The EMBO Journal vol.11 no.11 pp.4119-4129, 1992

A POU-A related region dictates DNA binding specificity of LFB1/HNF1 by orienting the two XL-homeodomains in the dimer

Licia Tomei, Riccardo Cortese and Raffaele De Francesco[1]

Istituto di Ricerche di Biologia Molecolare P.Angeletti, Via Pontina Km 30 600, 00040 Pomezia, Roma, Italy
[1]Corresponding author

Communicated by R.Cortese

LFB1/HNF1 regulates the hepatocyte-specific transcription of several genes, binding as a dimer to *cis*-acting elements that match the inverted palindrome GTTAATNATTAAC. The DNA binding domain of LFB1/HNF1 is characterized by a unique tripartite structure that includes an unusually long homeodomain (domain C), a region related to the POU-specific A-box (domain B) and a short N-terminal dimerization domain (domain A). We report that a recombinant peptide corresponding to the isolated homeodomain of LFB1/HNF1 binds as a monomer to a half-palindrome binding site, but shows diminished sequence specificity. Domain B, in addition to the homeodomain, is required and sufficient for proper recognition of LFB1/HNF1-responsive sites. A protein consisting of only these latter two domains is a monomer in solution, but forms dimers upon DNA binding. The protein-protein contacts established within the bound dimer restrain the orientation of the two homeodomains with respect to one another, thus contributing in a critical fashion to the recognition of the dyad symmetry-related LFB1/HNF1 sites. The DNA-independent dimerization domain (domain A) is required to increase the affinity of DNA binding, but does not influence the dimer geometry.

Key words: dimerization/DNA binding specificity/homeodomain/LFB1-HNF1/POU/transcription

Introduction

The machinery responsible for controlling the transcriptional activity of different genes gives cells the means to adjust the pattern of gene expression as a response to developmental and environmental stimuli. Through the interaction with specific short sequences located in the promoter and enhancer regions of a given gene, sequence-specific DNA binding proteins provide a fundamental mechanism for the regulation of transcription. The importance of these proteins in the regulation of gene expression has resulted in a concentrated effort to understand how these proteins locate and interact with their DNA binding sites.

LFB1/HNF1 is a DNA binding protein that has been implicated as a major determinant of hepatocyte-specific transcription of several genes. To date, LFB1/HNF1 has been shown to interact with defined sequences within the promoter region of 12 liver-specific genes in species ranging from *Xenopus* to humans (for a review see Mendel and Crabtree, 1991). The consensus binding site derived from these sequences is the palindrome GTTAAT(N)ATTAAC. As anticipated from the dyad symmetry of this site, LFB1/HNF1 binds to DNA as a dimer (Frain *et al.*, 1989). In addition, it has been shown that LFB1/HNF1 exists in solution as a dimer independently of the presence of DNA binding sites and that the dimer produced *in vitro* can exchange subunits at a rather rapid rate (Nicosia *et al.*, 1990).

The functional domains of LFB1/HNF1 were dissected by site-directed mutagenesis (Chouard *et al.*, 1990; Nicosia *et al.*, 1990): the residues required for the transcriptional activity of the molecule are located in the C-terminal part (amino acids 282-628), whereas the DNA binding activity maps in the first 281 amino acids.

Within the DNA binding domain of LFB1/HNF1, three indispensable regions have been identified, namely A, B and C (Nicosia *et al.*, 1990). Of these three regions, region A (amino acids 1-32) has been shown to be necessary and sufficient to bring about dimerization of the protein through a novel α-helical structure (De Francesco *et al.*, 1991; Pastore *et al.*, 1991). Region B (amino acids 100-184) and region C (amino acids 198-281) show limited homology respectively to the POU-A box and to the POU-homeodomains of POU proteins (Herr *et al.*, 1988). The putative homeodomain has an insertion of 21 amino acids between helix II and helix III, as compared with the canonical homeobox (Finney, 1990). For this reason this region of LFB1/HNF1 has also been referred to as the XL (extra large) homeobox (Nicosia *et al.*, 1990).

In this paper we address the question of how the different regions of LFB1/HNF1 contribute to high affinity and sequence-specific DNA binding and discuss the relevance of our results to the question of how homeobox-containing proteins in general, and POU proteins in particular, locate and interact with their DNA binding sites.

Results

The XL-homeodomain of LFB1/HNF1 binds to DNA as a monomer but shows relaxed specificity

Two polypeptides, one encompassing the complete DNA binding domain of LFB1/HNF1 as defined by Nicosia *et al.* (1990) (residues 1-281; BD) and the other corresponding to the region of homology to the homeodomain (residues 195-287; XL-HOM), were expressed in *Escherichia coli* and purified to homogeneity (Figure 1).

The DNA binding properties of the two polypeptides were initially examined using the imperfectly palindromic LFB1/HNF1 binding site contained within the promoter of the α1-antitrypsin gene (α1-AT; Monaci *et al.*, 1988). The relative affinities of the two polypeptides for this site were measured by gel retardation titrations in which a constant limiting amount of radiolabelled duplex oligonucleotide was

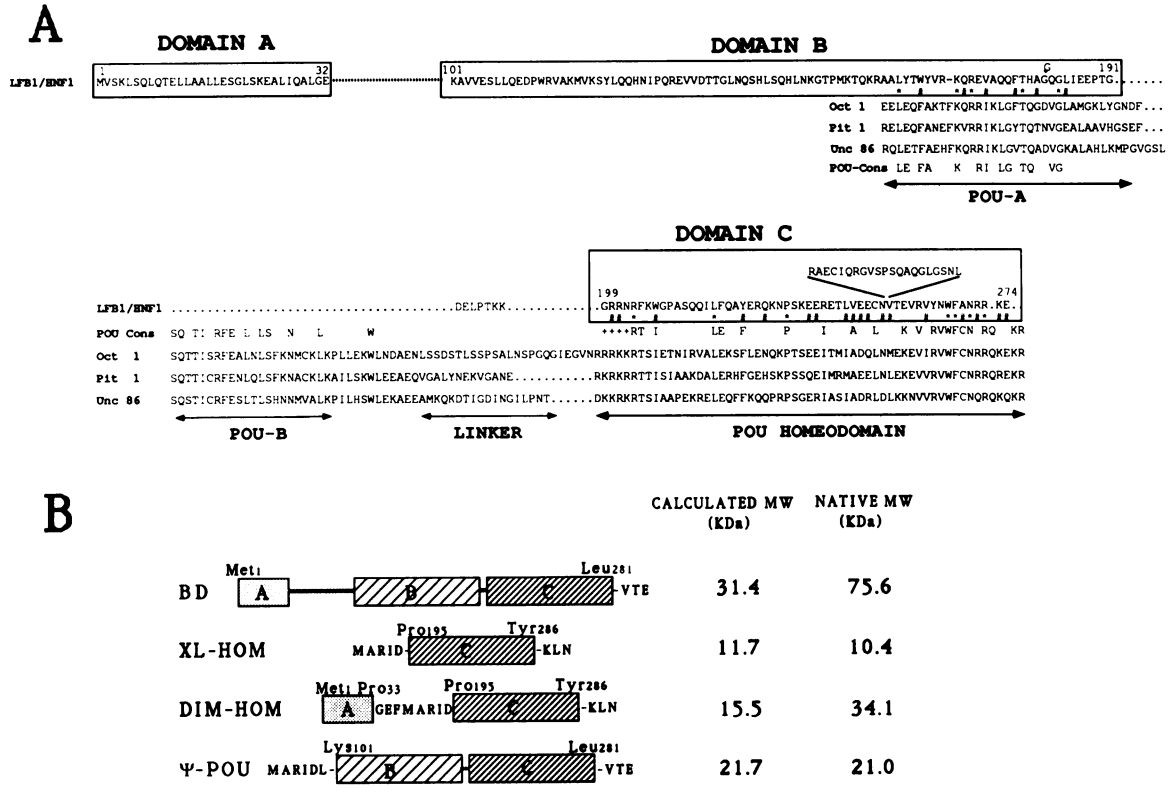

Fig. 1. Domain structure of the DNA binding portion of LFB1/HNF1 and the mutants used in this study. (**A**) Functional domains of the DNA binding portion of LFB1/HNF1 and alignment with representative POU proteins. * indicates residues that are identical between the POU protein family and LFB1/HNF1, while # indicates conservative substitutions or residues that are conserved between LFB1/HNF1 and at least one member of the POU family. (**B**) Diagram of the constructs that express the LFB1/HNF1 mutants in *E.coli*. Hatched and shaded boxes depict the functional domains described in panel A. Amino acid residues that are not present in the wild-type protein are indicated in the one-letter code. The calculated molecular weights were deduced from the cDNA sequence of the various mutants, whereas the native molecular weights were estimated by analytical gel-filtration chromatography on a Pharmacia Superdex 75 HR 10/30 column (see Materials and methods), assuming that all of the mutant proteins are globular. The elution volume (V_R) of the proteins used as molecular weight markers were as follows: bovine serum albumin, 9.05 ml; ovalbumin, 9.95 ml; carbonic anhydrase, 11.35 ml; myoglobin, 12.30 ml; cytochrome *c*, 13.5 ml; oxidized B-chain of bovine insulin, 15.60 ml. All of the LFB1 purified mutant proteins eluted as a single peak with V_R values of 8.90 ml (BD), 13.40 ml (XL-HOM), 10.70 ml (DIM-HOM) and 11.80 ml (Ψ-POU).

titrated with increasing amounts of protein in the absence of competitor DNA.

