FEBS Letters 410 (1997) 124–125 FEBS 18780

Hypothesis

A transcriptional modification motif encoded by homeobox and fork head genes

Sebastian M. Shimeld*

School of Animal and Microbial Sciences, The University of Reading, Whiteknights, Reading RG6 2AJ, UK

Received 18 April 1997; revised version received 13 May 1997

Abstract Homeodomain and fork head domain proteins are thought to act as transcription factors by binding to specific DNA target sequences and interacting with other proteins. Here I describe a motif which is present in members of both groups of transcription factor and which has been shown to modulate their ability to activate transcription. The presence of this motif in both homeodomain and fork head proteins indicates they may control transcription by a similar molecular mechanism, perhaps by interacting with the same cofactors.

Key words: Homeobox gene; Homeodomain; Fork head; HNF-3; Protein motif

1. Introduction

Most transcription factor families are defined by possession of particular DNA-binding domains. However many transcription factors also share other motifs, for example the hexapeptide of many homeodomain proteins which is believed to act as a binding site for Pbx proteins [1]. Such shared motifs can generally be interpreted as conservation of an ancestral motif, implying common descent of the genes by which they are encoded.

Here I describe an example of a functional motif in two unrelated transcription factor families: (a) the fork head domain-containing HNF-3 genes and (b) the homeobox genes of the en, Msx, NK-1, NK-2 and goosecoid classes. In vertebrates, the HNF-3 genes are expressed in the embryonic organiser and subsequently in midline cells of the ectoderm, mesoderm and endoderm, tissues with essential roles in the formation of the vertebrate body plan. Over-expression and targeted disruption of some of these genes has shown them to have a pivotal role in the specification of these tissues. The homeobox genes of the classes mentioned above include at least 10 members in mice and have been implicated in a wide range of developmental processes, including gastrulation, neurogenesis, central nervous system patterning and craniofacial morphogenesis. Because of these important roles, both groups of transcription factor have been the subject of extensive research since their original isolation.

2. Homeodomain and fork head proteins share a motif involved in transcription modification

Recently Smith and Jaynes (1996) [2] reported the identification of a conserved region (eh1) in all en, Msx, NK-1, NK-2 and gsc class homeodomain proteins. Analysis of eh1 from the *Drosophila* engrailed protein showed it to be capable of strongly repressing transcription when targeted in vivo by attachment to the DNA binding domain of an activator. Here I report that this region also shares sequence identity with a conserved motif present in HNF-3 class genes from vertebrates, protochordates and arthropods. HNF-3 genes contain a number of conserved motifs, including one previously designated region II which has been shown to activate transcription [3]. Fig. 1 shows the consensus sequence identified by Smith and Jaynes (1996) [2] aligned with that derived from region II of twelve HNF-3 class genes. The HNF-3 consensus sequence is centred around five invariant amino acids with the sequence HPFSI. Four of these are found in the eh1 consensus, including the central phenylalanine which

Accession	Gene	Alignment
J03177	Dm Fkh	S S **H P F S I** N R L L P
D38514	Bm SGF-1	A Q **H P F S I** T R L L P
X96519	amHNF-3-1	F T **H P F S I** S N L M S
Y09236	amHNF-3-2	F H **H P F S I** N S I I S
Z22762	Zf axial	F N **H P F S I** N N L M S
X65171	X pintallavis	G R **H P F S I** T Q L M S
L25637	X HNF3Beta	F N **H P F S I** N N L M S
M93658	X XFKH2	F N **H P F S I** N N L M S
L38904	C HNF3Beta	F N **H P F S I** N N L M S
X74936	M HNF3Alpha	F N **H P F S I** N N L M S
L10409	M HNF3Beta	F N **H P F S I** N N L M S
X74938	M HNF3Gamma	F N **H P F S I** N N L M S
	HNF-3 consensus	F N **H P F S I** N N L M S
	eh1 consensus	R A L P F S I D N I L S

Fig. 1. Sequence comparison of region II from 12 HNF-3 type genes. The five invariant residues are shown in bold. Underneath is a comparison of the consensus sequence derived from these proteins and the core of that identified in various homeodomain proteins by Smith and Jaynes (1996) [2]. Identical sites are boxed. To determine the significance of the identity between the HNF-3 and eh1 consensus sequences, the TREMBL database (release 17.2.97) was searched with FASTA [4] (version 3.0, ktup = 2, BLOSUM50 matrix) using the core eh1 consensus sequence of 12 amino acids. This identified 16 non-redundant hits, of which 9 were homeodomain proteins. Of the remaining hits, 43% (3/7) were HNF-3 proteins; however the highest E-value was 7.2, below the level of statistical significance. Similar searches using 12 amino acids stretches adjacent to the eh1 consensus of the *MshD* homeodomain gene identified between 0 and 18 non-redundant hits (average 11, $n = 4$). None of these were HNF-3 proteins. Abbreviations: Dm Fkh, *Drosophila melanogaster* fork head. Bm Sgf1, *Bombyx mori* silk gland factor 1. am, amphioxus. Zf, zebrafish. X, *Xenopus*. C, chick. M, mouse. For sources of sequences and a full alignment of HNF-3 proteins, see Shimeld (1997) [5].

---

*Fax: (44-1189) 316644.
E-mail: sshimeld@hgmp.mrc.ac.uk

PII S0014-5793(97)00632-7

is invariant in all the homeodomain proteins identified by Smith and Jaynes (1996) [2]. Statistical comparison between eh1 and HNF-3 sequences provides weak evidence for the significance of this identity (see Fig. 1 legend). There are, however, two additional factors which support its biological relevance: Firstly, in each type of transcription factor the motif shows evolutionary conservation across multiple proteins from widely separated taxa. This implies selective pressure in each type to maintain the motif sequence, suggesting it is not simply one of the chance similarities commonly encountered when comparing a short sequence with a large database. Secondly, the motif shows functional similarity, since it has been shown to be involved in transcriptional modification in both groups of transcription factor.

3. Conservation of an ancestral domain or convergent evolution?

Examination of published homeodomain and fork head protein sequences reveals many other members of these families which do not contain this motif. This suggests that the motif is convergently evolved, although it is also possible that it derives from an ancient recombination between ancestral fork head and homeobox genes, followed by subsequent loss in the subclasses where it is absent. The reason for this sequence identity, be it convergent or conserved, is not known, however the functional similarity provides a possible explanation: The sequence identity may reflect a similar molecular mechanism of transcriptional modification, perhaps by interaction with the same cofactor or group of cofactors.

References

[1] Mann, R.S.and Chan, S.K. (1996) Trends Genet. 12, 258–262.
[2] Smith, S.T. and Jaynes, J.B. (1996) Development 122, 3141–3150.
[3] Pani, L., Overdier, D.G., Porcella, A., Qian, X., Lai, E. and Costa, R.H. (1992) Mol. Cell. Biol. 12, 3723–3732.
[4] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[5] Shimeld, S.M. (1997) Dev. Biol. 183, 74–85.