionni, M., and McKnight, G. (1986). Cell 46: 151–154.
37. Patthy, L. (1990). Cell 61: 13–14.
38. Rogers, J. (1985). Nature 315: 458–459.
39. Rogers, J. H. (1986). Trends Genet. 2:223.
40. Dibb, N.J., and Newman, A.J. (1989). EMBO J. 8: 2015–2021.