The binding saturation curves obtained in this way for the two polypeptides are compared in Figure 2A: the first important observation is that the isolated XL-homeodomain of LFB1/HNF1 retained substantial DNA binding activity, with a K_d for the α1-AT site of $\sim 2 \times 10^{-9}$ M. This value is comparable to that of the *Antennapedia* homeodomain for its cognate binding site (Affolter *et al.*, 1990). The polypeptide encompassing the entire DNA binding domain (BD) was predominantly dimeric in solution (Nicosia *et al.*, 1990 and Figure 1B) and bound as a dimer to the α1-AT site with an apparent affinity that was more than one order of magnitude higher than that of the isolated XL-homeodomain: half saturation of the DNA probe by this polypeptide occurred at a protein concentration of $\sim 10^{-10}$ M.

Cooperativity analysis of the curves shown in Figure 2A (not presented) reveals important differences between the mode of binding to DNA of the two proteins. The degree of cooperativity shown by the BD protein on the α1-AT site (Hill slope = 1.5) suggests that an equilibrium exists between the monomeric and the dimeric forms of the protein in the lower concentration range ($< 10^{-9}$ M), the dimer being the species competent for high-affinity DNA binding. This behaviour was expected on the basis of our previous study (Nicosia *et al.*, 1990). Conversely, the absence of cooperativity for the XL-HOM polypeptide (Hill slope = 1), along with the evidence that XL-HOM is monomeric in solution (Figure 1B), indicates that the protein monomer can bind to the target sequence and that there is no enhancement of binding via protein–protein interactions. Consistent with the analysis of the cooperativity of DNA binding, dimers of BD bound to the α1-AT site in a single concerted step (Figure 2B, top panel). The isolated XL-homeodomain, in contrast, could no longer dimerize and occupied the symmetry-related binding site in two steps: the first retarded band to appear reflects the binding to only one half of the α1-AT site; the higher shift, observed only at the highest protein concentration used, suggests that a second XL-HOM molecule can subsequently fill the remaining half-site (Figure 2B, bottom panel).

In order to compare the DNA binding specificity of the BD and the XL-HOM proteins, we performed gel retardation experiments using, in addition to the α1-AT site, a second duplex oligonucleotide of the same length, but of a sequence unrelated to any known LFB1/HNF1 binding site (oligonucleotide C; Civitareale *et al.*, 1989). Whereas binding of BD to the C oligonucleotide was not detectable even at the highest protein concentration used (10^{-7} M),

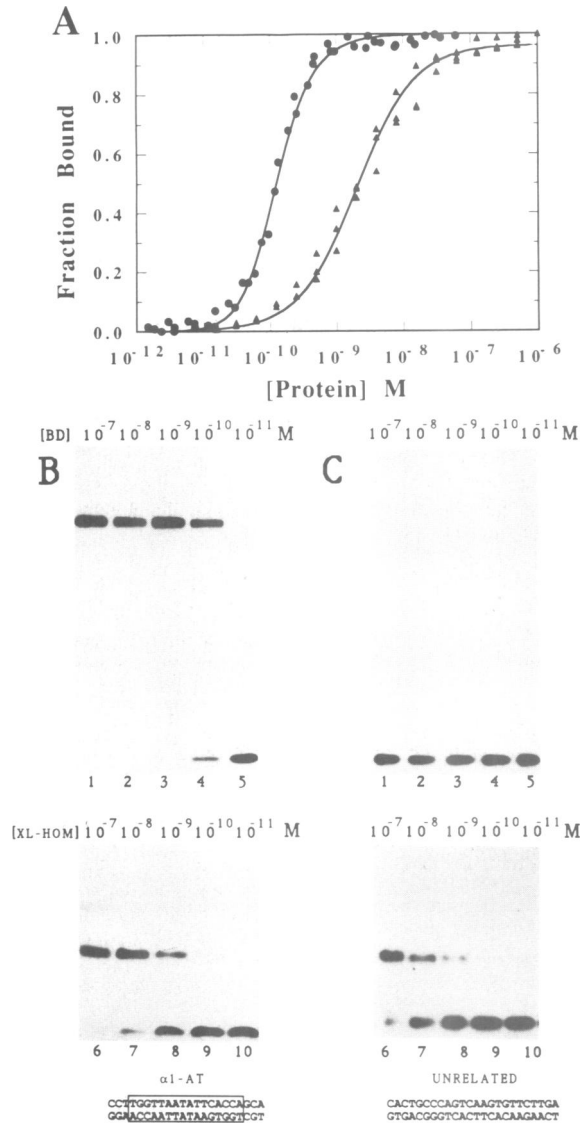

Fig. 2. Comparison of relative DNA binding affinities and specificities of the BD and the XL-HOM mutants. Various amounts of the purified BD and XL-HOM proteins were incubated with a constant limiting amount of labelled α1-AT and C duplexes in the absence of any non-specific competitor DNA. (**A**) Measurement of apparent binding constants of the intact DNA binding domain and isolated XL-homeodomain of LFB1/HNF1 for the α1-AT site. Bound and free DNA were separated on a bandshift gel and quantified by scintillation counting. Data from three independent experiments are shown for both the BD (●) and the XL-HOM (▲) proteins. The curves interpolating the data points were obtained by numerical simulation using the model described in Materials and methods with the following equilibrium constants: XL-HOM: $K_{d(XL\text{-}HOM)} = 2\times10^{-9}$ M, where $K_{d(XL\text{-}HOM)}$ is the equilibrium dissociation constant of the XL-homeodomain complex with the α1-AT site. BD: $K_{dim} = 5\times10^{-10}$ M; $K_{d(BD)} = 1.6\times10^{-11}$ M, where K_{dim} is the equilibrium dissociation constant for BD dimerization, and $K_{d(BD)}$ is the dissociation constant of the complex between the BD dimer and the α1-AT site. We note that in the BD case this is not the unique combination of values that generates a curve consistent with the experimental data. (**B** and **C**) These two panels compare representative mobility shift assays performed with the α1-AT (B) and the unrelated C (C) sites. Purified BD (lanes 1–5) and XL-HOM (lanes 6–10) proteins were used in the indicated amounts.

XL-HOM bound to the completely unrelated sequence only ~10-fold less efficiently than to the α1-AT site (Figure 2C). This observation suggests that the latter polypeptide possesses a very relaxed specificity of binding.

To assess more quantitatively the difference in sequence specificity between the BD and the XL-HOM proteins, we also performed binding competition experiments: the complex of each protein with the radiolabelled α1-AT site was challenged with increasing amounts of cold C oligonucleotide. We found that a 5000-fold molar excess of cold C over labelled α1-AT was required to compete 50% of the complex with BD. Conversely, only a 10-fold molar excess of cold competitor was enough to displace 50% of the XL-HOM complex with the α1-AT site (data not presented).

From this set of experiments, we conclude that the XL-homeodomain of LFB1/HNF1 retains the ability to form a stable complex with the α1-AT site, but it does so with an apparent affinity that is less than that of the intact binding domain. Furthermore, the isolated XL-homeodomain shows very little sequence specificity. The loss of sequence specificity is accompanied by the inability of this latter polypeptide to bind to DNA as a dimer.

The N-terminal dimerization domain is not sufficient to restrict specificity of the XL-homeodomain

The observation that the homeo-related domain of LFB1/HNF1 binds to DNA recognition sites with markedly relaxed specificity, raised the question of what other protein domains conferred stringency of binding.

We thought that the dimerization of the LFB1/HNF1 homeodomain might play a major role in discriminating between specific and non-specific DNA sequences. The observation that the DNA sequences of all the LFB1/HNF1 responsive elements characterized thus far do indeed fit to a palindromic consensus, lent support to this hypothesis.

In an earlier study we showed that the dimerization of LFB1/HNF1 is mediated by a 32 residue N-terminal region, named domain A (Nicosia *et al.*, 1990). In order to test whether dimerization of the XL-homeodomain through the A-domain was by itself sufficient to add specificity to DNA binding, we expressed in *E.coli* a mutant protein which bears the dimerization domain of LFB1/HNF1 fused at the N-terminus of the XL-homeodomain (mutant DIM-HOM, Figure 1B). The oligomeric composition of this protein was investigated by gel-filtration chromatography on a Superdex 75 column: the protein eluted at an apparent molecular weight of ~35 kDa, consistent with the formation of dimers. As a control, the XL-HOM peptide was found to be monomeric under the same conditions (Figure 1B).

Subsequently, we performed DNA binding experiments similar to the ones described above. A DNA–protein complex was always detected, the mobility of which was consistent with a stoichiometry of two molecules of protein bound per molecule of DNA probe, regardless of the oligonucleotide probe used (Figure 3). Thus, in spite of the effective protein dimerization, the DNA binding properties of DIM-HOM mirrored that of the monomeric XL-HOM rather than those of the dimeric BD: half-saturation of the α1-AT site was reached at a protein concentration of $\sim10^{-9}$ M and a stable DNA–protein complex was always observed also with the unrelated C oligonucleotide (Figure 3).

Bandshift competition experiments (data not presented) showed that the non-specific complex of DIM-HOM with the C oligonucleotide was ~50-fold less stable than the complex of the same protein with the α1-AT site. These observations suggest that dimerization by itself confers only

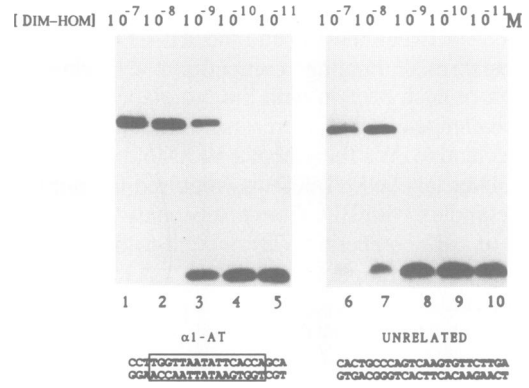

Fig. 3. DNA binding affinity and specificity of the DIM-HOM mutant. The indicated amounts of the purified DIM-HOM protein were incubated with a constant limiting amount of labelled α1-AT or C duplexes in the absence of any non-specific competitor DNA. Bound and free DNA were quantified as in Figure 2. The two panels show representative mobility shift assays performed with either the α1-AT (lanes 1–5) or the unrelated C (lanes 6–10) sites.

modest DNA binding specificity to the LFB1/HNF1 XL-homeobox.

The finding that a complex between the monomeric form of this protein and DNA was never observed throughout the range of concentrations investigated is in line with our previous observation that region A is sufficient for the dimerization of the LFB1/HNF1 protein (Nicosia *et al.*, 1990; De Francesco *et al.*, 1991).

Region B enhances site specificity and mediates the formation of DNA-stabilized protein dimers

To evaluate directly the contribution of the B region to site specificity, we characterized the DNA binding properties of a recombinant protein comprising only the B and the C regions of LFB1/HNF1. Because part of this region shows some sequence similarity to the POU-A box of POU proteins, we named this mutant Ψ-POU (Figure 1A and B).

The Ψ-POU protein was found to be monomeric in solution up to the micromolar concentration range, as judged by gel-filtration chromatography (Figure 1B). Even if this observation does not rule out the possibility of weak or transient protein–protein interactions, the monomer is expected to be the predominant species at the concentration used for our binding assays. A constant amount of the α1-AT radiolabelled oligonucleotide was titrated with increasing amounts of the purified protein (Figure 4A and B). Two DNA–protein complexes could be detected in the gel retardation experiments (Figure 4B). The mobility of each of the two complexes is consistent with one (complex I) and two (complex II) molecules respectively of the protein being bound to DNA.

At low protein concentrations, complex I was the predominant species; progressively increasing the protein concentration rapidly enhanced the formation of complex II over complex I. At saturating protein concentrations, complex II was the only species detectable. The relative amounts of DNA in complex I, complex II and the free probe were quantified and are summarized in Figure 4A.

Complex I initially increases but never exceeds 15% of the total DNA. These binding features indicate that two monomers of the mutant protein bind in a cooperative fashion to one molecule of the imperfect DNA palindrome (Tsai

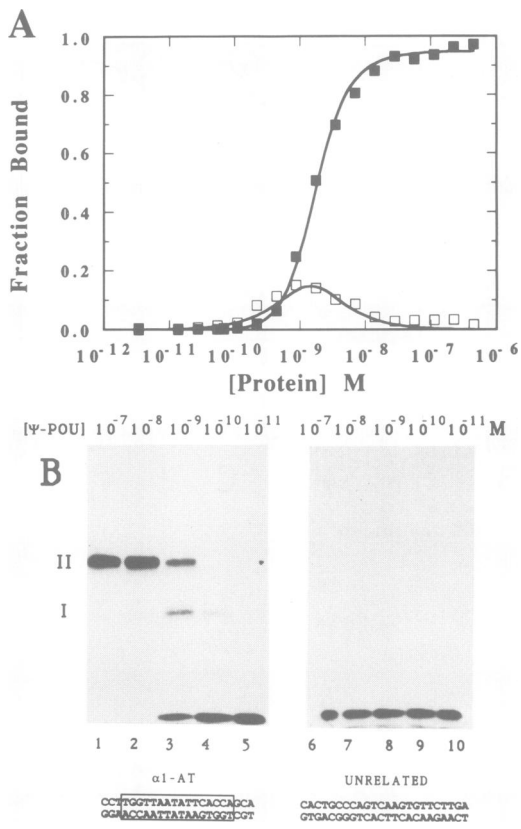

Fig. 4. Cooperative binding and specificity of the Ψ-POU mutant. Various amounts of the purified Ψ-POU protein were incubated with a constant limiting amount of labelled α1-AT or C duplexes in the absence of any non-specific competitor DNA. Bound and free DNA were quantified as in Figure 2. (**A**) Quantification of the fraction of α1-AT probe bound by the monomeric (□) and the dimeric (■) forms of the Ψ-POU protein. The curves interpolating the data points were obtained by numerical simulation using the model described in Materials and methods with the following equilibrium constants: $K_{d(I)} = 6.0\times10^{-9}$ M; $K_{d(II)} = 5.3\times10^{-10}$ M, where $K_{d(I)}$ and $K_{d(II)}$ are the macroscopic dissociation constants of the monomeric and dimeric complexes respectively of Ψ-POU with the α1-AT site. (**B** and **C**) These two panels compare representative mobility shift assays performed with the α1-AT (B) and the unrelated C (C) sites. Purified Ψ-POU protein was used in the indicated amounts.

et al., 1989): the occupation of a DNA half-site by the first Ψ-POU monomer increases the binding affinity of a second protein molecule to the remaining half-site. This cooperativity most probably arises from protein–protein contacts between adjacently bound proteins. It should be noticed that the Ψ-POU mutant protein is monomeric in solution, but predominantly dimeric when bound to DNA, despite the fact that it lacks the 'dimerization domain'. There is no paradox in this observation: the free energy at a protein interface required to form a dimer in solution must be in the range 10–15 kcal/mol in order to overcome the loss of translational and rotational degrees of freedom, whereas the free energy required for highly cooperative binding to adjacent half-sites is significantly lower (Harrison and Aggarwal, 1990).

The Ψ-POU protein did not bind significantly to the non-specific C oligo at any of the protein concentrations used in our experiments (Figure 4D). Furthermore, bandshift competition experiments (data not presented) indicated that a 500-fold molar excess of the cold C oligonucleotide is

required to compete away 50% of the complex of the Ψ-POU protein with radiolabelled α1-AT oligonucleotide.

Altogether, the experiments described above indicate that the B region of LFB1/HNF1 is implicated in restricting the specificity of DNA binding to that of the intact LFB1/HNF1 protein and that the same region is involved in the DNA-dependent dimerization of two protein molecules.

It is worth pointing out that the overall stability of the Ψ-POU protein complex with the α1-AT oligonucleotide, measured as the concentration of protein required to saturate 50% of the binding sites (10^{-9} M), did not differ significantly from that measured for the isolated XL-homeodomain with the same DNA site. In contrast, while the XL-HOM mutant bound tightly to non-specific DNA, the Ψ-POU protein failed to do so. This finding suggests that the increased sequence specificity displayed by the Ψ-POU over the isolated XL-homeodomain is not obtained by establishing extra contacts with specific DNA, but rather by preventing the interaction with the non-specific sequences. This observation, along with the finding that the Ψ-POU protein has diminished affinity for a LFB1/HNF1 half-site (compare lanes 1−4 of Figure 6B with the corresponding lanes of Figure 6D) points to a possible negative role of region B: sequence-specificity could be achieved through an 'inhibition' of DNA binding, mediated by the B-region, that can be overcome, on the α1-AT oligonucleotide, by the DNA-mediated dimerization of the protein. We tried to assess whether the inhibitory function depended on a portion of region B different from that involved in the DNA-dependent protein−protein interaction. To this end we constructed two mutants carrying, in addition to the XL-homeodomain, only the region of homology to the POU-A box (mutant POU-A-HOM) or the region comprising amino acids 101−155 of the protein (mutant ΔPOU-A). Interestingly, both of these mutant proteins have lost the ability to bind to any DNA site (data not shown), suggesting that the inhibitory function is retained by either mutant, while the presence of an intact B-region is required for efficacious protein−protein interaction.

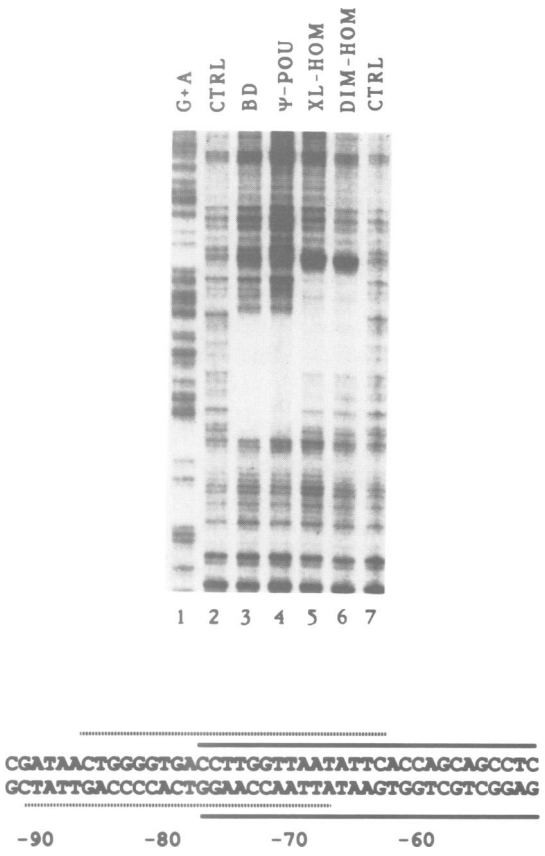

Fig. 5. DNase I footprinting of the LFB1/HNF1 mutants on the α1-antitrypsin promoter. DNase footprinting was carried out as described in Materials and methods. The final protein concentration in the binding mixture was 10^{-9} M for the BD mutant (lane 3) and 10^{-7} M for the Ψ-POU (lane 4), XL-HOM (lane 5) and DIM-HOM (lane 6) mutants. The digestion pattern is shown only for the top strand. Lanes 2 and 7 show the cleavage pattern of the unbound probe. The sequence corresponding to the footprint was determined by a G+A ladder (lane 1). Below the autoradiogram, the protection pattern obtained on both DNA strands using the BD and Ψ-POU mutants (solid line) is compared with that obtained using the XL-HOM and DIM-HOM mutants (dotted line).

Region B is responsible for the correct orientation of the two XL-homeodomains in the LFB1/HNF1 dimer

We wondered whether the different features of DNA binding displayed by the various mutant proteins in our bandshift assays reflected different ways of contacting the DNA duplex. In order to get a clearer picture of how the various polypeptides contact DNA, we performed DNase I protection experiments on a fragment derived from the α1-AT promoter. The result is shown in Figure 5. The BD (lane 3) and the Ψ-POU (lane 4) proteins yielded a protection pattern identical to that observed on this promoter using native LFB1/HNF1 extracted from rat liver nuclei (Monaci *et al.*, 1988). In contrast, a distinctly different region was protected by the XL-HOM (lane 5) and DIM-HOM (lane 6) mutants: only the most conserved 5′ half of the LFB1/HNF1 binding site was fully protected from DNase I activity. Furthermore, the boundaries of the footprint observed with these two latter mutants extended further upstream towards the 5′ end of the LFB1/HNF1 site. Independent gel retardation experiments (data not presented) revealed that this region contains an additional high-affinity binding site for the XL-HOM protein that is not recognized by the intact binding domain, nor is its sequence related to any known LFB1/HNF1 binding site.

The results just described reinforce the notion that the XL-HOM and DIM-HOM mutants have a relaxed or even altered specificity of DNA binding. This eventually results in profoundly altered geometry of the DNA−protein complex. In contrast, the mutant Ψ-POU protein can position the two XL-homeodomains on the symmetric DNA sites indistinguishably from the intact binding domain.

The above observation raised the possibility that the protein−protein interactions mediated by the presence of the B domain contributed critically to sequence specificity by holding one XL-homeodomain in a defined position relative to the other. In order to assess whether this was the case, we designed three additional binding sites, based on the LFB1/HNF1 consensus sequence: one bearing only one half of the consensus sequence (BHalf) and two in which the two idealized half-sites were organized as inverted (Pal) or direct (DirRep) repeats. In designing these latter two oligonucleotides, care was taken to mantain the two half-site binding surfaces on the same face of the DNA duplex, as they are normally found in the natural recognition sites (R.De Francesco, unpublished observation). The results of the gel retardation experiments, using these oligonucleotides as

probes, are presented in Figure 6. It should be noted that these experiments were carried out in the presence of a large excess of poly(dI-dC)·poly(dI-dC) in order to maximize the differences between specific and non-specific DNA binding. The apparent affinities of a given mutant for the various sites may therefore differ from those determined in the absence of competitor DNA.

We first analyzed the DNA binding properties of the intact LFB1/HNF1 binding domain on the three artificial sites (Figure 6A). Because of the strong protein−protein interactions mediated by the A-region (De Francesco *et al.*, 1991 and above), BD still bound to the half-site as a dimer (lanes 1−4). We believe that one of the two protein molecules in this complex contacts DNA non-specifically. The presence of a second half-site arranged as a direct repeat increases the stability of the DNA−protein complex by a factor of ~10 (lanes 5−8). However, a much more dramatic effect was observed when the two half-sites were arrayed in an inverted repeat as they are normally found in natural responsive elements: the BD protein displayed an apparent affinity for the Pal site that was about three orders of magnitude greater than for the half-site (lanes 9−12). These findings show that the LFB1/HNF1 dimer was capable of recognizing not only the information contained in the sequence of each half-site, but also the arrangement of the two half-sites with respect to each other. The isolated XL-homeodomain, in contrast, occupied the two half-sites in two independent steps and regardless of the orientation (Figure 6B, lanes 1−12).

As shown in Figure 6C, the DIM-HOM polypeptide, which is a dimer in solution, bound as a dimer to the half-site (lanes 1−4), although its affinity was less than that of either the inverted or the direct repeat (lanes 5−12). Since its apparent affinities for these two latter sites were nearly identical, this mutant seems to contact both half-sites independently of their arrangement. This behaviour, expected on the basis of the modest specificity displayed by the DIM-HOM dimer, suggests that the two XL-homeodomains are not constrained in this protein.

Finally, we studied the interaction of the Ψ-POU protein complex with the same three sites (Figure 6D). When challenged with the half-site (lanes 1−4), the Ψ-POU mutant forms a low-affinity complex in which the protein is predominantly monomeric. This is consistent with the finding that dimerization of this polypeptide is driven by the presence of an appropriate DNA site. Duplication of the half-site as a direct repeat (lanes 5−8) augmented to some extent the formation of the DNA-driven dimer, but was not sufficient to restore the tight binding observed when the two half-sites possess dyad symmetry as in the inverted repeats of the palindrome (lanes 9−12). This observation suggests that the orientation of the two XL-homeodomains in the bound dimer of Ψ-POU is restrained by the DNA-induced protein−protein interactions between two adjacent monomers. We believe that the dimer contacts formed by the Ψ-POU molecule are similar to those involved in setting the geometry of the BD dimer, since the two mutants displayed a similar rank order of DNA site affinity.

The DNA-independent dimerization domain of LFB1/HNF1 can be replaced by the GCN4 leucine zipper

The protein−protein contacts established within the Ψ-POU mutant were strong enough to define the structure of the bound dimer, yet insufficient to mantain its integrity when not associated with DNA. In contrast, the intact binding domain, BD, formed dimers in solution, regardless of the presence of specific DNA sites, and bound to the α1-AT site one order of magnitude more tightly.

It therefore seems as if the N-terminal dimerization domain had a specific role in increasing the affinity of the LFB1/HNF1 protein for the 2-fold symmetrical site, facilitating the formation of LFB1/HNF1 dimers in the

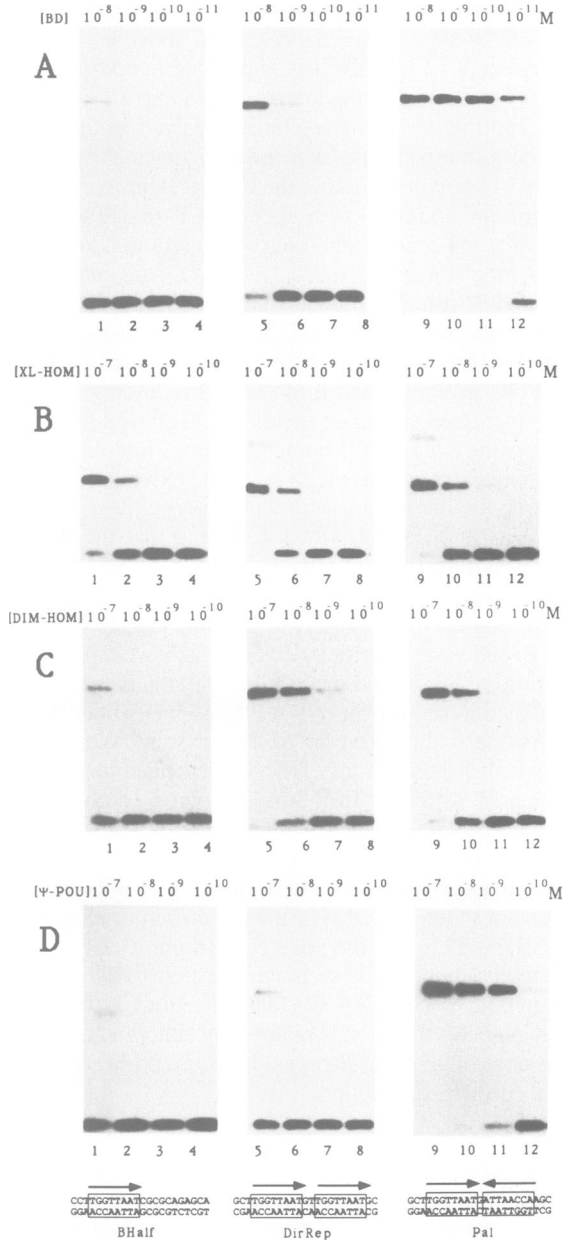

Fig. 6. Effect of the relative orientation of the half-sites on the DNA binding properties of the different LFB1/HNF1 mutants. The four panels show the relative binding affinities of the BD (**A**), XL-HOM (**B**), DIM-HOM (**C**) and Ψ-POU (**D**) mutant proteins for the half-site (BHalf, lanes 1−4), the direct repeat (DirRep, lanes 5−8) and the palindrome (Pal, lanes 9−12) duplex oligonucleotide respectively. The concentrations of protein used in each experiment are indicated at the top of each panel. Binding reactions were carried out in the presence of an excess of poly(dIdC)·poly(dIdC) and therefore the dissociation constants of the various complex estimated in this experiment cannot be compared directly with those determined from the experiments shown in Figures 2−4.

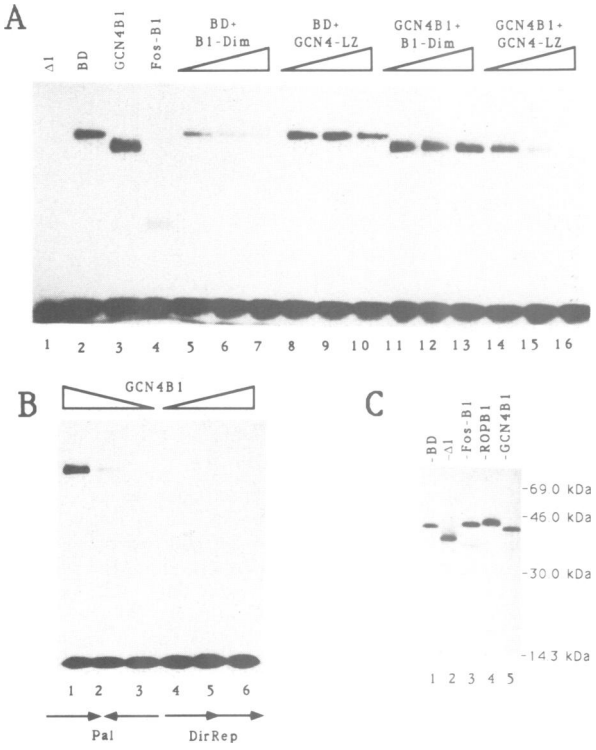

Fig. 7. DNA binding properties of the GCN4B1 and FosB1 dimerization domain substitution mutants of LFB1/HNF1. LFB1/HNF1 derivatives were produced *in vitro* by transcription and translation of the appropriate template as described in Materials and methods and evaluated for their ability to bind to the α1-AT, Pal and DirRep synthetic duplexes in bandshift experiments. (**A**) The dimerization domain deletion mutants Δ1 (lane 1) and the dimerization domain substitution mutants GCN4B1 (lane 3) and FosB1 (lane 4) were compared with the intact BD (lane 2) for their ability to bind to the α1-AT probe. The complexes formed by BD (lanes 5–10) and GCN4B1 (lanes 11–15) were challenged, 15 min prior to loading, with increasing amounts of B1-Dim and GCN4-LZ peptides, corresponding to the dimerization domains of LFB1/HNF1 and GCN4 respectively. The concentrations of competitor peptide used were 0.12 μM (lanes 5, 8, 11 and 14), 1.2 μM (lanes 6, 9, 12 and 15) and 12 μM (lanes 7, 10, 13 and 15). Similar amounts of protein, roughly equivalent to a final concentration of 8×10^{-11} M, were used in all the lanes, except for lanes 1 and 4, where the final concentration of protein in the bandshift mixture was ~4-fold higher. (**B**) The orientation of the two XL-homeodomains in the dimerization domain substitution mutant GCN4B1 was probed using the Pal (lanes 1–3) and the DirRep (lanes 4–6) duplexes in bandshift experiments. The protein concentration was $\sim10^{-11}$ M (lanes 1 and 6), 2×10^{-12} M (lanes 2 and 5) and 4×10^{-13} M (lanes 3 and 4). (**C**) The ^{35}S-labelled LFB1/HNF1 dimerization domain substitution mutants obtained by *in vitro* translation were visualized on a 12% SDS–PAGE. Lane 1, D1; lane 2, BD; lane 3, Fos-B1; lane 4, ROPB1; lane 5, GCN4B1.

absence of DNA. In order to assess whether the function of the N-terminal dimerization domain was limited to DNA-independent dimerization, or whether it also had a role in further defining the structure of the bound protein dimer, we constructed a series of chimeric proteins in which we substituted region A of LFB1/HNF1 with heterologous dimerization domains that have very different structural arrangements. If the LFB1/HNF1 dimerization domain imposed a structural constraint on the protein, then it would be expected that the resulting chimeric proteins would no longer recognize the LFB1/HNF1 binding sites.

We chose to use *in vitro*-translated proteins for the analysis of the activity of the dimerization domain substitution mutants. The reason for this choice was twofold. First, the quantification of the protein of interest does not require purification to homogeneity, thus speeding up the analysis of the mutants. Second, because of the low amount of the proteins translated *in vitro* and the presence of competitor DNA, we could clearly detect a DNA–protein complex only with those mutants that bind DNA with an affinity and a specificity comparable to that of the wild-type protein: a deletion mutant lacking the N-terminal dimerization domain (mutant Δ1, Nicosia *et al.*, 1990) produced only a very faint band (often observable only at very long exposure times) with the mobility expected for a protein monomer–DNA complex (Figure 7, lane 1).

We initially assembled a construct in which the sequence encoding the authentic 32 N-terminal amino acids of LFB1/HNF1-BD was removed and replaced with the sequence encoding the 32 amino acid leucine zipper from the yeast DNA binding protein GCN4. This domain is known to fold as a parallel coiled-coil (Oas *et al.*, 1990; Saudek *et al.*, 1991).

The chimeric protein (GCN4B1) was synthesized *in vitro*, incubated with the α1-AT site in the presence of an excess of non-specific competitor DNA and electrophoresed in a non-denaturing polyacrylamide gel. A retarded band was clearly seen (Figure 7, lane 3) which displayed a mobility similar to that of the LFB1/HNF1-BD–DNA complex generated by the same method (Figure 7, lane 2). As a negative control, we constructed a similar protein, but using the leucine zipper of the c-*fos* proto-oncogene (FosB1). The failure of this latter chimeric protein to bind DNA tightly as a dimer (Figure 7, lane 4), reflects the inability of the c-*fos* leucine zipper to self-dimerize (O'Shea *et al.*, 1989).

In addition we found that, whereas DNA binding by LFB1/HNF1-BD was competed by a molar excess of a synthetic peptide corresponding to its dimerization domain (B1-dim; Figure 7, lanes 5–7), GCN4B1 still bound to the cognate sequence under these conditions (Figure 7, lanes 8–10). Conversely, a peptide encompassing the leucine zipper region of GCN4 (GCN4-LZ; Saudek *et al.*, 1991) inhibited the DNA binding activity of GCN4B1 but not that of LFB1/HNF1-BD (Figure 7, compare lanes 11–13 with lanes 14–16).

The specificity of the GCN4B1 chimeric protein was further investigated using, in addition to the α1-AT binding site, two probes containing the perfect palindrome and the direct repeats. As depicted in Figure 7B, GCN4B1 paralleled the affinity of the wild-type LFB1/HNF1-BD protein for these two sites.

These results demonstrate that the 32 N-terminal amino acids are indeed responsible for protein dimerization and that dimerization is a requisite for high-affinity sequence-specific DNA binding by LFB1/HNF1. However, the native dimerization domain can be replaced by the leucine zipper domain of yeast GCN4 without loss of either DNA binding affinity or specificity.

The ROP protein, a dimer of antiparallel hairpins, can functionally replace the dimerization domain of LFB1/HNF1

The leucine zipper of GCN4 folds as a dimer of amphipathic α-helices arranged in parallel. It has previously been suggested by Pastore *et al.* (1991) that the dimerization domain of LFB1/HNF1 also adopted a parallel configuration.

In order to see whether the helical topology of the dimerization domain of LFB1/HNF1 was coupled to the

orientation of the two protein monomers on the DNA molecule, we designed an additional mutant protein in which the natural dimerization domain was substituted by the whole sequence of the ROP protein (ROPB1). ROP is a small dimeric protein that controls the replication of the ColE1 plasmid via regulation of RNA−RNA interactions (Helmer-Citterich *et al.*, 1988). Each monomer (63 amino acids) consists of two α-helices connected by a tight turn, the whole molecule forming a highly regular four-helix bundle arising from the antiparallel juxtaposition of the two monomer hairpins (Banner *et al.*, 1987). The packing of the hydrophobic residues of the ROP α-helices is prototypic for antiparallel coiled-coils.

While the ROP protein itself does not show any detectable DNA binding activity with our probe (not shown), the chimeric protein (ROPB1) bound to the α1-AT DNA site as tightly as the LFB1/HNF1-BD molecule (Figure 8A, lane 1 and 2). Cotranslation of the ROPB1 mRNA with that encoding the wild-type ROP protein resulted in a marked inhibition of DNA binding, consistent with the hypothesis that the chimeric protein dimerizes by virtue of the ROP folding unit (Figure 8A, lanes 3−5). In addition, as found for the GCN4B1 mutant, ROPB1 formed a tight complex with the DNA probe containing the idealized palindrome (Figure 8B, lanes 1−3), but failed to bind significantly to the inverted repeats (Figure 8B, lanes 4−6).

We draw the conclusion that dramatically altering the spatial arrangement of the dimerization domain is not paralleled by a substantial change of the ability of LFB1/HNF1 to bind to a specific DNA recognition sequence. This finding suggests that the N-terminal dimerization domain is flexibly linked to the rest of the protein and plays no role in determining the orientation of the protein monomers on the DNA binding site. Rather, the role of the dimerization domain of LFB1/HNF1 appears to be limited to guiding the formation of LFB1/HNF1 dimers even in the absence of the appropriate DNA binding site. In this way, the energy deriving from the strong protein−protein interactions can be used to counteract the loss of entropy associated with dimerization and formation of a ternary complex with DNA. The net result is an increased affinity for the palindromic DNA sites.

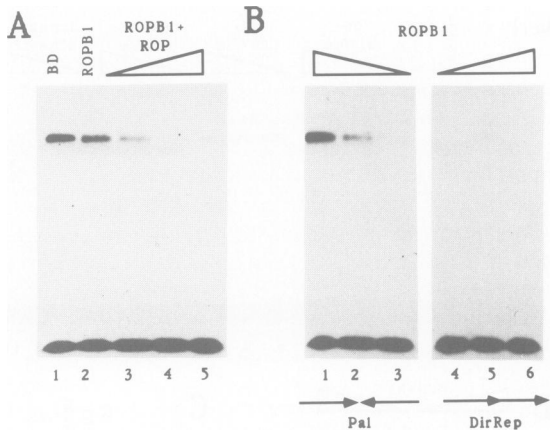

Fig. 8. DNA binding properties of the dimerization domain substitution mutant ROPB1. The ROPB1 dimerization domain substitution mutant was produced *in vitro* by transcription and translation of the appropriate template as described in Materials and methods and evaluated for its ability to bind to the α1-AT, Pal and DirRep synthetic duplexes in bandshift experiments. (**A**) The dimerization domain substitution mutant ROPB1 (lane 2) was compared with the intact BD (lane 1) for its ability to bind to the α1-AT probe. The concentration of each mutant protein was estimated to be $\sim 8\times10^{-11}$ M. The complexes formed by ROPB1 were competed by co-translating the mRNA encoding wild-type ROP protein in a ratio of 0.5:1 (lane 3), 1:1 (lane 4) or 3:1 (lane 5) with the mRNA of ROPB1. (**B**) The orientation of the two XL-homeodomains in the dimerization domain substitution mutant ROPB1 was probed using the Pal (lanes 1−3) and DirRep (lanes 4−6) duplexes in bandshift experiments. The protein concentration was $\sim 10^{-11}$ M (lanes 1 and 6), 2×10^{-12} M (lanes 2 and 5) and 4×10^{-13} M (lanes 3 and 4).

Discussion

Transcription factor LFB1/HNF1 is a sequence-specific DNA binding protein. All the information for high affinity and sequence-specific binding is contained within the N-terminal 281 amino acids: a recombinant protein spanning this region bound as a dimer to the LFB1/HNF1 responsive element contained within the promoter of α1-AT at concentrations of $<10^{-10}$ M and 5000-fold more tightly than to an unrelated sequence.

We investigated the mechanism by which these properties are brought about. Our conclusions are summarized below. All the characterized LFB1/HNF1 binding sites fit a palindromic consensus sequence in which two identical half-sites are arranged as inverted repeats. Using artificial binding sites containing ideal LFB1/HNF1 half-sites, arranged either in a palindrome as in the consensus binding site or as direct repeats, we demonstrated that the binding domain of LFB1/HNF1 requires two types of information: (i) the sequence of the two half-sites, and (ii) their relative arrangement on the DNA molecule. This finding suggests that the relative orientation of the two XL-homeodomains is constrained in the protein dimer.

We showed that region B contributes critically to site recognition by two mechanisms. First, the presence of region B seems to decrease the ability of the XL-homeodomain of LFB1 to bind non-specific DNA sequences, perhaps by direct masking of some of the contacts to the DNA backbone or via subtler conformational effects. This eventually results in a general 'inhibition' of DNA binding, mediated by the B-region, that can be relieved by the dimerization of the protein. An attractive possibility is that the presence of region B affects the conformation of the adjacent N-terminal basic residue cluster of the XL-homeodomain: the homologous region of the *Antennapedia* homeodomain, which is unstructured in the uncomplexed polypeptide (Billeter *et al.*, 1990), becomes ordered upon DNA binding and is responsible for several minor groove contacts, both to specific base pairs and to the phosphate backbone (Otting *et al.*, 1990; Percival-Smith *et al.*, 1990). Restricting the flexibility of this N-terminal arm could increase the protein sequence specificity by allowing only a subset of all the possible contacts.

Secondly, we were able to show that the dimer contacts mediated by the presence of region B, albeit not strong enough to drive DNA-independent dimerization, impose a 2-fold symmetry on the molecule, thus defining the reciprocal orientation of the two XL-homeodomains in the bound dimer. In this way, the B region increases the site-specificity of the LFB1/HNF1 XL-homeodomain severalfold

by permitting recognition of only those sequences containing properly oriented half-sites.

Profound alteration of the dimerization domain geometry influences neither the specificity nor the affinity of the LFB1/HNF1 protein for its binding sites. This finding suggests that the distal dimerization domain of LFB1/HNF1 is connected to the rest of the molecule via a flexible hinge region. The symmetry of the DNA binding-core of the protein is therefore uncoupled from the constraint imposed by the structure of the dimerization domain. Consistent with this conclusion, we showed in an earlier report (Nicosia *et al.*, 1990) that the 67 residue region between the A and the B domains could tolerate a variety of mutations.

In this scenario, the role of the N-terminal dimerization domain seems to be that of guiding the formation of protein dimers, even in the absence of DNA, in order to raise the affinity of the protein for specific DNA sites.

Another very important role of the dimerization domain of LFB1/HNF1 could be that of providing the molecular surface for the DNA-independent interaction with related transcription factors, such as LFB3/vHNF1 (De Simone *et al.*, 1991). Heterotypic dimerization has in fact recently emerged as an important way to expand the regulatory diversity of transcriptionally active proteins (Lamb and McKnight, 1991).

The properties of the isolated XL-homeodomain of LFB1/HNF1 are in many respects similar to those reported for a number of other isolated homeodomains. They all appear to bind to their DNA cognate sequences as monomers and display very limited binding specificity (Laughon, 1991). There is therefore a need to understand what gives homeobox-containing proteins the strict specificity necessary for the regulatory function *in vivo*.

In the case of the yeast $\alpha 2$ mating protein, the only other homeoprotein shown to dimerize, it has been demonstrated that the target specificity is determined by the interaction with a second protein, MCM1, which sets the spacing and the orientation of the two $\alpha 2$ homeodomains such that they coincide with the position of the half-sites in the natural operators (Smith and Johnson, 1992). In the case of LFB1/HNF1, we showed that region B plays the same role as MCM1 in constraining the orientation of the two homeodomains relative to one another. Noticeably, the B region acts on the same polypeptide and therefore *in cis* rather than *in trans* as in the case of yeast MCM1.

The POU class of transcription factors (Herr *et al.*, 1988) is another well characterized example of homeobox-containing proteins in which a second protein domain, namely the POU-specific domain, augments the site-specificity of the homeodomain (Rosenfeld, 1991). However, the mechanism by which the POU-specific domain contributes to site-specificity has not yet been fully elucidated. Region B shares some sequence homology with the POU-A box of the POU-specific domain (Figure 1A). In contrast to LFB1/HNF1, members of the POU gene family are largely monomeric in solution. Interestingly, while POU proteins can bind DNA as monomers (Rosenfeld, 1991 and references therein), Oct-2, Pit-1 and CF1-a exhibit cooperative DNA-dependent dimerization on certain *cis*-active elements (LeBowitz *et al.*, 1989; Ingraham *et al.*, 1990; Treacy *et al.*, 1991). In addition, Oct-1 and Pit-1 have been shown to bind cooperatively as a heterodimeric complex on a responsive element within the prolactin promoter (Voss *et al.*, 1991). The POU-specific domain of these proteins appears to be required for site-specific DNA binding, as well as for the DNA-dependent protein−protein interactions. Given the similarity of function with the LFB1/HNF1 B region, it is tempting to speculate that the POU-A box of the POU-specific domain contributes to the sequence specificity of these proteins by determining the orientation of the two POU-homeodomains in the DNA-bound dimer.

Finally, we would like to underscore the similarities of functional architecture that exist between LFB1/HNF1 and one of the better characterized bacterial helix−turn−helix bacterial protein: the repressor encoded by the *cI* gene of λ phage (reviewed in Sauer *et al.*, 1990). Like in the case of LFB1/HNF1, the λ repressor dimer is the fundamental DNA binding species: a repressor dimer binds to a partially symmetrical 17 bp operator site. The dimerization domain of the protein is located at the C-terminus and is connected via a flexible linker to the N-terminal DNA binding core. The isolated N-terminal DNA binding domain, like our Ψ-POU mutant, is monomeric in solution but can establish enough dimer contacts to bind operator sequences cooperatively as a dimer. Furthermore, the protein−protein contacts within the N-terminal DNA binding domain are sufficient to define the structure of the bound dimer: a chimeric protein in which the entire C-terminal moiety has been substituted by a leucine zipper dimerization domain functions as a repressor *in vivo* (Hu *et al.*, 1990).

It is interesting to note that the affinity of DNA binding depends critically on the dimerization energy of the protein, in LFB1/HNF1 as well as in λ repressor. As a consequence, dissociation into monomers seems to be the event that limits DNA binding at protein concentrations below the dimerization constant (Chadwick *et al.*, 1970; Figure 2A). This makes the dimerization domain an ideal target to modulate the DNA binding activity of these two proteins. Recently, the cDNA coding for DCoH, a protein factor that associates with LFB1/HNF1, was cloned (Mendel *et al.*, 1991). The protein encoded by this cDNA appears to selectively stabilize LFB1/HNF1 dimers and its engagement in a heterotetrameric complex eventually results in a strong enhancement of LFB1/HNF1 transcriptional activity. Based on our present data, we would like to suggest that this effect is mediated at least in part by the increased resistance of the dimer to dissociation. The extra dimer stability would be in turn coupled to the thermodynamic stability of the DNA−protein complex, resulting in a better site occupancy at the protein concentration present *in vivo*.

Materials and methods

Synthetic oligonucleotides and plasmid construction

The following oligonucleotides, corresponding to the indicated binding sites, and their respective complementary strands, were synthesized: α1-AT, CCTTGGTTAATATTCACCAGCA; C, CACTGCCCAGTCAAGTGTTCTTGA; BHalf, CCTTGGTTAATCGCGCAGAG; DirRep, GCTTGGTTAATGTTGGTTAATGC; and Pal, GCTTGGTTAATGATTAACCAAGC. The constructs pT7.7BD and pT7.7XL-HOM, containing respectively the LFB1/HNF1 DNA binding domain (amino acids 1−281) and the XL-HOM sequence (amino acids 195−286) in the prokaryotic expression vector pT7.7, were generously given by P.Monaci and A.Nicosia.

The plasmid pT7.7Ψ-POU contained the coding sequence of LFB1/HNF1 from residue 101 to 281 and was constructed as follows: the *Bgl*II−*Bam*HI fragment from mutant SM12 (Nicosia *et al.*, 1990), containig the LFB1/HNF1 coding sequence starting from amino acid 101 and including ~1 kb of the 3′ untranslated region, was cloned in the *Eco*RI/*Bam*HI sites of pT7.7, after filling in the *Bgl*II and the *Eco*RI ends using Klenow. A

double-stranded (ds) oligonucleotide (TAACTGAGTAGGATCCTACTCA-GTTA), containing stop codons in all reading frames, was subsequently cloned in place of the *MscI–HindIII* fragment.

To construct the plasmid pT7.7DIM-HOM, the *NcoI–EcoRI* fragment spanning the LFB1/HNF1 dimerization domain up to P_{33} was excised from the mutant IM2 (Nicosia *et al.*, 1990). After filling in both ends with Klenow enzyme, this fragment was inserted upstream of the XL-homeodomain after blunting the *NdeI* site of pT7.7XL-HOM.

The plasmid pT7.7POU-A-HOM was constructed as follows: the DNA fragment encoding the region between M_{154} and Y_{286} of wtLFB1/HNF1 was generated by PCR and cloned in the *NdeI* site of pT7.7 after filling in the ends with Klenow. The ds oligonucleotide containing stop codons in all reading frames was subsequently inserted as described for pT7.7Ψ-POU.

pT7.7ΔPOUA was generated in two steps: mutant SM8 (Nicosia *et al.*, 1990) was cut with *BamHI* and religated, producing the 'a' deletion mutant lacking all the residues between K_{155} and E_{188}. The DNA fragment encoding the region between K_{101} and Y_{286} of this latter mutant was generated by PCR and cloned in the *SmaI* site of pT7.7. The ds oligonucleotide containing stop codons in all reading frames was subsequently inserted as described for pT7.7Ψ-POU.

To construct the LFB1/HNF1 dimerization domain substitution mutants LEUB1 and ROPB1, two DNA fragments coding for the leucine zipper sequence from R_{249} to R_{291} of the GCN4 protein and for the complete sequence of the ColE1 ROP protein respectively were amplified by PCR using appropriate primers. Total yeast DNA and the plasmid pEXColE1ROP (a gift from D.Tsernoglou) were used as templates in the PCR reactions. The two fragments were inserted, after filling in with the Klenow polymerase, into the *NcoI/SphI* sites of the LFB1/HNF1 mutant Δ1 (Nicosia *et al.*, 1990) from which the second *NcoI* site present in the LFB1/HNF1 coding sequence at bp 133 was eliminated by site-directed mutagenesis.

Protein preparation

Proteins from pT7.7BD, pT7.7XL-HOM, pT7.7DIM-HOM and pT7.7Ψ-POU were expressed in the *E.coli* strain BL21(DE3) as described by Studier and Moffat (1986). After a 3 h induction with 0.4 mM IPTG, cells were harvested and lysed by a French pressure cell in a buffer containing 20 mM Tris–HCl, pH 8.8, 100 mM NaCl, 1 mM dithiothreitol and 1 mM EDTA (buffer A). After removing the cellular debris by high-speed centrifugation (20 min at 50 000 *g*), the proteins of interest were precipitated by adding ammonium sulfate to 40% (BD), 50% (Ψ-POU) or 60% (XL-HOM and DIM-HOM) saturation. The protein pellets were resuspended and dialysed against buffer A and subsequently purified on a heparin–Sepharose CL4B column eluted with a gradient from 0.1 to 1 M NaCl. The fractions containing the proteins of interest were pooled, dialysed against buffer A and further purified on a Mono S column using the same gradient as above. Proteins were >95% pure at this stage, as determined by SDS–PAGE. After salt removal by dialysis, proteins were stored in buffer A at −80°C. The concentration of the purified protein was determined by optical absorbance at 280 nm using an extinction coefficient calculated from the deduced protein sequence as described by Cantor and Schimmel (1980). The protein concentrations indicated in the DNA binding experiments have been corrected for the fraction of active protein. The active protein concentration was determined by gel retardation DNA titration at protein and DNA concentrations well above the observed dissociation constant. Typically, the fraction of active protein ranged between 60 and 90%.

The dimerization domain substitution mutants were synthesized using an *in vitro* transcription/translation system in the presence of [^{35}S]L-methionine as previously described (Nicosia *et al.*, 1990). ^{35}S-labelled translation products were analysed by SDS–PAGE, and the gel was treated with Intensify (New England Laboratories), dried and autoradiographed.

The concentration of the LFB1-BD protein synthesized *in vitro* was estimated by scintillation counting of the TCA-precipitable material and was typically found to be between 0.5 and 1.0 fmol/μl of translation mixture. The concentrations of all the other *in vitro* synthesized mutants relative to that of the LFB1-BD protein were estimated by densitometric scanning of the autoradiographs.

Gel filtration chromatography

The native molecular weights of the purified LFB1/HNF1 derivatives were determined on a Pharmacia Superdex 75 HR 10/30 pre-packed column in PBS buffer. The flow rate was 0.5 ml/min and the absorbance was recorded at 280 nm. Bovine serum albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (30 kDa), cytochrome *c* (12.5 kDa) and the oxidized B-chain of bovine insulin (3.5 kDa) were obtained from Sigma and used as molecular weight standards.

DNA binding studies

Gel retardation experiments were carried out by mixing the appropriate amount of protein with radiolabelled ds oligonucleotide (0.1–20 fmol) in 20 μl of a buffer containing 20 mM Tris–HCl, pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT and 0.1% low-fat dried milk. 100 ng of poly(dI-dC)·poly(dI-dC) were used as non-specific competitor DNA when specified. Binding reactions were incubated for 15 min at room temperature. 5 μl of a 20% Ficoll solution per reaction were then added and the samples were loaded on to a 6% native acrylamide gel (30:1 acrylamide:*bis*-acrylamide; 0.25×TBE). Gels were run at 10 V/cm at room temperature until the bromophenol blue dye had migrated 10 cm into the gel. Gels were subsequently dried and autoradiographed. When quantification was neccessary, radioactive bands corresponding to free and bound DNA were excised and counted in Cerenkov mode.

The experimental data were interpolated by numerical simulation using a personal computer spreadsheet program (KaleidaGraph 2.0, Synergy Software). The following models were assumed. For the isolated XL-homeodomain, XL-HOM, the experimental data were fitted to a simple single-site binding isotherm (Affolter *et al.*, 1990). For the intact binding domain (BD) it was assumed that two monomers combined to form a dimer, followed by binding of the dimer (BD_2) to the α1-AT site (D). The overall reaction is

$$2\,BD + D \overset{K_{dim}}{\rightleftharpoons} BD_2 + D \overset{K_{d(BD)}}{\rightleftharpoons} BD_2 \cdot D$$

where K_{dim} is the equilibrium dissociation constant for BD dimerization, and $K_{d(BD)}$ is the dissociation constant of the complex between the BD dimer and the α1-AT site. A thorough description of this model can be found in Sauer *et al.* (1990).

The binding curves of the Ψ-POU to the α1-AT site (D) were generated by a model that describes the consecutive interaction of two monomers with the two half-sites:

$$2\Psi\text{-POU} + D \overset{K_{a(I)}}{\rightleftharpoons} \Psi\text{-POU}\cdot D + \Psi\text{-POU} \overset{K_{a(II)}}{\rightleftharpoons} \Psi\text{-POU}_2\cdot D$$

where $K_{a(I)}$ and $K_{a(II)}$ are the macroscopic association constant for the first and the second monomers of Ψ-POU, respectively.

According to this model, the fraction of α1-AT duplex complex with one (ϕ_I) and two (ϕ_{II}) protein molecules, are given by the following equations, modified from Senear and Brenowitz (1991):

$$\phi_I = \frac{K_{a(I)}\cdot[\Psi\text{-POU}]}{1 + K_{a(I)}\cdot[\Psi\text{-POU}] + K_{a(I)}\cdot K_{a(II)}\cdot[\Psi\text{-POU}]^2}$$

$$\phi_{II} = \frac{K_{a(I)}\cdot K_{a(II)}\cdot[\Psi\text{-POU}]^2}{1 + K_{a(I)}\cdot[\Psi\text{-POU}] + K_{a(I)}\cdot K_{a(II)}\cdot[\Psi\text{-POU}]^2}$$

DNase I footprinting was performed essentially as described by Lichtsteiner *et al.* (1987). DNA binding reactions were carried out in 50 μl of a buffer containing 12.5 mM HEPES, pH 7.9, 0.5 mM DTT, 20% glycerol, 2% polyvinylalcohol, 0.1% Nonidet P40. After 15 min at room temperature, 50 μl of a solution containing 10 mM $MgCl_2$, 5 mM $CaCl_2$ and 0.02 units of DNase I (Boehringer Grade II) were added. Digestion was allowed for 1 min at 25°C and then stopped by adding 80 μl of a solution containing 20 mM EDTA, 1% SDS, 0.2 M NaCl and 250 μg/ml yeast tRNA. The DNA used in the footprinting assay was a fragment comprising the −139/+2 region of the rat α1-antytrypsin promoter cloned into the M5/M9 vector (P.Monaci and A.Nicosia, unpublished).

Acknowledgements

We are extremely grateful to Drs Paolo Monaci and Alfredo Nicosia for their gift of the pT7.7BD and pT7.7XL-HOM plasmid constructs, as well as for their helpful suggestions. We wish to thank Drs Joseph Jiricny, Marco Bianchi, Carlo Toniatti, and Maurizio Sollazzo for critically reading the manuscript. We also want to acknowledge Philippe Neuner for the synthesis of the oligonucleotides used in this work and Yves Cully for helping us in the preparation of the figures.

References

Affolter,M., Percival-Smith,A., Müller,M., Leupen,W. and Gehring,W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4093–4097.

Banner,D.W., Kokkinidis,M. and Tsernoglou,D. (1987) *J. Mol. Biol.*, **196**, 657–675.
Billeter,M., Qian,Y.Q., Otting,G., Müller,M., Gehring,W.J. and Wütrich K. (1990) *J. Mol. Biol.*, **214**, 183–197.
Cantor,C.R. and Schimmel,P.R. (1980) In Cantor,C.R. and Schimmel,P.R. (eds), *Biophysical Chemistry, Part II: Techniques for the Study of Biological Structure and Function*. W.H.Freeman and Co., New York, pp. 380–381.
Chadwick,P., Pirrotta,V., Steinberg,R., Hopkins,N. and Ptashne,M. (1970) *Cold Spring Harbor Symp. Quant. Biol.*, **35**, 283–294.
Chouard,T., Blumenfeld,M., Bach,I., Vandekerckhove,J., Cereghini,S. and Yaniv,M. (1990) *Nucleic Acids. Res.*, **18**, 5853–5863.
Civitareale,D., Lonigro,R., Sinclair,A. and Di Lauro,R. (1989) *EMBO J.*, **8**, 2537–2542.
De Francesco,R., Pastore,A., Vecchio,G. and Cortese,R. (1991) *Biochemistry*, **30**, 143–147
De Simone,V., De Magistris,L., Lazzaro,D., Gerstner,J., Monaci,P., Nicosia,A. and Cortese,R. (1991) *EMBO J.*, **10**, 1435–1443.
Finney,M. (1990) *Cell*, **60**, 5–6.
Frain,M., Swart,G., Monaci,P., Nicosia,A., Stampfli,S., Frank,R. and Cortese,R. (1989) *Cell*, **59**, 145–157.
Harrison,S.C. and Aggarwal,A.K. (1990) *Annu. Rev. Biochem.*, **59**, 933–969.
Helmer-Citterich,M., Anceschi,M.M., Banner,D.W. and Cesareni,G. (1988) *EMBO J.*, **7**, 557–566.
Herr,W., Sturm,R.A., Clerc,R.G., Corcoran,L.M., Baltimore,D., Sharp,P.A., Ingraham,H.A., Rosenfeld,M.G., Finney,M., Ruvkin,G. and Horvitz,H.R. (1988) *Genes Dev.*, **2**, 1513–1516.
Hu,C.J., O'Shea,E.K., Kim,P.S. and Sauer,R.T. (1990) *Science*, **250**, 1400–1403.
Ingraham,H.A., Flynn,S.E., Voss,J.W., Albert,V.R., Kapiloff,M.S., Wilson,L. and Rosenfeld,M.G. (1990) *Cell*, **61**, 1021–1033.
Lamb,P. and McKnight,S.L. (1991) *Trends Biochem. Sci.*, **16**, 417–422.
Laughon,A. (1991) *Biochemistry*, **30**, 11357–11367.
LeBowitz,J.H., Clerc,R.G., Brenowitz,M. and Sharp,P.A. (1989) *Genes Dev.*, **3**, 1625–1638.
Lichtsteiner,S., Wuarin,J. and Schibler,U. (1987) *Cell*, **51**, 963–967.
Mendel,D.B. and Crabtree,G.R. (1991) *J. Biol. Chem.*, **266**, 677–680.
Mendel,D.B., Khavari,P.A., Conley,P.B., Graves,M.K., Hansen,L.P., Admon,A. and Crabtree,G.R. (1991) *Science*, **254**, 1762–1767.
Monaci,P., Nicosia,A. and Cortese,R. (1988) *EMBO J.*, **7**, 2075–2087
Nicosia,A., Monaci,P., Tomei,L., De Francesco,R., Nuzzo,M., Stunnenberg,H. and Cortese,R. (1990) *Cell*, **61**, 1225–1236.
Oas,T.G., McIntosh,L.P., O'Shea,E.K., Dahlquist,F.W. and Kim,P.S. (1990) *Biochemistry*, **29**, 2891–2894.
O'Shea,E.K., Rutkowsky,R., Stafford,W.F. and Kim,P.S. (1989) *Science*, **245**, 646–648.
Otting,G., Qian,Y.Q., Müller,M., Affolter,M., Gehring,W.J. and Wüthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.
Pastore,A., De Francesco,R., Barbato,G., Castiglione Morelli,M.A., Motta,A. and Cortese,R. (1991) *Biochemistry*, 30, 148–153.
Percival-Smith,A., Müller,M., Affolter,M. and Gehring,W.J. (1990) *EMBO J.*, **9**, 3967–3974.
Rosenfeld,M.G. (1991) *Genes Dev.*, **5**, 897–907.
Saudek,V., Pastore,A., Castiglioni Morelli,M.A., Frank,R., Gausepohl,H, and Gibson,T. (1991) *Protein Engng*, **4**, 519–529.
Sauer,R.T., Jordan,S.R. and Pabo,C.O. (1990) *Adv. Protein Chem.*, **40**, 1–61.
Senear,D.F. and Brenowitz,M. (1991) *J. Biol. Chem.*, **266**, 13661–13671.
Smith,D.L. and Johnson,A.D. (1992) *Cell*, **68**, 133–142.
Studier,F.W. and Moffatt,B.A. (1986) *J. Mol. Biol.*, **189**, 113–130.
Treacy,M.N., Xi,H. and Rosenfeld,M.G. (1991) *Nature*, **350**, 577–584.
Tsai,S.Y., Tsai,M.-J. and O'Malley,B.W. (1989) *Cell*, **57**, 443–449.
Voss,J.W., Wilson,L. and Rosenfeld,M.G. (1991) *Genes Dev.*, **5**, 1309–1320.

Received on June 11, 1992; revised on July 27, 1992